



# HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA



*Luis Alberto Duicela Guambi*  
2023



# HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA

*Luis Alberto Duicela Guambi*

**2023**

---

Herramientas estadísticas para la investigación agropecuaria

Universidad Técnica de Babahoyo

ISBN: 978-9942-606-24-2 (eBook)

ISBN: 978-9942-606-25-9 (Impreso)

Editado por:

Universidad Técnica de Babahoyo

Avenida Universitaria Km 2.5 Vía a Montalvo

Teléfono: 052 570 368

© Reservados todos los derechos 2023

Babahoyo, Ecuador

[www.utb.edu.ec](http://www.utb.edu.ec)

E-mail: [editorial@utb.edu.ec](mailto:editorial@utb.edu.ec)

ISBN: 978-9942-606-24-2



ISBN: 978-9942-606-25-9



Este texto ha sido sometido a un proceso de evaluación por pares externos.

Diseño y diagramación, montaje y producción editorial

Universidad Técnica de Babahoyo

Babahoyo – Los Ríos – Ecuador

Impreso: GRAFICAS ISAX

Tiraje: 500 Ejemplares

Guayaquil - Ecuador

*Queda prohibida la reproducción total o parcial de este libro en forma idéntica o modificada por cualquier medio mecánico o electrónico incluyendo fotocopia, grabación o cualquier sistema de almacenamiento o recuperación de información sin la autorización del autor.*

## Luis Alberto Duicela Guambi



Ingeniero agrónomo de la Escuela Superior Politécnica de Chimborazo (Ecuador), Magister en Ciencias Agrícolas del Instituto Tecnológico y de Estudios Superiores de Monterrey (México) y Doctor en Ciencias Agrarias de la Universidad del Zulia (Venezuela). Con formación en Alta gerencia en la Universidad Internacional SEK, en Formulación y gestión de proyectos en la Escuela Politécnica Nacional y en Investigación participativa en la Universidad de Wageningen. Extensionista del Ministerio de Agricultura y Ganadería (1983-1986), Investigador agropecuario del programa de Cacao y Café de la Estación Experimental Tropical Pichilingue del Instituto Nacional de Investigaciones Agropecuarias (1986-1998), Jefe de la División Técnica del Consejo Cafetalero Nacional (1998-2014), Asesor de la Unidad de Investigación de la Universidad Estatal del Sur de Manabí (2015) y Docente de la Carrera de Ingeniería Agrícola de la Escuela Superior Politécnica Agropecuaria de Manabí Manuel Félix López (2016...). Docente invitado en programas de maestría de la Universidad Nacional de Loja, Universidad Técnica Estatal de Quevedo, Universidad Técnica de Babahoyo y Escuela Superior Politécnica Agropecuaria de Manabí. Coordinador del Grupo de Investigación en Fitogenotecnia de la ESPAM MFL y Miembro de la Red Universitaria de Investigación y Desarrollo Cafetalero.

Registro SENESCYT: REG-INV 1601713

ORCID : <https://orcid.org/0000-0002-9326-8545>

Celular : 0983345824

Correo : [lduicela@gmail.com](mailto:lduicela@gmail.com)



**Dedicatoria:**

A mis hijos e hijas



## **Agradecimiento:**

Agradezco a las instituciones y autoridades que han apoyado mi trabajo investigativo principalmente al Instituto Nacional de Investigaciones Agropecuarias, al Ministerio de Agricultura y Ganadería, al Instituto Interamericano de Cooperación para la Agricultura, a la Escuela Superior Politécnica Agropecuaria de Manabí Manuel Félix López, a la Universidad Técnica de Manabí, a la Universidad Técnica Estatal de Quevedo y a la Universidad Técnica de Babahoyo. Además, agradezco a la Cooperación Alemana al Desarrollo y a las empresas Solubles Instantáneos C.A., Compañía de Elaborados de Café C.A. y Dublinsa S.A.

Mi agradecimiento imperecedero a la maestros y amigos, unos aportando a mi formación científica y otros acompañándome en las tareas investigativas, entre ellos a Gustavo Enríquez Calderón, Jaime Tola Cevallos, Ignacio Sotomayor Herrera, Freddy Amores Puyutaxi, Francisco Mite Vivar, Nelson Motato Alarcón, Juan Rodríguez Dillon, Pedro Ramírez Torres, Rey Gastón Loo Solorzano, Luis Fernando Plaza Avellán, Wiston Javier Morales, Mercedes Carranza Patiño, Emma Torres Navarrete, Adolfo Sánchez Laiño, Jenny Torres Navarrete, Willian Paúl Chilán Villafuerte, Santiago Jaramillo, Raúl Quijije Pinargote, Jorge Mendoza Mora, Cristina Maldonado Camposano, Juan Alberto Vera, Gabriel Suárez y Víctor Julio Esquivel Valverde. De manera particular, agradezco infinitamente al señor Jorge Salcedo Benítez, presidente de SICA, por el permanente apoyo a muchos de mis proyectos.

A las autoridades de la Escuela Superior Politécnica Agropecuaria de Manabí, Drs. Miryam Félix López, María Piedad Ormaza Murillo, Gabriel Antonio Navarrete Schettini, Ángel Monserrate Guzmán Cedeño, Leonardo Vera Macías y Mg. Sc. Sofía Velásquez Cedeño. A las autoridades de la Universidad Técnica de Babahoyo, Dr. Marcos Oviedo Rodríguez, Dra. María Bernarda Ruilova Cueva, Dra. Betty Mazacón Roca y M. Sc. Ramiro Galarza Bravo; así como a los Drs. Walter Oswaldo Reyes Borja, Luis Antonio Alcívar Torres, Walter Sandoya Villafuerte y Efraín Velasteguí López.

Muchas gracias al equipo de la editorial de la UTB, especialmente al Lcdo. Carlos Proaño, así como, a don Victorino Jaramillo Robles de Gráficas Isax.

Luis Alberto Duicela Guambi  
Autor





## PRESENTACIÓN

*“Un día llegará en que el pensamiento estadístico será tan necesario para ejercer la ciudadanía con eficiencia, como la capacidad de leer y escribir”* es la célebre frase del inglés Herbert George Wells, usada como argumento por uno de mis maestros, el Doctor Leonardo Corral Dávalos, para destacar la importancia de la estadística.

La sistematización de los apuntes sobre las “Herramientas estadísticas para la investigación agropecuaria” fue motivada por cuatro hechos: el haber participado en equipos de investigación en INIAP, GIZ, COFENAC, MAG-PROMSA, ESPAM MFL, UTEQ, UTM y UTB; el haber ejercido la docencia de la asignatura de estadística, en pregrado y posgrado; el haber sido consultor en planificación participativa de la investigación y el haber tenido como maestros de estadística a Leonardo Corral Dávalos (ESPOCH) y Gabino de Alba Flores (ITESM).

La investigación científica, apoyado en el análisis estadístico, usando técnicas paramétricas y no paramétricas, tiene la finalidad de contribuir a la generación de soluciones tecnológicas a los problemas del sector agropecuario. En este escenario, se requiere integrar los niveles investigativos: exploratorio, descriptivo, relacional, experimental, predictivo y aplicativo, en forma dinámica y ordenada, con enfoque sistémico. Las herramientas estadísticas desarrolladas en este documento, tratan de proporcionar una referencia de los procedimientos de análisis y de la toma de decisiones.

Las decisiones estadísticas, rigurosamente tienen que cumplir tres condiciones: Tener significación biológica expresada en la descripción del comportamiento de poblaciones y muestras en ambientes diferenciados, evidenciar la significación estadística determinada a través de las pruebas de hipótesis y proyectar significancia económica cuantificada a través de los beneficios y utilidad práctica.

Se reconoce que este documento tiene limitaciones de distinto tipo como la no inclusión de muchas pruebas estadísticas, la reducida argumentación de los resultados y la falta de la discusión y teorización. El benevolente aporte de los Grupos de investigación y de los investigadores, a quienes está dirigido el presente trabajo, contribuirá a superar las limitaciones conceptuales y metodológicas, para mejorar el proceso de construcción del conocimiento.

Calceta, septiembre del 2023

Luis Alberto Duicela Guambi, Ph. D.

Docente de Bioestadística de la Carrera de Ingeniería Agrícola  
Escuela Superior Politécnica Agropecuaria de Manabí Manuel Félix López



## PRÓLOGO

Pareciera lógico expresar que las habilidades en el área de la estadística, tanto descriptiva como inferencial, se adquieren con la experiencia. Los fundamentos de esta área del conocimiento se adaptan muy bien al enfoque de las aplicaciones que tendría la estadística. Durante mi experiencia docente de estadística, en Costa Rica y Panamá, he sido un convencido de que el mejor ambiente para el aprendizaje, y por ende del dominio de estas técnicas matemático-estadísticas, es la enseñanza tomando en consideración la utilidad práctica. Si bien el autor hace un esfuerzo por circunscribir la temática a la agricultura, lo cierto del caso es que la estadística, en todas sus modalidades, es una disciplina que se aplica a todas las áreas propias del quehacer humano. La vida misma, si se desea optimizar, se le puede dar un enfoque estadístico, dado que nuestras vivencias están saturadas de variables que siempre nos plantea su cuantificación para tomar las mejores decisiones. Lo anterior supone ver la vida misma desde un punto de vista de pensamiento estadístico.

El enfoque de la aplicación de la estadística muestra a los interesados en aprender y aplicar la estadística la razón por la cual se considera importante y emocionante a la vez. A lo largo del contenido temático del libro, el autor trata, no solo de seleccionar cuidadosamente los temas, sino también de incluir una serie de ejercicios y casos donde se analiza y toma decisiones.

Las *Herramientas estadísticas para la investigación Agropecuaria*, se empezó a gestar a raíz de la participación del autor, en diversos equipos de investigación, la práctica de la docencia, durante su formación profesional y el haber contado con excelentes profesores como el Dr. Gabino de Alba. Estos hechos contribuyeron a que el Dr. Luis Alberto Duicela Guambi generara una experiencia importante, que se sistematiza en el presente libro.

El enfoque que el autor le imprime al libro es de hecho una realidad; es decir, la obra pretende convertirse en una herramienta útil para buscar soluciones tecnológicas a las diferentes situaciones del sector agropecuario. Lo anterior a partir de cualquier enfoque que se le quiera dar a la investigación con base en el método científico. De lo anterior se desprende que la estadística nos aporta la estrategia para estudiar una población a partir de una muestra y desde sus inicios condiciona su tamaño a las características propias de la misma población, finita o infinita, así como al error proyectado y el nivel de confianza con que pretendemos estimar los parámetros poblacionales.

Finalmente considero que cualquier esfuerzo dirigido a buscar soluciones tecnológicas apropiadas para producir alimentos, como el que plantea el presente libro, debe ser bien visto

por cualquier usuario potencial. Hoy más que nunca las sociedades requieren producir alimentos no solo en cantidades adecuadas sino también inocuos y que su producción se lleve a cabo en armonía con el ambiente.

El presente libro nos presenta las herramientas necesarias para poder seleccionar objetivamente aquellas tecnologías, enmarcadas dentro de estos modernos modelos de producción alimentaria sostenible: económicamente rentable, ambientalmente amigable y socialmente justa.

Víctor Julio Esquivel Valverde, Ing. Agr. M. C.  
Colegio de Ingenieros agrónomos de Costa Rica  
Registro: 1813  
Correo: vjesquivel@gmail.com

## TABLA DE CONTENIDO

<b>CAPÍTULO I: INTRODUCCIÓN A LA INVESTIGACIÓN CIENTÍFICA</b> .....	<b>1</b>
1.1 AGRICULTURA Y ESTADÍSTICA .....	1
1.2 CONOCIMIENTO CIENTÍFICO.....	2
1.2.1 MÉTODO CIENTÍFICO .....	3
1.2.2 MÉTODOS AUXILIARES DEL MÉTODO CIENTÍFICO .....	5
1.2.3 NIVELES DE INVESTIGACIÓN .....	6
1.2.4 TIPOS DE INVESTIGACIÓN .....	8
1.2.4.1 POR LA INTERVENCIÓN DEL INVESTIGADOR .....	8
1.2.4.2 POR LA PLANIFICACIÓN DE LAS MEDICIONES .....	8
1.2.4.3 POR EL NÚMERO DE OBSERVACIONES EN EL TIEMPO .....	8
1.2.4.4 POR EL NÚMERO DE VARIABLES ANALÍTICAS .....	9
1.2.4.5 POR EL GRADO DE APLICABILIDAD .....	9
1.2.4.6 POR LA NATURALEZA DE LOS DATOS .....	9
1.2.4.7 POR EL GRADO DE MANIPULACIÓN DE LAS VARIABLES .....	9
1.2.4.8 POR LA PARTICIPACIÓN SOCIAL .....	10
1.2.4.9 POR LA CONTRIBUCIÓN A LA FORMACIÓN ACADÉMICA .....	10
1.3 PLANIFICACIÓN DE LA INVESTIGACIÓN .....	11
1.3.1 PROBLEMÁTICA DE INVESTIGACIÓN .....	11
1.3.1.1 CASO DE ANÁLISIS: SECTOR CAFETALERO ECUATORIANO .....	12
1.3.2 OBJETIVOS DE INVESTIGACIÓN .....	13
1.3.3 FORMULACIÓN DE LA PROPUESTA DE INVESTIGACIÓN .....	14
1.3.3.1 PROGRAMA DE INVESTIGACIÓN .....	14
1.3.3.2 PROYECTO DE INVESTIGACIÓN .....	15
1.3.3.3 ENSAYO DE INVESTIGACIÓN .....	20
1.3.4 ESTRUCTURA DE LA PROPUESTA DE INVESTIGACIÓN .....	21
<b>CAPÍTULO II: POBLACIONES Y MUESTRAS</b> .....	<b>23</b>
2.1 CLASIFICACIÓN DE LAS POBLACIONES .....	23
2.1.1 POBLACIÓN REAL Y POBLACIÓN ESTADÍSTICA .....	23
2.1.2 POBLACIÓN FINITA Y POBLACIÓN INFINITA .....	23
2.1.3 POBLACIÓN ACCESIBLE Y POBLACIÓN NO ACCESIBLE .....	23
2.1.4 POBLACIÓN ACTUAL Y POBLACIÓN FUTURA .....	24
2.2 PARÁMETROS Y ESTADÍSTICOS .....	24
2.3 TAMAÑO ÓPTIMO DE MUESTRA.....	26
2.3.1 TAMAÑO ÓPTIMO EN POBLACIONES INFINITAS .....	28
2.3.2 TAMAÑO ÓPTIMO EN POBLACIONES FINITAS .....	29
2.3.3 TAMAÑO ÓPTIMO DE MUESTRA PARA CONTEOS .....	30
2.4 MÉTODOS DE MUESTREO .....	31
2.4.1 MUESTREO NO PROBABILÍSTICO .....	31
2.4.2 MUESTREO PROBABILÍSTICO .....	31

2.4.2.1 MUESTREO ALEATORIO SIMPLE .....	32
2.4.2.2 MUESTREO SISTEMÁTICO .....	32
2.4.2.3 MUESTREO ESTRATIFICADO .....	33
2.4.2.4 MUESTREO POR CONGLOMERADOS .....	35
2.5 TÉCNICAS DE MUESTREO .....	36
2.6 VARIACIÓN .....	38
2.6.1 VARIABILIDAD BIOLÓGICA .....	38
2.6.2 VARIABILIDAD EN PROCESOS .....	38
2.6.3 ESCALAS DE MEDICIÓN .....	39
2.7 DATOS Y VARIABLES .....	40
2.7.1 TIPOS DE DATOS .....	41
2.7.2 TRANSFORMACIONES DE DATOS .....	42
2.7.2.1 TRANSFORMACIÓN DE DATOS ORDINALES A PORCENTAJES .....	42
2.7.2.2 TRANSFORMACIÓN DE DATOS A INCREMENTOS PORCENTUALES .....	43
2.7.2.3 TRANSFORMACIÓN A VALORES ESCALAR AL MÁXIMO .....	43
2.7.2.4 TRANSFORMACIONES A RANGOS .....	44
2.7.2.5 TRANSFORMACIONES MULTIPLICATIVAS Y DIVISORIAS .....	45
2.7.2.6 TRANSFORMACIONES LOGARÍTMICAS .....	46
2.7.2.7 TRANSFORMACIONES A RADICALES O POTENCIAS .....	47
2.7.2.8 TIPIFICACIÓN DE DATOS .....	47
2.7.3 CÁLCULO DE DATOS PERDIDOS .....	48
2.7.3.1 ESTIMACIÓN CON LA FÓRMULA DE YATES .....	49
2.7.3.2 CÁLCULO DE PARCELA PERDIDA CON LA MEDIA GEOMÉTRICA .....	50
2.8 VALIDACIÓN DE LA CALIDAD DE LOS DATOS .....	50
2.8.1 DIAGRAMA DE DISPERSIÓN .....	50
2.8.2 LÍMITES DE CONTROL .....	52
2.8.3 CONTROL DE DATOS ATÍPICOS CON LA RELACIÓN MEDIA Y DESVIACIÓN TÍPICA .....	53
2.8.4 DETECCIÓN DE DATOS ATÍPICOS CON EL RANGO INTERCUARTÍLICO .....	53
2.9 PRUEBAS ESTADÍSTICAS .....	54
2.9.1 PRUEBAS DE UNA COLA Y DE DOS COLAS .....	55
2.9.2 ERRORES EN LAS PRUEBAS ESTADÍSTICAS .....	55
2.9.3 POTENCIA DE LA PRUEBA ESTADÍSTICA .....	56
2.10 DISTRIBUCIONES DE PROBABILIDADES .....	57
2.10.1 DISTRIBUCIONES CONTINUAS .....	58
2.10.1.1 DISTRIBUCIÓN NORMAL .....	58
2.10.1.2 DISTRIBUCIÓN UNIFORME .....	62
2.10.2 DISTRIBUCIONES DISCRETAS .....	62
2.10.2.1 DISTRIBUCIÓN UNIFORME DISCRETA .....	62
2.10.2.2 DISTRIBUCIÓN BINOMIAL .....	63
2.10.2.3 DISTRIBUCIÓN GEOMÉTRICA .....	66
2.10.2.4 DISTRIBUCIÓN DE POISSON.....	67
<b>CAPÍTULO III: INVESTIGACIÓN EXPLORATORIA .....</b>	<b>68</b>
3.1. ELABORACIÓN DE INSTRUMENTOS.....	69

---

---

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

---

3.2.1	COEFICIENTE DE CONFIABILIDAD ALFA DE CRONBACH .....	71
3.2.2	COEFICIENTE DE CONFIABILIDAD DE RULÓN .....	72
3.3	DIAGNÓSTICO .....	73
3.3.1	PROBABILIDAD DE SUCESOS SEGÚN LA LEY DE LAPLACE .....	74
3.3.2	SENSIBILIDAD Y ESPECIFICIDAD .....	75
3.3.3	FRECUENCIAS DE HECHOS NOVEDOSOS .....	76
3.4	PRUEBA DE TAMIZAJE .....	79
3.4.1	PROBABILIDADES CONDICIONALES Y MARGINALES .....	80
3.4.2	PRUEBA EXACTA DE FISHER .....	81
3.5	CASOS DE INVESTIGACIÓN EXPLORATORIA .....	85
<b>CAPÍTULO IV: INVESTIGACIÓN DESCRIPTIVA .....</b>		<b>87</b>
4.1	ANÁLISIS DE FRECUENCIAS .....	87
4.2	DIAGRAMA DE “TALLOS Y HOJAS” .....	91
4.3	MEDIDAS DE TENDENCIA CENTRAL .....	93
4.3.1	MEDIA SIMPLE .....	93
4.3.2	MEDIA ARITMÉTICA .....	93
4.3.3	MEDIA GEOMÉTRICA.....	94
4.3.4	MEDIA ARMÓNICA .....	95
4.3.5	MEDIA PONDERADA .....	95
4.3.6	MEDIANA .....	96
4.3.7	MEDIANA EN DATOS AGRUPADOS .....	97
4.3.8	MODA .....	98
4.4	MEDIDAS DE DISPERSIÓN .....	98
4.4.1	ANÁLISIS DE CUANTILES .....	98
4.4.1.1	CUARTILES .....	99
4.4.1.2	QUINTILES .....	100
4.4.1.3	DECILES .....	101
4.4.1.4	PERCENTILES .....	101
4.4.2	DIAGRAMA DE “CAJA” .....	102
4.4.3	RANGO .....	103
4.4.4	VARIANZAS DE MUESTRAS Y DE LAS POBLACIONES .....	104
4.4.5	VARIANZA EN DATOS AGRUPADOS .....	105
4.4.6	DESVIACIÓN ESTÁNDAR .....	106
4.4.7	ERROR ESTÁNDAR .....	108
4.4.8	INTERVALO DE CONFIANZA DEL PARÁMETRO $\mu$ .....	109
4.4.9	COEFICIENTE DE VARIACIÓN DE LA MUESTRA .....	112
4.4.10	VARIACIÓN RELATIVA.....	113
4.4.11	PRECISIÓN .....	113
4.5	MEDIDAS DE CONCENTRACIÓN .....	114
4.5.1	ÍNDICE DE GINI .....	114
4.5.2	CURVA DE LORENZ.....	115
4.6	MEDIDAS DE FORMA DE LA DISTRIBUCIÓN .....	116
4.6.1	COEFICIENTE DE ASIMETRÍA .....	116



4.6.2	COEFICIENTE DE CURTOSIS.....	118
4.7	CONDICIONES PARA USAR LAS TÉCNICAS PARAMÉTRICAS .....	119
4.7.1	PRUEBAS DE LA NORMALIDAD .....	120
4.7.1.1	PRUEBA DE JARQUE Y BERA DE LA NORMALIDAD .....	120
4.7.1.2	PRUEBA DE KOLMOGOROV-SMIRNOV DE LA NORMALIDAD .....	122
4.7.2	PRUEBAS DE LA HOMOCEDASTICIDAD .....	124
4.7.2.1	PRUEBA F DE HOMOGENEIDAD DE LAS VARIANZAS .....	125
4.7.2.2	PRUEBA DE LEVENE DE HOMOGENEIDAD DE LAS VARIANZAS .....	126
4.7.3	ADITIVIDAD DE LOS EFECTOS .....	128
4.7.4	PRUEBA DE RACHAS DE ALEATORIEDAD .....	128
4.7.5	PRUEBA DE LOS SIGNOS DE ALEATORIEDAD .....	130
4.8	PRUEBAS DE LA INDEPENDENCIA .....	131
4.8.1	PRUEBA DE LA INDEPENDENCIA CON LA <i>t</i> de WELCH .....	131
4.8.2	PRUEBA U DE MANN-WHITNEY .....	133
4.8.3	PRUEBA DE LA INDEPENDENCIA CON CHI CUADRADO .....	135
4.9	INTRODUCCIÓN AL ANÁLISIS MULTIVARIADO .....	138
4.9.1	ANÁLISIS DE COMPONENTES PRINCIPALES .....	139
4.9.2	ANÁLISIS FACTORIAL .....	141
4.9.3	ANÁLISIS DE CORRESPONDENCIAS .....	144
4.9.4	ESCALAMIENTO MULTIDIMENSIONAL .....	149
4.9.5	ANÁLISIS DE CLÚSTER .....	151
<b>CAPÍTULO V:</b>	<b>INVESTIGACIÓN RELACIONAL.....</b>	<b>155</b>
5.1	ANÁLISIS DE FRECUENCIAS EN DOS SENTIDOS .....	157
5.2	COEFICIENTE DE CORRELACIÓN .....	158
5.2.1	COEFICIENTE DE CORRELACIÓN DE PEARSON .....	159
5.2.2	COEFICIENTE DE CORRELACIÓN POR RANGOS DE SPEARMAN .....	161
5.3	PRUEBAS DE CONCORDANCIA .....	163
5.3.1	PRUEBA CUADRADA .....	163
5.3.2	COEFICIENTE DE CONTINGENCIA DE PEARSON .....	167
5.3.3	COEFICIENTE DE CONCORDANCIA DE KAPPA DE COHEN .....	168
5.3.4	PRUEBA CHI CUADRADA PARA PROPORCIONES .....	170
5.3.5	COMPARACIÓN DE ATRIBUTOS CON CHI CUADRADA .....	171
<b>CAPÍTULO VI:</b>	<b>INVESTIGACIÓN EXPERIMENTAL.....</b>	<b>171</b>
6.1	ORGANIZACIÓN DEL EXPERIMENTO .....	173
6.1.1	FACTORES Y NIVELES .....	173
6.1.2	UNIDAD EXPERIMENTAL .....	175
6.1.3	DISEÑOS BÁSICOS .....	176
6.1.4	REGISTRO DE VARIABLES EXPERIMENTALES .....	177
6.1.5	ANÁLISIS ESTADÍSTICO .....	177
6.2	ANÁLISIS DE VARIANZA PARAMÉTRICO .....	178
6.2.1	EXPERIMENTO EN DISEÑO COMPLETAMENTE AL AZAR .....	180
6.2.1.1	EXPERIMENTO DCA DESBALANCEADO .....	181
6.2.1.2	EXPERIMENTO DCA BALANCEADO .....	183

6.2.2 EXPERIMENTO CON DISEÑO EN BLOQUES AL AZAR .....	186
6.2.3 EXPERIMENTO CON DISEÑO EN CUADRADO LATINO .....	188
6.2.4 EXPERIMENTO CON DISEÑO EN LÁTICE .....	192
6.2.5 EXPERIMENTO CON DISEÑO IRRESTRICTO AL AZAR .....	197
6.2.6 EXPERIMENTOS FACTORIALES .....	200
6.2.6.1 EXPERIMENTO BIFACTORIAL .....	201
6.2.6.2 EXPERIMENTO TRIFACTORIAL.....	203
6.2.6.3 EXPERIMENTO EN PARCELA DIVIDIDA.....	208
6.2.7 EXPERIMENTOS COMPLEJOS .....	211
6.2.7.1. EXPERIMENTO AB + 1 .....	211
6.2.7.2 EXPERIMENTO AB + 1 + 1 .....	212
6.3 DISEÑOS FACTORIALES INCOMPLETOS .....	216
6.3.1 EXPERIMENTO EN DISEÑO “SAN CRISTÓBAL” .....	216
6.3.2 EXPERIMENTOS MULTIFACTORIALES CON EL MÉTODO TAGUCHI.....	221
6.3.2.1 DISEÑO DE TRES FACTORES CON DOS NIVELES .....	222
6.3.2.2 EXPERIMENTO DE SIETE FACTORES EN DOS NIVELES .....	225
6.3.2.3 EXPERIMENTO DE CUATRO FACTORES EN TRES NIVELES .....	230
6.3.3 ANÁLISIS DE VARIANZA EN LOS DISEÑOS TAGUCHI .....	232
6.4 ANÁLISIS DE COVARIANZA .....	235
6.4.1 COVARIANZA EN LA INVESTIGACIÓN PECUARIA .....	237
6.4.2 COVARIANZA EN LA INVESTIGACIÓN AGRONÓMICA .....	242
6.5 COMPARACIÓN DE MEDIAS .....	245
6.5.1 USO DE LA PRUEBA “t DE STUDENT” DE UNA MUESTRA .....	246
6.5.2 USO DE t PARA SEPARAR MEDIAS INDEPENDIENTES .....	247
6.5.3 COMPARACIÓN DE DOS MEDIAS RELACIONADAS .....	249
6.5.4 DIFERENCIA MÍNIMA SIGNIFICATIVA .....	251
6.5.5 PRUEBA DE DUNCAN .....	255
6.5.6 PRUEBA DE TUKEY .....	257
6.5.7 PRUEBA DE NEWMAN-KEULS .....	259
6.5.8 COMPARACIONES ORTOGONALES .....	261
6.6 ANÁLISIS DE VARIANZA NO PARAMÉTRICO .....	266
6.6.1 ANÁLISIS DE VARIANZA POR RANGOS DE KRUSKAL-WALLIS .....	266
6.6.2 ANÁLISIS DE VARIANZA POR RANGOS DE FRIEDMAN .....	273
<b>CAPÍTULO VII: INVESTIGACIÓN PREDICTIVA .....</b>	<b>281</b>
7.1 INTERVALOS DE CONFIANZA DEL PARÁMETRO $\mu$ .....	281
7.2 MEDIAS MÓVILES .....	283
7.3 MODELACIÓN MATEMÁTICA .....	285
7.3.1 MODELO LINEAL.....	288
7.3.2 MODELO POLINÓMICO DE GRADO 2 .....	292
7.3.2.1 MAXIMIZACIÓN Y OPTIMIZACIÓN .....	293
7.3.3 MODELO POLINÓMICO DE GRADO 3 .....	297
7.3.4 MODELO MULTIVARIADO.....	299

<b>CAPÍTULO VIII: INVESTIGACIÓN APLICATIVA.....</b>	<b>303</b>
8.1 VALIDACIÓN DE TECNOLOGÍAS .....	303
8.2 INTRODUCCIÓN A LA PROGRAMACIÓN LINEAL .....	305
8.3 INVESTIGACIÓN DEL AMBIENTE .....	307
8.3.1 USO DE GRÁFICAS RADIALES EN ANÁLISIS AMBIENTAL.....	308
8.3.2 EVALUACIÓN DE IMPACTO AMBIENTAL .....	309
8.4 ANÁLISIS DE LA SOSTENIBILIDAD DE LAS FINCAS .....	312
8.4.1 DIMENSIONES, INDICADORES Y SUBINDICADORES .....	313
8.4.1.1 DIMENSIÓN ECONÓMICA (E) .....	313
8.4.1.2 DIMENSIÓN AMBIENTAL (A) .....	313
8.4.1.3 DIMENSIÓN SOCIO CULTURAL (S) .....	313
8.4.2 MEDICIÓN DE LA SOSTENIBILIDAD .....	314
8.4.3 INDICADORES DE LA SOSTENIBILIDAD .....	314
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>319</b>
<b>ANEXOS .....</b>	<b>325</b>
Anexo 1: Glosario estadístico .....	327
Anexo 2: Valores críticos del Coeficiente de correlación de Spearman.....	331
Anexo 3: Lectura de valores críticos p de Z en la distribución normal .....	332
Anexo 4: Lectura de valores críticos Chi cuadrada para varios niveles de significación .....	333
Anexo 5: Lectura de los valores críticos t de Student para distintos grados de libertad y niveles de significación .....	334
Anexo 6: Lectura de los valores críticos de F con $\alpha = 0,05$ .....	335
Anexo 7: Lectura de Amplitud Estudentizada de Tukey (Q) con $\alpha = 0,05$ y $0,01$ .....	336

ÍNDICE DE GRÁFICOS

Gráfico 1. Población finita $N = 20$ y los parámetros media y desviación estándar .....	25
Gráfico 2. Dos muestras aleatorias de $n = 4$ y los principales estadísticos .....	25
Gráfico 3. Funciones de distribución y de densidad para datos cuantitativos tipificados .....	48
Gráfico 4. Dispersograma de la altura de planta de 32 clones de café robusta .....	51
Gráfico 5. Dispersograma de residuos para identificar “puntos atípicos” .....	52
Gráfico 6. Monitoreo de defectos físicos del grano en una planta de semillas .....	52
Gráfico 7. Identificación de datos atípicos basado en los cuartiles y el rango intercuartílico .....	54
Gráfico 8. Hipótesis, pruebas, zona de rechazo de hipótesis nula y distribución normal .....	55
Gráfico 9. Propiedades de la distribución normal estandarizada .....	59
Gráfico 10. Contenidos máximos y mínimos de azufre en suelos de 110 fincas del cantón Loreto .....	78
Gráfico 11. Caracterización química de los suelos de la Amazonía norte del Ecuador .....	79
Gráfico 12. Probabilidades condicionales y marginales de dos sucesos fitosanitarios .....	81
Gráfico 13. Frecuencias absolutas de altura de planta en 51 cafetos de 60 meses de edad .....	90
Gráfico 14. Diagrama de “tallos y hojas” para datos de producción en una muestra de 30 plantas .....	92
Gráfico 15. Clasificación de los ingresos anuales de 829 familias de agricultores .....	99
Gráfico 16. Diagrama de caja para la serie de datos de producción g.planta-1 .....	102
Gráfico 17. Curva normal o campana gaussiana .....	107
Gráfico 18. Relación entre la media muestral y su error típico .....	108
Gráfico 19. Medias $\pm$ Errores típicos de tres clones de café robusta .....	109
Gráfico 20. Intervalos de confianza de $\mu$ (95 %) para ganancia de peso en dos grupos cerdos .....	111
Gráfico 21. Curva de Lorenz de la distribución salarial en una empresa agrícola .....	116
Gráfico 22. Curvas asimétricas que distorsionan la campana gaussiana.....	116
Gráfico 23. Curvas de las series de datos según los tipos de curtosis .....	118
Gráfico 24. Dispersograma de los datos de producción de cuatro clones de café robusta.....	124
Gráfico 25. Análisis de componentes principales de los parámetros productivos en 12 fincas ganaderas .....	141
Gráfico 26. Reducción de variables en un estudio ganadero mediante análisis factorial .....	143
Gráfico 27. Análisis de correspondencias de las variables asociadas a la producción ganadera .....	145
Gráfico 28. Análisis de correspondencias de los atributos de la calidad de peces de la especie “vieja azul ...	147
Gráfico 29. Análisis de correspondencias entre pH, materia orgánica y nutrimentos en suelos amazónicos ..	149
Gráfico 30. Calidad de taza de los cafés robustas en función de los métodos de beneficio .....	154
Gráfico 31. Similitud en calidad de taza en función de los genotipos de café robusta .....	154
Gráfico 32. Similitud en taza de los cafés robustas de Orellana y Sucumbíos, según métodos de beneficio ....	155
Gráfico 33. Diagramas de los tipos de correlaciones entre variables y su aproximación al coeficiente $r$ .....	159
Gráfico 34. Comportamiento de una variable de respuesta como efecto de tres tratamientos .....	117
Gráfico 35. Relaciones entre niveles de estudio por los factores nitrógeno y fósforo .....	219
Gráfico 36. Superficie de respuesta a los niveles de nitrógeno y fósforo en la producción .....	220
Gráfico 37. Modelo polinómico de grado 2 de la producción en función del nitrógeno .....	220
Gráfico 38. Modelo polinómico de grado 2, de la producción en función del fósforo .....	221
Gráfico 39. Diseños ortogonales sugeridos para la investigación agropecuaria y agroindustrial .....	221

Gráfico 40. Densidad del substrato para el cultivo de hongos comestibles en función de tres factores y dos niveles de estudio .....	225
Gráfico 41. Calidad de taza del café soluble en función de siete factores con dos niveles en estudio.....	227
Gráfico 42. Porcentaje de almendras bien fermentadas en función de los factores de la poscosecha .....	229
Gráfico 43. Rendimiento de pulpa de celulosa en función de cuatro factores con tres niveles en estudio .....	232
Gráfico 44. Regresión entre pesos inicial y final de cuatro camadas de cerdos.....	238
Gráfico 45. Cambios observados entre los pesos inicial y final de cuatro camadas de cerdos tratados con cinco dietas .....	239
Gráfico 46. Diferencias entre pesos finales observados y pesos finales corregidos por covarianza .....	241
Gráfico 47. Modelos lineales de la variable producción de grano en función de las plantas.parcela <sup>-1</sup> .....	244
Gráfico 48. Diagrama de la separación de medias usando la DMS en experimentos desbalanceados .....	253
Gráfico 49. Diagrama de la separación de medias de genotipos de sorgo con la DMS .....	254
Gráfico 50. Distribuciones de las series de datos en dos tratamientos comparados .....	266
Gráfico 51. Dispersión de los datos y posición de las medianas de la variable peso al nacimiento de cerdos en camadas de distinto tamaño .....	272
Gráfico 52. Relación entre media $\pm$ error típico y los límites de confianza del parámetro $\mu$ .....	282
Gráfico 53. Modelos lineal y cuadrático que se ajustan a la función insumo producto .....	288
Gráfico 54. Producción neta de pasto en función de las dosis de bioestimulante .....	290
Gráfico 55. Producción de azúcar blanca en función de la materia prima .....	290
Gráfico 56. Incertidumbre de las respuestas posibles en la zona inexploradas .....	292
Gráfico 57. Relación entre la función producción y la función económica .....	293
Gráfico 58. La producción de maíz en función de la fertilización nitrogenada .....	295
Gráfico 59. Modelo característico del comportamiento de una plaga en función del plaguicida .....	299
Gráfico 60. Maximización de beneficios usando máquinas A y B para producir M y N .....	307
Gráfico 61. Análisis de variables ambientales en cuatro comunidades rurales en un momento 1.....	309

ÍNDICE DE FIGURAS

<i>Figura 1. Características fundamentales del conocimiento científico .....</i>	2
<i>Figura 2. Los seis pasos del método científico para construir conocimiento con su argumentación.....</i>	3
<i>Figura 3. Niveles de investigación científica y su relación con los objetivos de estudio .....</i>	6
<i>Figura 4. Causas y consecuencias de la crisis del sector agropecuario en el Ecuador .....</i>	11
<i>Figura 5. Árbol de problemas del sector cafetalero ecuatoriano .....</i>	12
<i>Figura 6. Árbol de objetivos del sector cafetalero ecuatoriano .....</i>	13
<i>Figura 7. Diagrama causa - efecto de los factores determinantes de la producción agrícola .....</i>	16
<i>Figura 8. Árbol de problemas del cultivo de cacao Nacional en la provincia de Manabí .....</i>	17
<i>Figura 9. Identificación de objetivos de un programa de investigación en cacao Nacional en Manabí .....</i>	18
<i>Figura 10. Construcción de la idea del proyecto de investigación alineado a un programa .....</i>	19
<i>Figura 11. Relaciones entre título, objetivos, hipótesis y resultados esperados del proyecto .....</i>	20
<i>Figura 12. Estructura referencial de una propuesta de investigación agropecuaria .....</i>	22
<i>Figura 13. Diagrama de un muestreo aleatorio simple .....</i>	32
<i>Figura 14. Muestreo sistemático en un campo experimental uniforme .....</i>	33
<i>Figura 15. Población humana y subpoblaciones (estratos) por género .....</i>	34
<i>Figura 16. Estratificación de un terreno cultivado en base al grado de pendiente.....</i>	35
<i>Figura 17. Muestreo por conglomerados en seis comunidades rurales de un territorio .....</i>	36
<i>Figura 18. Distintas formas de dispersión de las poblaciones de organismos vivos .....</i>	36
<i>Figura 19. Mapeo de la cobertura arbórea en sistemas cafetaleros .....</i>	37
<i>Figura 20. Técnicas de muestreo usadas en la investigación agrícola .....</i>	37
<i>Figura 21. Mediciones de variables de interés en poblaciones finitas .....</i>	40
<i>Figura 22. Escala Likert aplicada para valorar la importancia del café en el desayuno .....</i>	40
<i>Figura 23. Tipos de datos cuantitativos y cualitativos .....</i>	41
<i>Figura 24. Identificación de los árboles de interés genético a nivel de campo .....</i>	77
<i>Figura 25. Estudios a lo largo del tiempo y en diversos escenarios .....</i>	87
<i>Figura 26. Técnicas de análisis multivariados de mayor uso en la investigación agropecuaria .....</i>	138
<i>Figura 27. Frecuencias de semillas amarillas y verdes en F2 derivadas del cruce monohíbrido .....</i>	164
<i>Figura 28. Frecuencias observadas en F2 de semillas derivadas de un cruce dihíbrido .....</i>	166
<i>Figura 29. Distribución de las unidades experimentales y algunas formas de la parcela útil .....</i>	175
<i>Figura 30. Elementos de muestreo y generación de datos .....</i>	175
<i>Figura 31. Distribución de unidades experimentales en un diseño de Bloques al azar .....</i>	186
<i>Figura 32. Diagrama de un experimento en cuadrado latino 4 x 4 .....</i>	189
<i>Figura 33. Diagrama de un experimento en parcelas divididas de dos factores por tres niveles .....</i>	209
<i>Figura 34. Diseño San Cristóbal para ensayo de dos factores en cuatro niveles .....</i>	217
<i>Figura 35. Impacto de un proyecto agroecológico sobre la sobrevivencia de lombrices en el suelo.....</i>	310
<i>Figura 36. Triángulo de la sostenibilidad a partir de mediciones en fincas .....</i>	317

ÍNDICE DE CUADROS

Cuadro 1. Programa de investigación “Mejora de la productividad del café” .....	15
Cuadro 2. Análisis de frecuencias para la priorización de los factores causales de la baja productividad....	17
Cuadro 3. Fórmulas más usadas para el cálculo de parámetros y estadígrafos .....	26
Cuadro 4. Valores críticos de Z asociados a los niveles de significación más usados en bioestadística .....	27
Cuadro 5. Cálculo de las proporciones de las submuestras en un muestreo estratificado .....	35
Cuadro 6. Tipos de dato, escalas de medición y tipos de variable, relaciones y ejemplos .....	39
Cuadro 7. Áreas del conocimiento, elemento de estudio, variables, escala de medición, tipos de datos y unidad de medida en estudios agrosociales y ambientales .....	41
Cuadro 8. Transformación de datos ordinales a porcentajes .....	42
Cuadro 9. Transformación de datos cuantitativos a porcentaje en función del testigo .....	43
Cuadro 10. Transformación de datos cuantitativos a valor Escalar al Máximo .....	44
Cuadro 11. Transformación de datos originales a rangos, con y sin valores repetidos .....	45
Cuadro 12. Transformación de datos originales con factores multiplicativos y divisorios .....	45
Cuadro 13. Transformación de datos cuantitativos a valores logarítmicos y con notación científica .....	46
Cuadro 14. Transformación de datos porcentuales a radicales y potencias .....	47
Cuadro 15. Transformación de datos a los valores Z y cálculo de las funciones densidad y distribución .....	48
Cuadro 16. Datos de un ensayo de cinco tratamientos y cuatro repeticiones con una “parcela perdida” .....	49
Cuadro 17. Tipos de error estadístico en las pruebas de hipótesis .....	56
Cuadro 18. Datos de una encuesta de seis ítems a productores de tomate, usando la escala ordinal de 1 a 5, referida al uso de insecticidas y cálculos básicos .....	70
Cuadro 19. Matriz de resultados de una encuesta de seis variables con respuestas dicotómicas .....	72
Cuadro 20. Resultados del diagnóstico en el laboratorio contrastado con el diagnóstico visual en campo ...	76
Cuadro 21. Matriz de puntajes de atributos de interés económico en árboles de naranja .....	79
Cuadro 22. Prueba de tamizaje de agroquímicos midiendo la germinación de una especie sensible en substratos contaminados .....	80
Cuadro 23. Modelo de una tabla de contingencia 2 x 2 .....	82
Cuadro 24. Calidad de los experimentos relacionados con la selección del diseño .....	82
Cuadro 25. Estado de preñez de las novillas en función de los métodos de inseminación .....	83
Cuadro 26. Estado de salud de los trabajadores agrícolas clasificados por edades .....	84
Cuadro 27. Análisis de frecuencias de la variable producción en una muestra de 30 plantas .....	89
Cuadro 28. Análisis de frecuencias de altura (AP) en una muestra de 51 plantas de café .....	90
Cuadro 29. Frecuencias absolutas (FA) y relativas (FR) de los promedios de altura de planta en 32 clones de café robusta y clasificación cualitativa según el porte .....	91
Cuadro 30. Comparación de las medias de altura de planta de tres clones de café robusta .....	92
Cuadro 31. Análisis de frecuencias de la variable producción g.planta-1 en una muestra de 30 plantas .....	97
Cuadro 32. Análisis de cuartiles de los ingresos anuales de 829 hogares de caficultores .....	100
Cuadro 33. Análisis de quintiles de los ingresos anuales en una muestra de 829 hogares de caficultores.....	100
Cuadro 34. Análisis de deciles de los ingresos anuales en 829 hogares de agricultores .....	101
Cuadro 35. Datos unitarios de peso de aguacates y cálculo de los cuadrados .....	105



---

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

---

Cuadro 36. Cálculo de la varianza del peso de los aguacates .....	105
Cuadro 37. Clases, marca de clase ( $M$ ), frecuencia absoluta ( $FA$ ) y cálculos de la media ponderada y varianza .....	106
Cuadro 38. Fórmulas para calcular los límites de confianza de $\mu$ en series de datos numéricos y categóricos ...	110
Cuadro 39. Salarios mensuales de los trabajadores de una empresa agrícola .....	115
Cuadro 40. Frecuencias de las remuneraciones de 10 trabajadores en una empresa agrícola y procedimiento de cálculo del índice de Gini .....	115
Cuadro 41. Cálculo del Coeficiente de asimetría ( $A$ ) de la variable peso del fruto .....	117
Cuadro 42. Cálculo del Coeficiente de curtosis para peso en una muestra de 9 frutos .....	119
Cuadro 43. Prueba de la normalidad de cuatro clones de café robusta con el método de Jarque-Bera .....	121
Cuadro 44. Análisis de frecuencias absolutas de los clones de café robusta en estudio .....	123
Cuadro 45. Prueba KS de la normalidad de los datos de rendimiento del clon LE-A1 de café robusta .....	123
Cuadro 46. Prueba $F$ de la homocedasticidad de cuatro clones de café robusta .....	125
Cuadro 47. “Levene-data” ( $Y$ ) de la producción de tres clones experimentales de café robusta .....	127
Cuadro 48. Análisis de varianza de las diferencias entre medias y medianas o data de Levene .....	127
Cuadro 49. Croquis del huerto con las dos especies frutales .....	129
Cuadro 50. Ventas en las tiendas A y B, diferencias, signos y rachas de los signos .....	130
Cuadro 51. Prueba $t$ de Welch de la independencia entre dos clones de café robusta .....	132
Cuadro 52. Frutos por planta y transformación a rangos de dos tratamientos con bioestimulantes .....	134
Cuadro 53. Modelo de tabla de contingencia $2 \times 2$ de muestras relacionadas .....	135
Cuadro 54. Posición de las hojas infestadas en las plantas y valores observados y esperados .....	135
Cuadro 55. Cálculo de los valores Chi-cuadrado para posición de las hojas infestadas .....	136
Cuadro 56. Triunfos observados en las carreras de caballos según carriles en la pista circular .....	137
Cuadro 57. Triunfos observados y esperados de 144 carreras de caballos y cálculo de Chi cuadrada ....	137
Cuadro 58. Variables productivas en 12 fincas ganaderas .....	139
Cuadro 59. Valores propios y proporciones de la contribución a la variación total de las variables productivas asociadas a la ganadería y magnitud de los autovectores en los componentes principales 1 y 2 .....	140
Cuadro 60. Magnitud de los autovectores en los componentes principales 1 y 2 .....	141
Cuadro 61. Coeficientes de correlaciones entre variables productivas en fincas ganaderas .....	142
Cuadro 62. Análisis de las comunalidades entre variables productivas en las fincas ganaderas .....	142
Cuadro 63. Análisis de componentes principales en el AF confirmatorio .....	143
Cuadro 64. Variables productivas de 12 fincas ganaderas .....	144
Cuadro 65. Frecuencias absolutas ( $FA$ ) de los atributos de calidad en faenamiento de peces de la especie “vieja azul” en función a cuatro dietas .....	146
Cuadro 66. Promedios de los contenidos de nutrientes, materia orgánica y pH del suelo de 150 fincas cafetaleras de la amazonia norte .....	147
Cuadro 67. Contribución a la variación explicada mediante Chi cuadrado .....	148
Cuadro 68. Valores de peso de las variables en las coordenadas .....	148
Cuadro 69. Mortalidad, eficacia y rangos de eficacia en función del uso de agroquímicos .....	150
Cuadro 70. Distancias entre los niveles de eficacia de los productos .....	151
Cuadro 71. Escalamiento multidimensional de la eficacia de agroquímicos usando la distancia euclídea ...	151



---

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

---

Cuadro 72. Calidad de taza de genotipos de café robusta, en función de cuatro métodos de beneficio ....	153
Cuadro 73. Análisis de frecuencias absolutas (FA), relativas (FR) y acumuladas (FAA y FRA) de las variables altura de planta y producción, en una muestra aleatoria de 40 plantas de café robusta ..	158
Cuadro 74. Altura y número de pares de ramas en cafetos en una muestra aleatoria de nueve plantas .....	160
Cuadro 75. Datos de las variables altura de planta y número de ramas, cuadrados y productos .....	160
Cuadro 76. Transformación a rangos de los datos altura de planta y pares de ramas y diferencias de rangos ...	162
Cuadro 77. Frecuencias observadas y esperadas de plantas con semillas clasificadas por forma y color y cálculo de $\chi^2$ .....	166
Cuadro 78. Evaluación a 50 aspirantes a becas estudiantiles por dos evaluadores .....	169
Cuadro 79. Frecuencias observadas y esperadas de insectos vivos y muertos en función de las dosis de insecticida y cálculo de Chi cuadrada .....	170
Cuadro 80. Cualidades de los candidatos percibidas por la ciudadanía en una encuesta .....	172
Cuadro 81. Factores y niveles para cultivares (A) y densidades poblacionales (B) .....	174
Cuadro 82. Diseño de los tratamientos en dos experimentos unifactoriales .....	174
Cuadro 83. Diseño de los tratamientos en un ensayo bifactorial .....	174
Cuadro 84. Diseños experimentales básicos y sus principales características .....	176
Cuadro 85. Referencias para diseñar experimentos totalmente al azar .....	180
Cuadro 86. Larvas de broca del café por recipiente en función de cuatro dietas artificiales .....	181
Cuadro 87. Resultado del ADEVA en un experimento en DCA desbalanceado .....	182
Cuadro 88. Peso de los racimos en función de los tratamientos de frecuencias de riego .....	183
Cuadro 89. ADEVA del ensayo de frecuencias de riego .....	184
Cuadro 90. Alternativas de alimentación de las ovejas y ganancias de peso .....	184
Cuadro 91. Resultados del ADEVA en ensayo DCA balanceado .....	185
Cuadro 92. Producción por parcela en 4 variedades de sorgo en experimento en bloques al azar .....	187
Cuadro 93. ADEVA del experimento de variedades de sorgo forrajero .....	188
Cuadro 94. Distribución de las jaulas y tratamientos en cuatro períodos de estudio .....	190
Cuadro 95. Pesos de los huevos y cálculo de sumas y medias de columnas (jaulas) e hileras (períodos) ....	190
Cuadro 96. Datos de los pesos de los huevos y cálculo de sumas y promedios de dietas y jaulas .....	191
Cuadro 97. ADEVA del diseño en cuadrado latino .....	192
Cuadro 98. Promedios de rendimiento (kg.parcela-1) de 25 variedades de maíz en diseño LT 5 x 5 .....	194
Cuadro 99. Promedio de rendimiento de las variedades por repeticiones; suma, promedio y varianza, valor escalar al máximo (EAM) y priorización de selección .....	195
Cuadro 100. Promedios de rendimiento de grano los bloques incompletos alojados en las repeticiones .....	195
Cuadro 101. ADEVA del experimento en Látice de 25 variedades .....	196
Cuadro 102. Promedios de magnesio en hojas de fréjol en función de tres tratamientos de fertilización edáfica .....	198
Cuadro 103. Medias de Mg en follaje de fréjol como respuesta a tres alternativas de fertilización edáfica ..	199
Cuadro 104. Datos de producción del experimento bifactorial VF .....	201
Cuadro 105. Análisis de la varianza del experimento .....	203
Cuadro 106. Producción de forraje de cuatro genotipos de sorgo en dos ciclos de cultivo y aplicación de bioestimulante .....	204

---

---

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

---

<i>Cuadro 107. Sumas y medias de genotipos y bloques del experimento .....</i>	205
<i>Cuadro 108. Sumas y medias de genotipos y bioestimulantes .....</i>	205
<i>Cuadro 109. Sumas y medias de genotipos y ciclos de cultivo .....</i>	205
<i>Cuadro 110. Sumas y medias de bioestimulantes y ciclos de cultivo .....</i>	206
<i>Cuadro 111. Sumas y medias de genotipos, ciclos de cultivo y bioestimulantes .....</i>	206
<i>Cuadro 112. Análisis de la varianza del ensayo de cuatro genotipos de sorgo forrajero en dos ciclos de cultivo y aplicación de bioestimulantes .....</i>	207
<i>Cuadro 113. Promedios de producción de cacao de los tratamientos desglosados por factores y niveles ....</i>	209
<i>Cuadro 114. Sumas y promedios del rendimiento en función de las densidades y frecuencias de riego .....</i>	210
<i>Cuadro 115. Sumas y promedios de las parcelas grandes y de las repeticiones .....</i>	210
<i>Cuadro 116. Análisis de la varianza del ensayo en parcela dividida .....</i>	211
<i>Cuadro 117. Promedios de producción por parcela del experimento AB + 1.....</i>	213
<i>Cuadro 118. Análisis de las interacciones V x F sobre la producción de maíz (kg.parcela<sup>-1</sup>) .....</i>	214
<i>Cuadro 119. Análisis de varianza del experimento AB + 1 .....</i>	214
<i>Cuadro 120. Promedios de producción de experimento VF + 1 + 1 .....</i>	215
<i>Cuadro 121. Cálculo de las medias de la interacción V.F y testigos .....</i>	215
<i>Cuadro 122. Análisis de varianza del experimento VF + 1 + 1 .....</i>	216
<i>Cuadro 123. Factores y niveles en estudio del tipo (4)<sup>2</sup> .....</i>	217
<i>Cuadro 124. Producción de un experimento en diseño San Cristóbal de 7 tratamientos con cuatro repeticiones .....</i>	218
<i>Cuadro 125. Efectos principales de los factores N y P en los distintos niveles en estudio .....</i>	218
<i>Cuadro 126. Interacciones entre los factores N y P (Respuesta en g.planta<sup>-1</sup>) .....</i>	218
<i>Cuadro 127. Diseño factorial 2<sup>3</sup> clásico .....</i>	222
<i>Cuadro 128. Modelo de diseño ortogonal L<sub>4</sub>(2)<sup>3</sup> .....</i>	223
<i>Cuadro 129. Factores y niveles, según diseño L<sub>4</sub>(2)<sup>3</sup> .....</i>	223
<i>Cuadro 130. Diseño ortogonal L<sub>4</sub>(2)<sup>3</sup> .....</i>	223
<i>Cuadro 131. Promedios de la variable de respuesta “densidad sustrato” (kg.m<sup>3</sup>) en función del diseño L<sub>4</sub>(2)<sup>3</sup> .....</i>	224
<i>Cuadro 132. Respuesta de “densidad del sustrato” (kg.m<sup>3</sup>) en función del diseño L<sub>4</sub>(2)<sup>3</sup> .....</i>	224
<i>Cuadro 133. Diseño ortogonal L<sub>8</sub>(2)<sup>7</sup> .....</i>	226
<i>Cuadro 134. Factores y niveles en estudio en un diseño L<sub>8</sub>(2)<sup>7</sup> .....</i>	226
<i>Cuadro 135. Calidad sensorial del café solubles en función del diseño L<sub>8</sub>(2)<sup>7</sup> .....</i>	227
<i>Cuadro 136. Efecto de los factores y niveles en estudio sobre la calidad de taza .....</i>	227
<i>Cuadro 137. Factores y niveles en estudio de un experimento L<sub>8</sub>(2)<sup>7</sup> .....</i>	228
<i>Cuadro 138. Almendras bien fermentadas de cacao (%) como respuesta al experimento L<sub>8</sub>(2)<sup>7</sup>.....</i>	229
<i>Cuadro 139. Almendras de cacao fermentadas (%) en función de los factores/niveles en estudio .....</i>	229
<i>Cuadro 140. Diseño ortogonal L<sub>8</sub>(3)<sup>4</sup> de cuatro factores en tres niveles .....</i>	230
<i>Cuadro 141. Factores y niveles en estudio del diseño L<sub>8</sub>(3)<sup>4</sup> para obtener pulpa de celulosa a partir de Agave .....</i>	231
<i>Cuadro 142. Rendimiento de pulpa de celulosa para papel en función de los factores agua (A), tiempo (B) y concentraciones de ácidos acético (C) y clorhídrico (D) en tres niveles.....</i>	231

---

---

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

---

Cuadro 143. Extracción de pulpa de celulosa en función de cuatro factores en tres niveles de estudio .....	231
Cuadro 144. Extracción de pulpa de celulosa de Agave en función de los factores agua (A), tiempo (B), concentración de ácido acético (C) y concentración de ácido clorhídrico (D).....	233
Cuadro 145. Estadígrafos de la variable extracción de pulpa de celulosa a partir de Agave .....	234
Cuadro 146. Análisis de la varianza de la característica “Extracción de pulpa de celulosa” en Agave mediante el método organosolv .....	235
Cuadro 147. Extracción de pulpa de celulosa en Agave, en función de cuatro factores en tres niveles ....	235
Cuadro 148. Pesos inicial y final de cerdos de distintas camadas alimentadas con cinco dietas .....	237
Cuadro 149. Estadísticos básicos de la variable peso por animal, al inicio y al final del ensayo .....	240
Cuadro 150. Análisis de varianza de los pesos inicial y final de los animales con la corrección por covarianza .....	241
Cuadro 151. Producción de maíz ( $Y = \text{kg.parcela}^{-1}$ ) en función de la fertilización y densidad poblacional ..	243
Cuadro 152. Análisis de regresión entre densidad y producción de grano por parcela en maíz .....	244
Cuadro 153. Estadísticos de la covarianza entre producción ( $\text{kg.parcela}^{-1}$ ) y número de plantas ( $\text{parcela}^{-1}$ ) .....	244
Cuadro 154. Plantas $\text{parcela}^{-1}$ , medias de producción observada y predichas y residuos en el modelo lineal ....	245
Cuadro 155. Análisis de varianza de la producción corregida por la covarianza densidad .....	245
Cuadro 156. Promedios de peso de los racimos de banano en función de dos niveles de fertilización compleja .....	247
Cuadro 157. Estadísticos básicos del tamaño del fuste (m) en los rodales A y B .....	248
Cuadro 158. Producción de grano dos variedades de sorgo en 10 fincas .....	250
Cuadro 159. Comparaciones entre pares de tratamientos basado en la diferencia mínima significativa .....	252
Cuadro 160. Sumas y promedios de producción de grano de cuatro genotipos de sorgo forrajero .....	253
Cuadro 161. Separación de medias de cuatro genotipos de sorgo forrajero usando la DMS .....	254
Cuadro 162. Promedios de rendimiento de 25 variedades de maíz ordenados de mayor a menor .....	254
Cuadro 163. Separación de medias de producción de forraje de genotipos de sorgo con la prueba de Duncan .....	256
Cuadro 164. Separación de medias de producción de forraje seco de cuatro genotipos usando la prueba Tukey .....	258
Cuadro 165. Separación de medias del experimento bifactorial, variedades x fertilización, con la prueba Tukey .....	258
Cuadro 166. Separación de medias de cuatro genotipos de sorgo usando la prueba Newman-Keuls.....	260
Cuadro 167. Pesos promedios de racimos de plátano en función de las frecuencias de riego .....	262
Cuadro 168. Contrastes ortogonales del peso del racimo de plátano en función a las frecuencias de riego.....	263
Cuadro 169. ADEVA del peso de racimo de plátano con inclusión de los contrastes de frecuencias de riego ....	263
Cuadro 170. Separación de medias mediante comparaciones ortogonales del peso de huevos (Dg) en función de cuatro dietas .....	264
Cuadro 171. Sumas, promedios y valores escalares al máximo del contenido de magnesio en las hojas de frejol en función de tres tratamientos de fertilización edáfica .....	264
Cuadro 172. Separación de medias mediante comparaciones ortogonales del contenido de Mg en hojas de fréjol en función de tres tratamientos de fertilización edáfica .....	264
Cuadro 173. Comparaciones ortogonales para la separación de medias en un experimento $AB + I$ .....	265

---

---

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

---

Cuadro 174. Comparaciones ortogonales para la separación de medias en un experimento $AB + 1 + 1$ .....	265
Cuadro 175. Peso (libras.animal <sup>-1</sup> ) de los cerdos ( $A_i$ ) al nacimiento en camadas de distinto tamaño .....	267
Cuadro 176. Rangos ( $R$ ) de peso de los cerdos al nacimiento en ocho camadas de distinto tamaño .....	268
Cuadro 177. Número de larvas vivas de broca del café por recipiente en función de cuatro dietas artificiales y transformación a rangos .....	269
Cuadro 178. Procedimiento de cálculo de los rangos para el análisis de varianza KW .....	269
Cuadro 179. Cálculo del comparador estadístico de Kruskal-Wallis .....	272
Cuadro 180. Separación de medias con el estadístico CKW en ensayo de dietas artificiales para broca del café .....	273
Cuadro 181. Datos del conteo de insectos vivos en plantas, antes (AA) y después de la aplicación (14DDA) de un bioplaguicida .....	275
Cuadro 182. Diferencias del número de insectos vivos, antes (AA) y 14 después de la aplicación de un bioplaguicida y transformaciones a rangos por bloques .....	275
Cuadro 183. Comparación entre rangos de los tratamientos en la prueba de Friedman .....	277
Cuadro 184. Resultados de la comparación entre tratamiento con la prueba de Friedman .....	278
Cuadro 185. Calificaciones sensoriales asignadas a cuatro muestras de arroz por cinco evaluadores ....	278
Cuadro 186. Rangos asignados a cuatro muestras de arroz en evaluación sensorial .....	278
Cuadro 187. Comparación de cuatro genotipos de arroz usando el método de Friedman .....	280
Cuadro 188. Datos de producción de maíz duro en una muestra aleatoria de 12 fincas .....	282
Cuadro 189. Ventas mensuales y anual, en decenas, de un producto alimenticio de la marca A. ....	283
Cuadro 190. Ventas de un producto alimenticio y cálculo de las medias móviles con sus desviaciones absolutas .....	284
Cuadro 191. Efecto de la aplicación de bioestimulante sobre la producción de pasto .....	289
Cuadro 192. Análisis de la varianza de la regresión de la variable producción de biomasa en función de la aplicación de bioestimulante .....	290
Cuadro 193. Rendimiento azucarero en función de la materia prima .....	292
Cuadro 194. Datos de rendimiento del maíz en función de la fertilización nitrogenada .....	294
Cuadro 195. Relación entre variable explicativa ( $X$ ) y productos resultantes observado ( $Y$ ) y predicho ( $Y_p$ ) .....	295
Cuadro 196. Componentes del modelo polinómico de grado 3 relacionado al número de insectos muertos en función de las dosis de insecticida .....	297
Cuadro 197. Dosis de abono, infestación de una plaga, producción y valores escalares al máximo .....	300
Cuadro 198. Matriz de correlaciones lineales bivariadas .....	300
Cuadro 199. Análisis de regresión de la producción en función de las dosis de fertilizante e incidencia de la plaga $W$ .....	300
Cuadro 200. Comparación de tres alternativas tecnológicas para la producción de maíz en 10 localidades .....	304
Cuadro 201. Maquinarias A y B, productos M y N y restricciones de uso de las máquinas .....	306
Cuadro 202. Mediciones del deterioro ambiental en cuatro comunidades rurales .....	308
Cuadro 203. Magnitud e importancia como variables asociadas a la medición del impacto ambiental .....	311
Cuadro 204. Criterios ambientales para el ordenamiento territorial de la zona de playa de San Clemente ..	311
Cuadro 205. Impactos ambientales de las acciones antropogénicas en la "Propuesta de ordenamiento y uso turístico de la zona de playa de San Clemente" .....	312

---

<i>Cuadro 206. Criterios de evaluación de las dimensiones económica, ambiental y socio cultural e importancia .....</i>	<i>314</i>
<i>Cuadro 207. Grado de importancia y calificaciones de las variables económicas .....</i>	<i>315</i>
<i>Cuadro 208. Grado de importancia y calificaciones de las variables ambientales .....</i>	<i>316</i>
<i>Cuadro 209. Grado de importancia y calificaciones de las variables socio culturales .....</i>	<i>316</i>

# CAPÍTULO I: INTRODUCCIÓN A LA INVESTIGACIÓN CIENTÍFICA

## 1.1 AGRICULTURA Y ESTADÍSTICA

La agricultura es la actividad humana que implica la producción de alimentos, forraje, fibra y otros productos vegetales y animales, a través del cultivo de plantas y la cría de animales para sustentar la vida en el planeta tierra. La agricultura moderna ha experimentado avances tecnológicos significativos en aspectos como: el desarrollo de prácticas eficientes de manejo del cultivo, la mejora genética vegetal y animal, la fabricación de maquinaria agrícola y la gestión del suelo agrícola y uso del agua, entre otros. La agricultura tiene un impacto significativo sobre el medio ambiente, la economía y la seguridad alimentaria a nivel mundial, por lo tanto, resulta crucial el diseño de una eficiente gestión para asegurar la sostenibilidad en la producción de alimentos.

En la actualidad, se han diversificado los sistemas de producción, evidenciándose varias tipologías de agricultura, como: tradicional, convencional, orgánica, agroecológica, de precisión, regenerativa, biogénica y sostenible. La seguridad alimentaria, el cambio climático, la generación de empleo e ingresos, el desarrollo territorial sostenible, la conservación de los recursos naturales, la innovación tecnológica, la eficiencia productiva y la preservación del acervo cultural y del patrimonio genético son los grandes temas que tratan en las investigaciones relacionadas a las ciencias de la vida.

Los investigadores usan cotidianamente las hojas de cálculo, procesadores de texto, graficadores y programas estadísticos para analizar e interpretar series de datos en estudios planeados y tomar las decisiones estadísticas pertinentes, controlando los errores y asumiendo los riesgos de su falibilidad. El uso de los métodos estadísticos se ha intensificado, en todos los campos como son las ciencias biológicas, sociales, físicas e ingeniería. En el siglo XXI, especialmente en los últimos 10 años, se ha intensificado el uso de programas computacionales de análisis estadístico y más recientemente, el auge de aplicaciones basadas en la inteligencia artificial.

La agricultura y la estadística están estrechamente vinculados, prácticamente surgieron al mismo tiempo con la observación de los hechos. En la sistematización teórica se indica que el término “estadística” proviene del italiano *statista* y del latín *status* que significa estado. La estadística, operativamente, trata sobre la planificación de las investigaciones, en todos los campos del saber humano, los métodos y técnicas de la recolección, organización y análisis de datos cualitativos y cuantitativos, la interpretación de los resultados, la toma de decisiones y la diseminación de los logros (Suárez y Tapia, 2018, p. 12). La bioestadística se define como la aplicación de los métodos estadísticos en las ciencias de la vida y de la salud (García, 2011, p. 14).

La importancia de la estadística se resume en la frase del inglés Herbert George Wells, referida por Lind et al. (2012): “*Un día llegará en que el pensamiento estadístico será tan necesario para ejercer la ciudadanía con eficiencia, como la capacidad de leer y escribir*” (p. 2). En el escenario actual de intensos cambios del entorno agroambiental y de agudizamiento de los problemas complejos como la escasez de alimentos, las ciencias agropecuarias y conexas podrán contribuir con soluciones tecnológicas efectivas, solo basado en el conocimiento científico con apoyo de la estadística.



## 1.2 CONOCIMIENTO CIENTÍFICO

Las ciencias son las diversas ramas del saber humano que se distinguen por los métodos de búsqueda en objetos de estudio concretos. La palabra ciencia proviene del latín *Scientia* que significa: conocimiento que, según Rojas (2006, p. 5), tiene como características ser racional, sistemático y verificable. La ciencia es racional porque se fundamenta en la razón al explicar los fenómenos; es sistemática porque tiene argumentación teórica, orden y coherencia metodológica y es verificable porque es repetible y reproducible perfeccionando los métodos y resultados (Hernández et al., 2010, p. 205).

Al tratar sobre la evolución histórica de las ciencias surgen cuestiones como: ¿De qué se alimentaba el hombre nómada en la comunidad primitiva? y ¿Qué factores incidieron para hacerse sedentario?. Sin duda, las enfermedades humanas, la escasez de alimento y los fenómenos meteorológicos, fueron las preocupaciones de los “hombres de ciencia”, en aquellos tiempos, como son en la actualidad.

En los últimos años, motivados por el COVID-19, surgió la integración de científicos del mundo para buscar la cura de la enfermedad. También hay otras preocupaciones entre los hombres de ciencia, además de las enfermedades humanas catastróficas, como: cambio climático, degradación del suelo, contaminación del agua, pérdida de la biodiversidad e inseguridad alimentaria, entre otras.

En la Figura 1, se observa la interrelación entre las características del conocimiento científico: es racional porque prevalece la razón y puede explicarse; es universal porque es repetible y comunicable, es sistemático porque es metódico y fundamentado, es fiable porque hay una probabilidad de equivocarse, de perfeccionarse y verificarse; además, tiene objetividad porque se busca la exactitud y la precisión en los procedimientos y en los resultados.

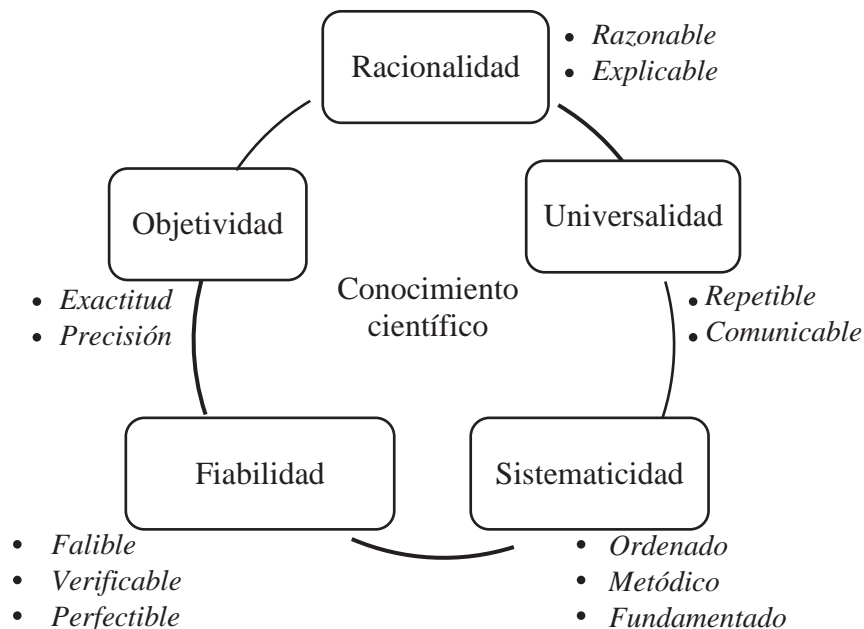


Figura 1. Características fundamentales del conocimiento científico

1.2.1 MÉTODO CIENTÍFICO

En la sociedad se ha desarrollado la pedagogía como métodos para transmitir y enseñar las ciencias y el método científico para crear ciencia (Cifuentes, 2016, p. 68; Quezada, 2021, p. 54). La investigación científica, según Ponce (2018, p. 6) y Briceño et al. (2021, p. 12) es un proceso sistemático, formal, inteligente y controlado de búsqueda de la verdad por medio del método científico. El método científico se define como el conjunto de técnicas y procedimientos para la obtención de un conocimiento teórico con validez científica, usando instrumentos fiables que no dan lugar a la subjetividad, se conoce como método científico (Briceño et al., 2021, p. 22). Éste ha evolucionado desde su nivel primigenio hasta la actualidad donde se usa la inteligencia artificial para planear, realizar el seguimiento de las investigaciones, analizar y elaborar reportes.

El método “prueba y error” se enmarca en la investigación exploratoria y es considerada de naturaleza heurística, que significa basado en la experiencia, es el procedimiento primigenio del método científico que se basa en pruebas sucesivas de alternativas, aciertos y no aciertos, para tomar la decisión de aceptar o rechazar una hipótesis.

La investigación tiene como objetivo generar conocimiento científico basado en la rigurosa aplicación del método científico, el mismo que consta de seis pasos interrelacionados y a partir de sus resultados se pueden desarrollar soluciones tecnológicas para beneficiar a la humanidad (Briceño et al., 2021, p. 12). Los pasos del método científico son: observación de los hechos, formulación de la hipótesis, prueba de hipótesis, análisis de datos e interpretación de resultados, teorización del hallazgo y comunicación (Figura 2).

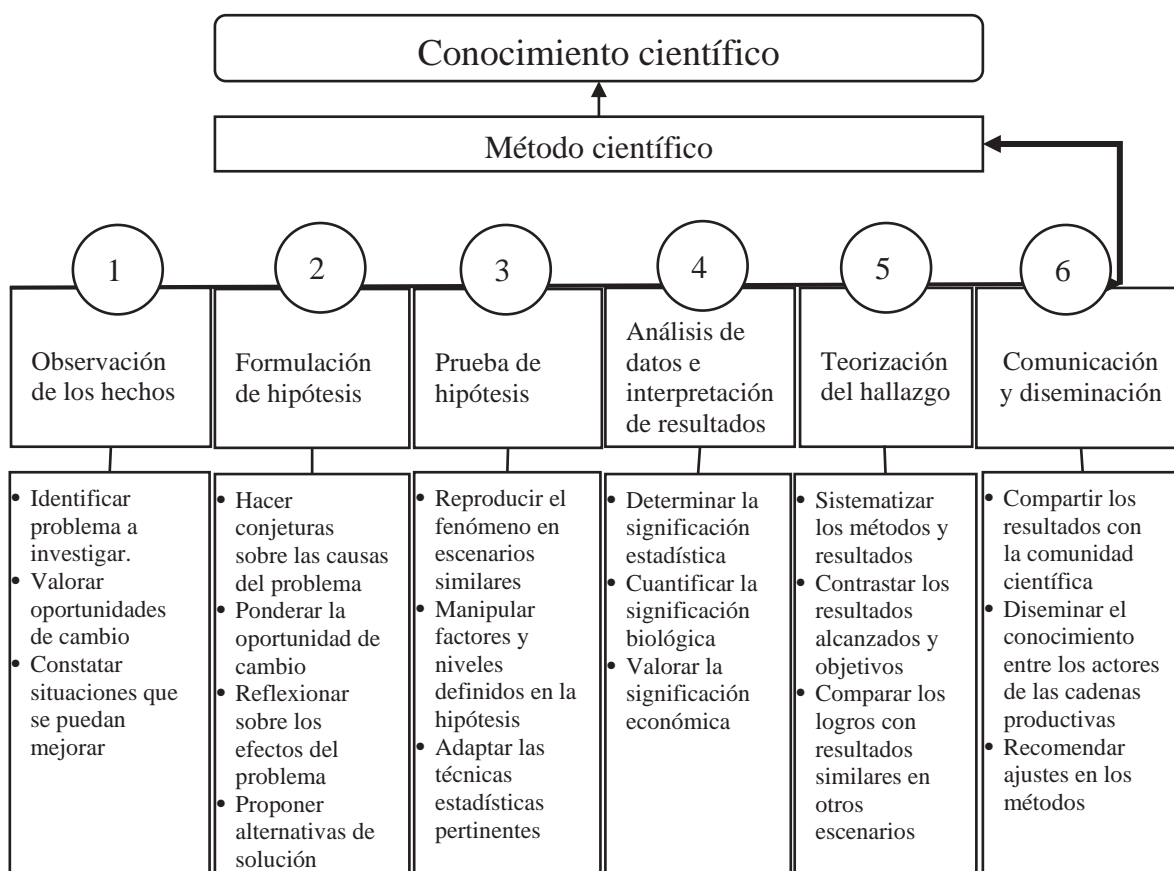


Figura 2. Los seis pasos del método científico para construir conocimiento con su argumentación



**Observación de los hechos.-** Esta es una actividad humana que se refiere a usar la vista y los otros sentidos, ya sea directamente o con instrumentos de medición que posibilitan obtener información de hechos y fenómenos. Todas las personas observan cotidianamente hechos y fenómenos; los investigadores observan, valoran detalles y miden las variables de interés en las muestras y poblaciones con el objetivo de estudiarlas.

La presencia de síntomas de deficiencias de nutrientes en las plantas, que el experto observa y valora, es un hecho. Por ejemplo: el amarillamiento de las hojas es un indicativo de la deficiencia de nitrógeno. Para probar esa conjetura, debe realizar el análisis químico del suelo. Es un hecho que la mayor parte de los suelos dedicados al cultivo de cacao en el litoral ecuatoriano son deficientes en nitrógeno, boro y azufre. Otro hecho: Las pérdidas a la economía causadas por el moko del plátano (*Ralstonia solanacearum raza 2*). A partir de la observación de los hechos se formulan la hipótesis, en distintos niveles de complejidad.

**Formulación de la hipótesis de investigación.-** Es la conjetura que conlleva a analizar el problema e identificar posibles soluciones o vislumbrar oportunidades de cambiar un hecho. ¿De qué manera se puede medir los efectos del problema?, ¿Cuáles son los factores causales del problema?, ¿Hay soluciones válidas?, ¿Cómo han abordado otros investigadores el mismo problema?, ¿Es posible resolver el problema?

La hipótesis de investigación, por lo tanto, es una idea provisoria de la solución al problema, definido como potencial logro. Por ejemplo: “El uso de variedades mejoradas de arroz y la eficiente fertilización nitrogenada contribuyen al incremento significativo de los rendimientos”.

**Prueba de hipótesis.-** En los ensayos experimentales se prueban hipótesis y se toman decisiones con una probabilidad de equivocarse. La prueba descifra el fenómeno, verifica la idea provisoria y con esta información se toma la decisión pertinente, de aceptar o de rechazar la hipótesis (Badii y Guillen, 2009, p. 186). La prueba estadística es la condición para tomar decisiones confiables. Estas pruebas pueden basarse en técnicas paramétricas y no paramétricas, en función del objetivo del estudio, de la forma de la distribución de la población y del tamaño de la muestra (Urdaneta y Urdaneta, 2016; Badii et al., 2004, p. 43).

**Análisis de datos e interpretación de resultados.-** El análisis de datos se realiza usando herramientas estadísticas como Excel<sup>1</sup> o programas estadísticos (software) que facilitan los procesos investigativos sobre los hechos, fenómenos o sucesos físicos, biológicos o agrosociales. La interpretación de resultados conlleva a tomar las decisiones estadísticas pertinentes con un nivel de confianza concreto.

**Teorización del hallazgo.-** La investigación debe concluir con un informe técnico, redactado en forma concreta con argumentos tomados de literatura especializada, vigente y pertinente. Los resultados tienen que compararse con otros estudios, de similares objetivos, realizados por otros investigadores en otros escenarios (espacio) y momentos (tiempo).

**Comunicación y diseminación.-** Los resultados de toda investigación tienen que diseminarse por distintos medios a la comunidad científica, a los profesionales, estudiantes y productores. A los científicos se les comunica a través de los artículos científicos, a los profesionales y estudiantes mediante documentos técnicos y a los productores mediante documentos divulgativos y audiovisuales. Además, la diseminación comprende las ponencias, cursos, seminarios y conferencias dirigidas a las personas interesadas o grupos de interés, que son parte de la sociedad.

---

<sup>1</sup> Excel, desarrollado por Microsoft, es un programa que permite editar hojas de cálculo con funciones matemáticas, financieras y estadísticas, facilitando la elaboración de cálculos, gráficos y tablas.

### 1.2.2 MÉTODOS AUXILIARES DEL MÉTODO CIENTÍFICO

En el desarrollo de las distintas ciencias se han generado diferentes formas de abordar los problemas y de proponer soluciones válidas. Los métodos que usan técnicas diferenciadas para observar los hechos, formular las hipótesis, realizar las pruebas, analizar datos, teorizar los resultados y diseminar los resultados se catalogan como métodos auxiliares del método científico. Entre estos se encuentran los métodos: analítico, sintético, inductivo, deductivo, analógico, modelación y dialéctico.

**Método analítico.**- Se basa en la desagregación del todo en sus partes para realizar una valoración segmentada. Los análisis físicos y químicos de suelos, por ejemplo, permite diagnosticar la situación edáfica, estos resultados se comparan con patrones preestablecidos, sobre los contenidos en niveles bajo, medio o alto; permitiendo definir la relación entre problema-factor de estudio-variable de respuesta (Rodríguez y Pérez, 2017, p. 182). El método analítico requiere de observación constante y con la experimentación se determina el comportamiento de las partes y del todo.

**Método sintético.**- Se fundamenta en el concepto de síntesis. En el método analítico se desagrega el todo en sus partes y en la síntesis se reconstruye todo. Solo si se han superado las etapas de análisis y síntesis, se pueden hacer juicios y dar explicaciones. Un juicio, en el método sintético, es la afirmación o negación de las relaciones entre objetos, variables o fenómenos. Se aplica en la elaboración de las conclusiones, las que implican una estrategia basada en el razonamiento inductivo o deductivo.

**Método inductivo.**- Se basa en la observación de hechos y a partir de un razonamiento proyectivo se sacan conclusiones, que son los juicios con un nivel de significación concreto. La inducción siempre lleva a obtener conclusiones empíricas (observación y experimentación). Ejemplo: El agua hierve con el calor, la leche hierve con el calor; por lo tanto, todo líquido hierve con el calor (Rodríguez y Pérez, 2017, p. 183).

**Método deductivo.**- Es una forma de explicar la realidad a partir de leyes o teorías generales hacia casos particulares. La conclusión de un razonamiento deductivo está incluida en las premisas. Es muy útil cuando no se pueden observar las causas de un hecho. No genera por sí mismo nuevo conocimiento, trata de verificar conocimiento previo. Por ejemplo: Las plantas con semillas en vainas son leguminosas, la soya tiene vaina, por lo tanto es una leguminosa.

**Método analógico.**- Consiste en encontrar dos situaciones o sistemas similares o análogos. Si en uno de los sistemas se obtiene una conclusión, en el otro sistema seguro que se obtendrá un resultado similar. El método analógico toma una experiencia pasada y compara con una actual para llegar a conclusiones. No se pueden comparar dos cosas solo por su parecido, debe haber una relación (Rodríguez y Pérez, 2017, p. 187).

**Modelación.**- Entre un modelo y un fenómeno siempre hay una analogía. La modelación es una reproducción esquematizada de la realidad. Un fenómeno puede involucrar varios modelos. Las constantes y las variables de un modelo se interpretan en base a una teoría científica. Ejemplos de modelos: el modelo del átomo en química, el modelo de célula vegetal en biología y el modelo de ADN de Watson-Crick en genética. La modelación comprende también la construcción de modelos matemáticos o modelos institucionales (Rodríguez y Pérez, 2017, p. 188). Ciertamente, la teoría y práctica de la bioestadística se fundamenta en la modelación matemática.

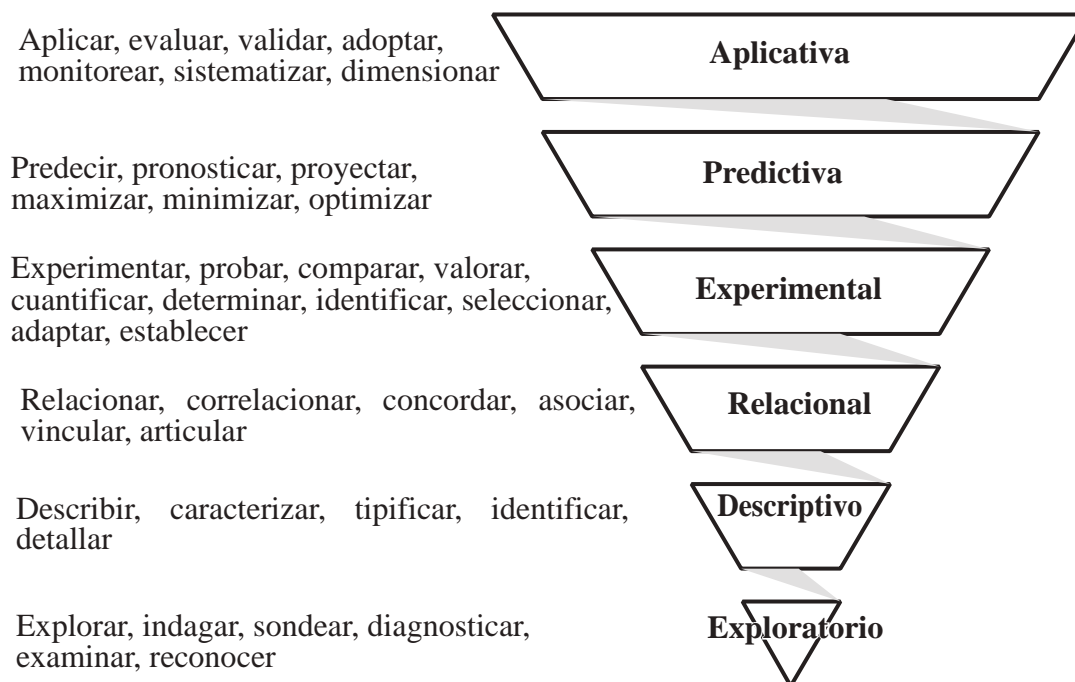
**Método dialéctico.**- El desarrollo de la naturaleza, de la sociedad y del pensamiento humano es resultante de las contradicciones existentes en ellas (Ponce, 2018, p. 4). Las características

del método dialéctico son: (1) la unidad y contradicción son parte consustancial de la naturaleza, de la sociedad y del pensamiento humano, (2) el movimiento perenne es indicativo de que todo se transforma, (3) la sucesión cíclica de cambios cuantitativos y cualitativos es la clave de la evolución; y, (4) las contradicciones internas en la naturaleza, sociedad y pensamiento humano constituyen la fuerza decisiva para el desarrollo (Izaguirre, 2014, p. 129).

### 1.2.3 NIVELES DE INVESTIGACIÓN

Una agenda de investigación institucional busca resolver los problemas reales que afronta la sociedad, o un segmento de ella, en distintos ámbitos, desde lo simple hacia lo complejo, enmarcado en distintos niveles investigativos, adaptando los métodos y las técnicas a los distintos escenarios y momentos.

Las explicaciones acerca de los niveles investigativos frecuentemente son confusas y contradictorias, según afirma Rojas (2015, p. 2). Esta situación motivó a adoptar la propuesta de Espinoza y Ochoa (2020, p. 96) y de Supo (2017, p. 9) quienes categorizan la investigación en seis niveles: exploratorio, descriptivo, relacional, explicativo, predictivo y aplicativo. En la Figura 3, se indica la relación entre los niveles investigativos y los verbos apropiados que definen los objetivos de los distintos estudios.



Fuente: Supo (2017)

Figura 3. Niveles de investigación científica y su relación con los objetivos de estudio

**Investigación exploratoria.-** Es aquella que trata sobre un fenómeno, situación o hecho desconocido o donde hay poca información. En la investigación exploratoria no hay hipótesis, se exploran escenarios para tener una idea del porqué un hecho o fenómeno, usando técnicas descriptivas como: análisis de frecuencias, elaboración de histogramas y en situaciones sanitarias, el análisis de la especificidad y sensibilidad.

Los resultados contribuyen a definir las hipótesis de investigación descriptiva. El procedimiento valora una situación dada, identifica un objeto de estudio y facilita la explicación de potenciales oportunidades.

**Investigación descriptiva.-** Trata sobre un fenómeno, situación o hecho donde falta información detallada sobre los factores que originan las variaciones en la forma, estructura y caracteres medibles usando escalas nominales, ordinales o de intervalo constante. En el nivel investigativo descriptivo si hay hipótesis y conjeturas de potenciales cambios. Se detallan los escenarios y circunstancias.

Los estudios descriptivos posibilitan diagnosticar, tipificar, caracterizar y describir fenómenos, situaciones, hechos y cambios en las variables de interés, usando técnicas univariadas, bivariadas o multivariadas. Los resultados de la investigación descriptiva contribuyen a definir las hipótesis de investigaciones de los niveles relacional y explicativo.

**Investigación relacional.-** Se orienta a identificar y establecer relaciones entre las variables, sin la manipulación de factores. En este nivel investigativo si hay hipótesis, supuestos y conjeturas de asociaciones estadísticas entre variables.

Se valora la asociación estadística entre fenómenos, situaciones o hechos, usando diversas técnicas paramétricas y no paramétricas sin abordar la causalidad. Son estudios del tipo “Cuasi experimental” cuyos resultados ayudan a plantear las hipótesis sobre los niveles experimental y predictivo.

**Investigación experimental.-** Tiene el propósito de explicar la causalidad del fenómeno, situación o hecho. Se caracteriza porque hay manipulación de los factores en estudio. En este nivel investigativo si hay hipótesis y conjeturas de potenciales causas y efectos.

Se crean escenarios con aproximación a la realidad para probar factores y niveles que expliquen los hechos usando diversas técnicas estadísticas. Los resultados de la experimentación conllevan a definir las hipótesis de los niveles predictivo y aplicativo.

**Investigación predictiva.-** Tiene el propósito de prever los comportamientos de fenómenos, situaciones y hechos (variables dependientes), basado en el comportamiento de variables independientes. En el nivel predictivo si hay hipótesis y supuestos de potenciales causas y efectos, en escenarios fijos o dinámicos. Se construyen modelos matemáticos y gráficos para representar los cambios en el comportamiento de las variables dependientes en función de factores causales y sus niveles en estudio.

Estas investigaciones crean escenarios potencialmente probables y respuestas de predicción, con aproximación a la realidad. Se proyecta a escenarios orientados a maximizar la producción, minimizar los costos y optimizar los beneficios. En la construcción de modelos matemáticos se usan las técnicas de regresión. Los resultados de estas investigaciones contribuyen a definir las hipótesis de estudios del nivel aplicativo.

**Investigación aplicada.-** Tiene el propósito de validar, verificar y aplicar las alternativas integradas como soluciones tecnológicas, cumpliendo las premisas de las significancias biológica, estadística y económica.

En este nivel investigativo, las hipótesis denotan los potenciales cambios en el sector agroproductivo. Un ejemplo de este nivel investigativo es la evaluación de impacto ambiental o las evaluaciones ex post de proyectos donde se usan generalmente herramientas de análisis multicriterio para su valoración.

#### 1.2.4 TIPOS DE INVESTIGACIÓN

Los estudios investigativos se enmarcan en un tipo de investigación concreto. Los tipos de investigación se clasifican por la intervención del investigador, por la planificación de las mediciones, por el número de observaciones a lo largo del tiempo, por el número de variables analíticas, por el grado de aplicación, por la naturaleza de los datos y por el grado de manipulación de las variables (Briceño et al., 2012, p.14; Supo y Cavero, 2014, pp. 51-56). En la actualidad, se han ampliado los tipos de investigación (Quezada, 2021, p. 31), destacándose por la participación social y por la contribución a la formación académica.

##### 1.2.4.1 POR LA INTERVENCIÓN DEL INVESTIGADOR

Los estudios en función de la intervención del investigador se clasifican en observacionales y experimentales.

**Investigación observacional.**- Se trata de estudios sin intervención del investigador, sean exploratorios, descriptivos y relacionales. Ejemplo: *Incidencia de moniliasis en clones de cacao EETP-800 y EETP 801.*

**Investigación experimental.**- Trata de los estudios con manipulación de los factores y niveles en estudio. Pueden tratarse de investigaciones prospectivas, transversales o longitudinales con análisis de la varianza y regresiones. Ejemplo: *Efecto de las enzimas pectolíticas en la fermentación y en la calidad organoléptica del cacao.*

##### 1.2.4.2 POR LA PLANIFICACIÓN DE LAS MEDICIONES

Por la planificación de las mediciones, la investigación puede ser prospectiva y retrospectiva.

**Investigación prospectiva.**- Se fundamenta en la información primaria y los datos que provienen de mediciones planeadas con participación directa del investigador. Ejemplo: *Calidad sensorial de híbridos de arroz en Daule, Ecuador.*

**Investigación retrospectiva.**- Utiliza datos de registros preexistentes, de mediciones donde no hubo participación del investigador (información secundaria). Ejemplo: *Promedios de calificaciones de los estudiantes de agronomía en los últimos cinco años.*

##### 1.2.4.3 POR EL NÚMERO DE OBSERVACIONES EN EL TIEMPO

Los estudios de este tipo pueden ser transversales o longitudinales.

**Investigación transversal.**- Los experimentos se instalan simultáneamente en varias localidades de interés, en un momento concreto (por ejemplo: ciclo de cultivo de invierno). En los estudios no experimentales o cuasi experimentales, se toman los datos pertinentes en un momento concreto, con fines de comparación. Pueden ser exploratorios, descriptivos, relacionales o experimentales. Ejemplo: *Comparación de 13 variedades de arroz en tres localidades de Manabí, en la época seca del 2021.*

**Investigación longitudinal.**- Trata de mediciones en estudios a lo largo del tiempo, pudiendo ser de tipo relacional (sin intervención) o experimental y predictivo cuando hay intervenciones (manipulación de un factor limitante). Ejemplo del primer caso: Curva epidemiológica de la roya del café arábigo en Piñas, provincia El Oro. Ejemplo del segundo caso: Incidencia de escoba de bruja en función de la poda y aplicación de clorotalonil en el cultivo de cacao.



#### 1.2.4.4 POR EL NÚMERO DE VARIABLES ANALÍTICAS

En función del número de variables, la investigación puede ser descriptiva o analítica.

**Investigación descriptiva.**- Se indica explícitamente, en el título y objetivo del estudio, una o más variables experimentales. Los análisis de frecuencias y las medidas de tendencia central, de dispersión y de forma son estudios descriptivos. Ejemplo: *“Uso eficiente de maquinaria agrícola en la producción de maíz en cinco organizaciones de agricultores”*.

**Investigación analítica.**- De modo explícito, se indica, en el título y objetivo general, las variables de interés propuesta en el objetivo. Por ejemplo: *“Efecto del bioestimulante ADMF sobre la incidencia de Fusarium oxysporum R1 en el cultivo de plátano”*.

#### 1.2.4.5 POR EL GRADO DE APLICABILIDAD

Según el grado de aplicabilidad, la investigación se clasifica como: básica y aplicada.

**Investigación básica.**- También identificada como pura, tiene el objetivo de generar conocimiento, sin prever aplicabilidad. Ejemplos: 1) *Algoritmo para predecir la radiación solar en un territorio;* y, 2) *Identificación de una proteína.*

**Investigación aplicada.**- Tiene el objetivo de encontrar aplicaciones prácticas para el conocimiento generado en la perspectiva de generar beneficios directos. Ejemplos: 1) *Marcadores moleculares en genotipos de maíz de alto contenido de proteína;* y, 2) *Optimización de dosis de azufre en suelos dedicados al cultivo de cacao.*

#### 1.2.4.6 POR LA NATURALEZA DE LOS DATOS

Por la naturaleza de los datos, según Briceño et al. (2012), la investigación puede ser cualitativa o cuantitativa (p. 16).

**Investigación cualitativa.**- Los datos cualitativos del estudio se obtienen con el uso de escalas nominales u ordinales y para analizarse deben ser transformados. Por ejemplo: 1) *Niveles de satisfacción en el empleo en la Fabril,* 2) *Diagnóstico agrosocial en las comunidades rurales del cantón Puyango, Loja.*

**Investigación cuantitativa.**- Se basa en el uso de datos cuantitativos discretos o continuos, obtenidos mediante mediciones con escalas de intervalo constante. donde se evidencia un total control de los factores y los resultados son generalizables. Ejemplo: *Efecto de los fertilizantes edáficos a base de nitrógeno y azufre sobre el rendimiento de plátano var. Barraganete, en el cantón El Carmen, Manabí.*

#### 1.2.4.7 POR EL GRADO DE MANIPULACIÓN DE LAS VARIABLES

En función de la manipulación de las variables, la investigación puede clasificarse como: experimental, cuasi experimental o no experimental.

**Investigación experimental.**- En estos estudios, la variable independiente (X) es manipulada en el propósito de reproducir un efecto medible sobre la variable de respuesta (Y). Los datos cuantitativos deben provenir de muestras representativas y la distribución normal. Ejemplo: *Efecto de las dosis nitrógeno de origen orgánico sobre el incremento de la productividad del maíz blanco cristalino.*

**Investigación cuasi experimental.**- En estudios de este tipo, las variables independientes no son manipuladas. Los datos pueden ser cuantitativos o cualitativos. En este tipo se ubican los estudios observacionales: exploratorios, descriptivos y correlacionales. Ejemplo 1:

*Relación entre altitud de la zona de cultivo y la calidad sensorial en cafés arábigos* (Duicela, Velázquez y Farfán, 2017, p. 76). Ejemplo 2: *Evaluación correlacional de la asociación cacao-leguminosas en la sostenibilidad del cacao (Theobroma cacao L.), en Santo Domingo de los Tsáchilas* (Anzules et al., 2015, p. 268).

**Investigación no experimental.**- En estos estudios no se controlan los factores en estudio ni las variables de respuesta. Puede tratarse de estudios exploratorios, descriptivos y correlacionales, y los datos cualitativos o cuantitativos. Ejemplo: *Tipificación de los sistemas de producción prevalentes en las comunidades rurales del cantón Loreto, Orellana*.

#### 1.2.4.8 POR LA PARTICIPACIÓN SOCIAL

La integración de los actores de las cadenas productivas en el diagnóstico de la problemática, en la planificación de estrategias para buscar soluciones en forma colectiva y en la ejecución define a la investigación como convencional y participativa.

**Investigación convencional.**- Es el enfoque de la investigación planeada, ejecutada y evaluada por equipos de investigación que buscan generar conocimiento y soluciones tecnológicas a demandas concretas de la sociedad. En estos procesos, los actores de las cadenas productivas son sujetos pasivos, reciben la información proporcionada por la academia o las instituciones de investigación y la adoptan cuando constatan sus beneficios.

**Investigación participativa.**- Es el enfoque inclusivo de los actores de las cadenas productivas en los procesos de formulación y gestión de la investigación, cubriendo todas las etapas, como: diagnóstico indagativo, identificación y priorización de problemas y objetivos, planificación, ejecución, seguimiento y evaluación (Geilfus, 1997; De Boef y Thijssen, 2007). Las instituciones de investigación y la academia se convierten en facilitadores de procesos de la acción colectiva que busca soluciones tecnológicas a las demandas concretas de las comunidades y de la sociedad.

#### 1.2.4.9 POR LA CONTRIBUCIÓN A LA FORMACIÓN ACADÉMICA

A nivel de las instituciones de educación superior (IES), la investigación se clasifica como formativa y generativa.

**Investigación formativa.**- Es el conjunto de actividades informales o formales, que incluyen tesis de pregrado y posgrado, revisiones bibliográficas de temas de interés, incluso usando las aplicaciones de la inteligencia artificial, trabajos de aplicación experimental de las asignaturas, participación de los estudiantes en los programas “semillero de investigadores”. Estas actividades de investigación, se planean y ejecutan cotidianamente como parte del proceso “enseñanza-aprendizaje”.

**Investigación generativa.**- Es el conjunto de programas, proyectos y ensayos que tienen la finalidad de generar soluciones tecnológicas a los problemas del campo agropecuario, enmarcados en el mandato institucional de hacer investigación.

La formulación y gestión de proyectos está a cargo de los grupos de investigación y de los docentes investigadores. En estos procesos institucionales se integran los proyectos de investigación, multidisciplinarios y transdisciplinarios, alineados a la política pública, concretada en el plan nacional de desarrollo, las líneas de investigación de la SENESCYT y los objetivos de desarrollo sostenible de las Naciones Unidas.

### 1.3 PLANIFICACIÓN DE LA INVESTIGACIÓN

La investigación se desarrolla en dos momentos claves: la planificación y la ejecución. La planificación comprende desde la identificación de la “idea del proyecto”, la definición de objetivos y la elaboración de la propuesta, actividades que en conjunto constituyen la etapa de formulación. La gestión empieza con el análisis y aprobación de la propuesta, la ejecución y la evaluación *ex post*. Transversal a las actividades de ejecución hay una actividad que se conoce como seguimiento, que lo realizan los delegados institucionales con el propósito de verificar el cumplimiento del cronograma, del presupuesto de los logros del proyecto.

La planificación de la investigación agropecuaria garantiza el rigor científico, la relevancia temática y la aplicabilidad práctica, en el proceso de generación de conocimiento científico y de soluciones tecnológicas para resolver los problemas de producción y productividad, atendiendo la creciente demanda de alimentos.

#### 1.3.1 PROBLEMÁTICA DE INVESTIGACIÓN

En el análisis de los problemas del sector agroproductivo debe haber participación de los productores, investigadores, extensionistas y asesores técnicos para que el diagnóstico tenga objetividad y haya involucramiento social en la búsqueda de soluciones. En la Figura 4, se expone un diagrama referencial de la problemática del sector agropecuario.

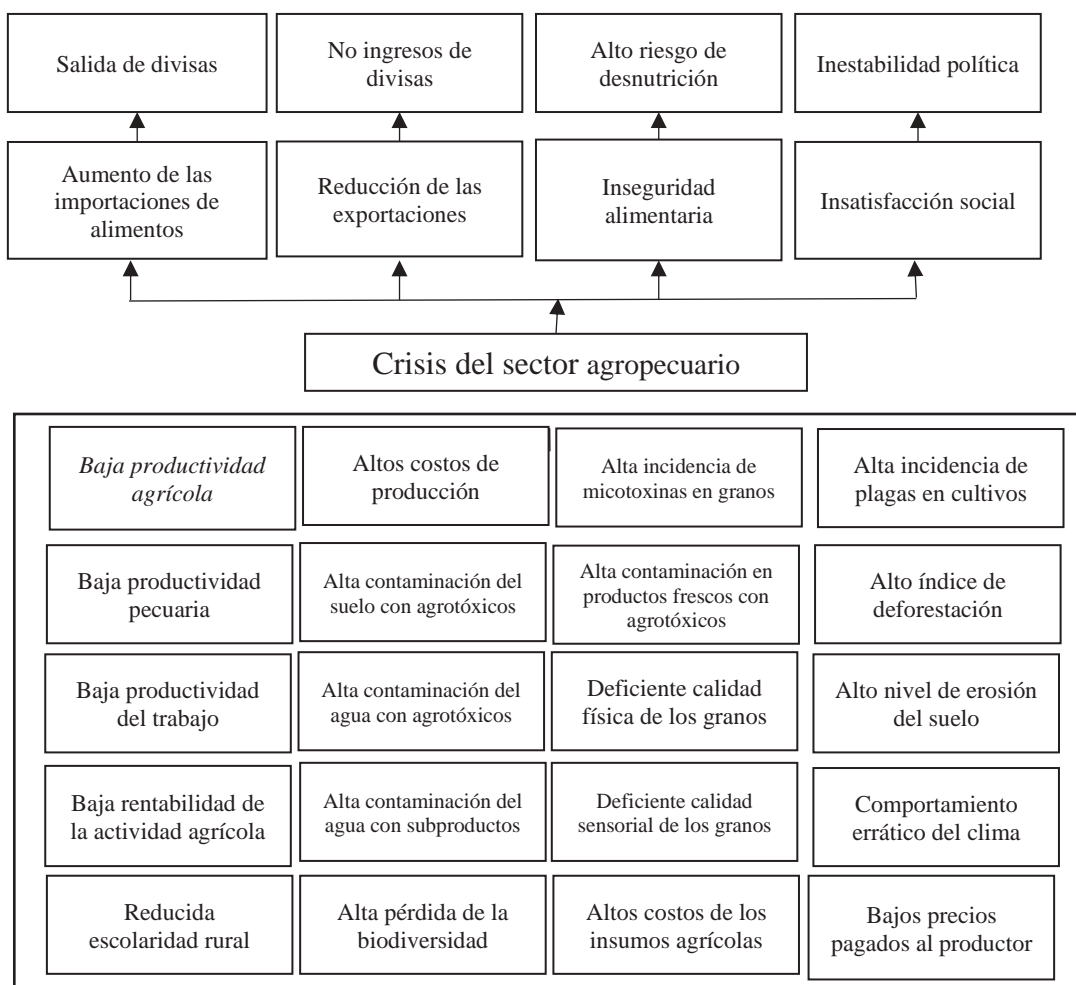


Figura 4. Causas y consecuencias de la crisis del sector agropecuario en el Ecuador



1.3.1.1 CASO DE ANÁLISIS: SECTOR CAFETALERO ECUATORIANO

El análisis del sector agropecuario conlleva a tratar problemáticas concretas por rubro de interés, por ejemplo: sector cafetalero (Figura 5). Los actores de la cadena cafetalera con el apoyo de las instituciones públicas y privadas, elaboraron un árbol de problemas, basado en un análisis de relaciones de causalidad, organizando y jerarquizando las relaciones “causa - efecto” sobre el problema central y sus consecuencias para el país.

El problema central del sector cafetalero es la “Baja producción nacional”, la producción estimada es de 300 mil sacos de 60 kg, mientras la demanda supera los 2 millones de sacos. Más de un millón de sacos son requeridos por la industria de café soluble, que se cubre con la importación del grano desde Vietnam, Brasil, Indonesia y Costa de Marfil.

La baja producción nacional se debe a causas de primer orden, como: baja productividad, reducción del área cafetalera, prevalencia de cafetales viejos, falta de crédito y comportamiento errático del clima. La baja productividad tiene las siguientes causas: uso de genotipos de origen desconocido, deficiente manejo del suelo y agua, alta incidencia de broca y otros problemas sanitarios, falta de tecnologías en sistemas productivos diferenciados y la no diseminación de las tecnologías disponibles.

La solución a la “baja productividad” requiere de un enfoque multidisciplinario y la organización de los productores de cafés arábigos y robustas, para dar respuestas válidas a la problemática del sector cafetalero.

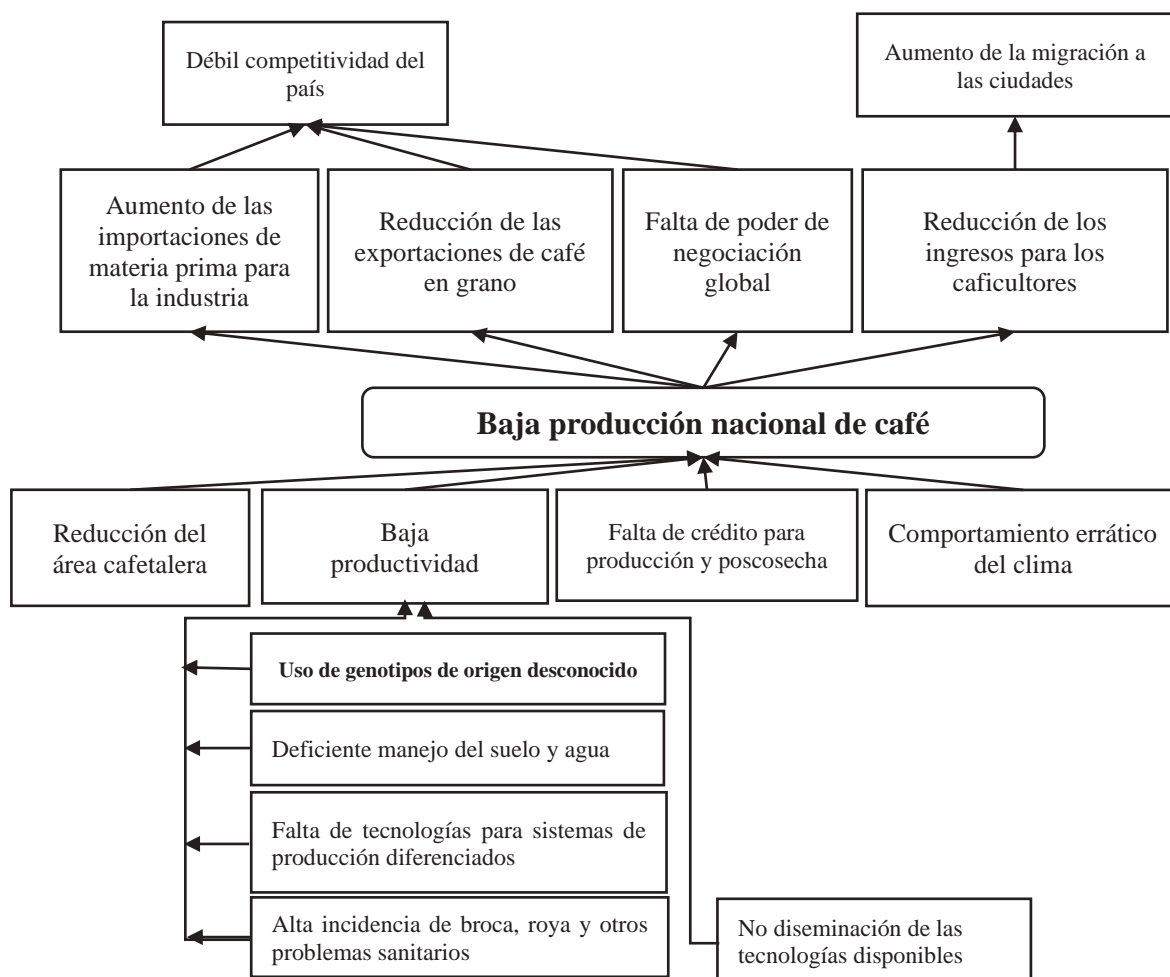


Figura 5. Árbol de problemas del sector cafetalero ecuatoriano

1.3.2 OBJETIVOS DE INVESTIGACIÓN

La elaboración de un árbol de objetivos, contrastando con el contenido del árbol de problemas, es una técnica de aplicación del mapa conceptual, identificando con claridad el ámbito de intervención. El árbol de objetivos se genera como un mapa conceptual en base al análisis medio-fin a partir del diagrama del árbol de problemas (Figura 6).

Cada factor causal se desagrega a un nivel operativo específico. En el ejemplo: entre los factores causales de la “baja productividad”, se enfatizan en el uso de genotipos de origen desconocido, deficiente manejo del suelo y agua, falta de tecnologías para sistemas de producción diferenciados, alta incidencia de broca, roya y otros problemas y la no diseminación de las tecnologías disponibles. En este momento se identifica si el ámbito de intervención se orienta hacia un proyecto de investigación (I) o de desarrollo (D).

A partir del análisis de los factores causales del problema central se derivan los otros problemas de menor jerarquía que se interpretan como “factores limitantes”. Por ejemplo: si el problema es baja productividad, se deduce que hay factores limitantes. Al planear un estudio, el “factor limitante se traduce a “factor en estudio”. En este momento se puede definir un objetivo concreto de la investigación.

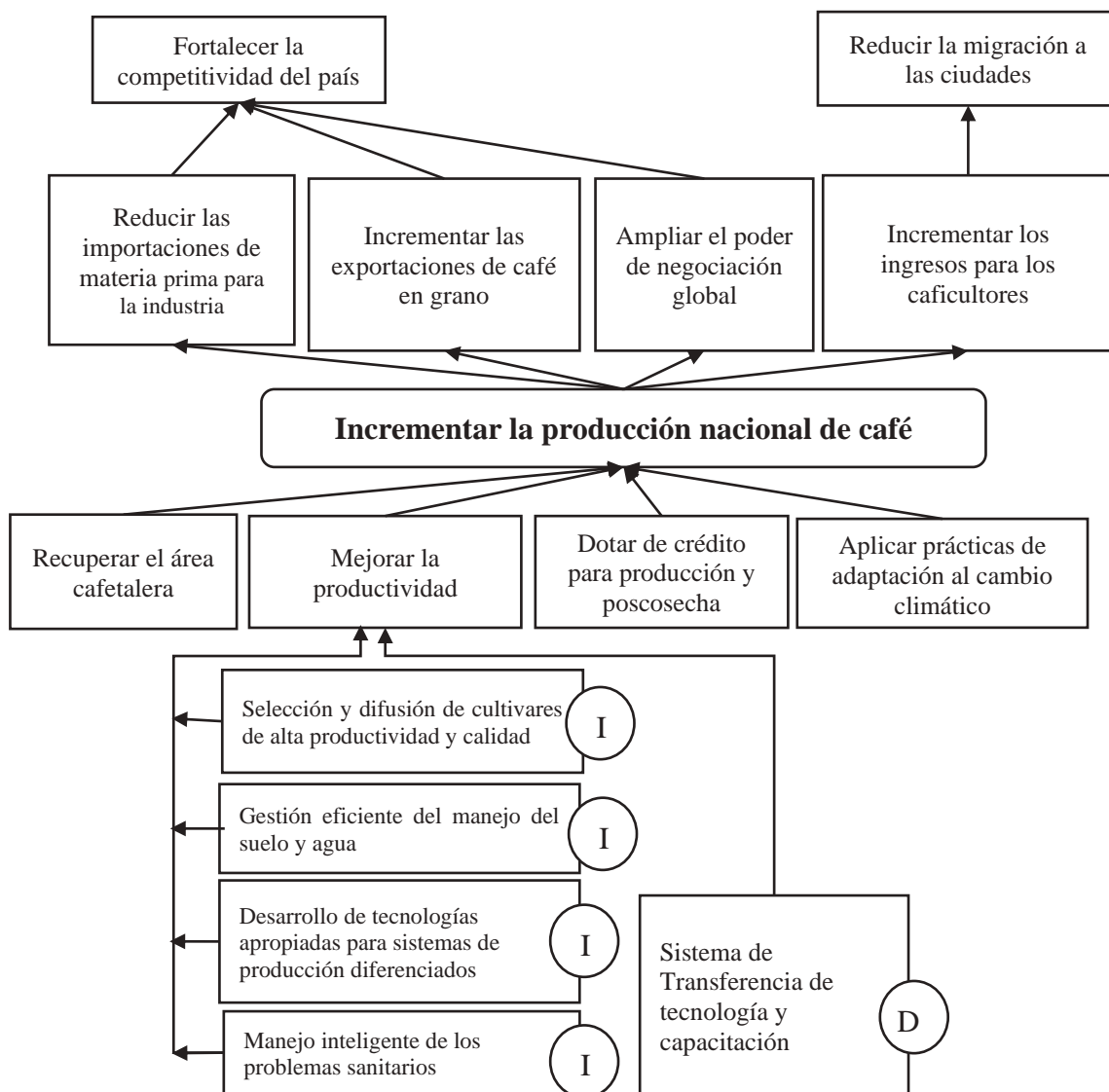


Figura 6. Árbol de objetivos del sector cafetalero ecuatoriano

### 1.3.3 FORMULACIÓN DE LA PROPUESTA DE INVESTIGACIÓN

El plan de investigación institucional se integra por el conjunto de programas. Un programa está compuesto de proyectos. Un proyecto está conformado por ensayos y estudios de distintos niveles investigativos. Un ensayo involucra una prueba estadística que tiene el objetivo de aceptar o de rechazar una hipótesis, en tanto que el término estudio no exige de una prueba estadística, porque principalmente abordan los niveles exploratorio y descriptivo.

Un plan de investigación de una institución de educación superior (IES) debe alinearse, en forma explícita, a las prioridades y líneas institucionales, a las políticas del plan nacional de desarrollo, a las líneas de investigación de la Secretaría Nacional de Educación Superior, Ciencia y Tecnología (Acuerdo Nro. SENESCYT 2022-016) y a los objetivos de desarrollo sostenible de la Organización de las Naciones Unidas [ONU] (ONU, 2015). En lo legal, tiene que cumplir el mandato de la Constitución política, la ley de educación superior y el estatuto universitario, con sus normativas.

La formulación y gestión de los planes, programas y proyectos de investigación están bajo la responsabilidad de los órganos de planificación institucional, de las autoridades y de sus delegados, que tienen el mandato de impulsar la investigación científica.

#### 1.3.3.1 PROGRAMA DE INVESTIGACIÓN

La planificación de programas de investigación es a largo plazo y tiene el propósito de generar soluciones tecnológicas a los problemas del del agro. En el caso del café, un problema de primer orden es la “baja productividad”, por lo tanto, se define como objetivo: “incrementar la productividad” y para lograrlo se requiere de un “programa de mejora de la productividad del café”.

En base al árbol de objetivos, el programa estaría conformado por los proyectos:

- Selección y difusión de cultivares de alta productividad
- Gestión eficiente del manejo del suelo y agua
- Desarrollo de tecnologías apropiadas para sistemas de producción cafetaleros
- Manejo inteligente de problemas fitosanitarios
- Sistema de transferencia de tecnología y capacitación

Los proyectos de investigación y ensayos tienen que priorizarse en función de la demanda interna (productores) y externa (necesidades del mercado), de los talentos y de los recursos disponibles. Si los actores de la cadena productiva priorizan el proyecto: “Selección y difusión de cultivares de alta productividad”, cabe el análisis, reflexión y decisiones sobre cuestiones como: ¿Qué requiere el país? (demanda), ¿Dónde se deben probar los nuevos genotipos?, ¿Se cubren los dominios de recomendación de mayor importancia?, ¿Hay probabilidades de éxito?, ¿Cuáles son las experiencias previas del equipo de investigación?.

Un dominio de recomendación es el territorio que comparte problemáticas y condiciones biofísicas parecidas, donde se pueden aplicar los resultados de la investigación con mayor confiabilidad, pues se ajustan a las circunstancias de los productores y responden a sus demandas tecnológicas específicas (Lores, 2008, p. 6). La valoración de la situación conlleva a la revisión de literatura, para conocer detalles como: ¿Cuál es la densidad poblacional apropiada? (si se desconoce, hay que experimentar), ¿Cuál es la frecuencia de riego óptima?, ¿Cuál es la fertilización óptima?, entre otras.

En el Cuadro 1, se presenta una matriz de proyectos que podrían conformar un programa “Mejora de la productividad del café”. ¿Por dónde empezar?

## HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA

Cuadro 1. Programa de investigación “Mejora de la productividad del café”

Proyectos	Ensayos y actividades/proyecto			
Selección y difusión de cultivares	Bancos de germoplasma de café	Selección de variedades de café	Desarrollo de híbridos de café	Estabilidad genética
Manejo del suelo y agua	Optimización de la fertilización en cafetales	Enmiendas y acondicionadores de suelos en cafetales	Fertilización foliar y uso de bioestimulantes	Sistemas de riego en cafetales
Sistemas de producción	Densidades poblacionales en cafetales	Asociación de cultivos en cafetales	Sistemas agroforestales en cafetales	Captura de carbono en SAF cafetaleros
Manejo inteligente de problemas sanitarios	Manejo integrado de plagas insectiles	Control integrado de enfermedades foliares	Estudios entomológicos especiales	Estudios epidemiológicos especiales
Transferencia de tecnología y capacitación	Formación de capacitadores en caficultura sostenible	Escuela de campo en caficultura sostenible	Formación de semilleros y viveristas	Eventos sobre la ciencia del café

### 1.3.3.2 PROYECTO DE INVESTIGACIÓN

Los proyectos I + D + i son aquellos que tienen una connotación integradora y multidisciplinaria orientada a la generación de soluciones tecnológicas y a la mejora continua de procesos, productos o servicios (Investigación), a la aplicación de las alternativas en distintos escenarios (Desarrollo) y la puesta en valor presente y en valor futuro los beneficios de la adopción (innovación).

El árbol de objetivos orienta el trazado de la ruta de la investigación. Una herramienta de uso práctico, complementaria para analizar un rubro agrícola u otra área del conocimiento, es el diagrama de causa-efecto que ayuda a identificar el ámbito de intervención de un proyecto con un enfoque holístico (Báez et al., 2010, p. 67). El análisis del diagrama sugiere los temas del programa o proyectos para el corto, mediano y largo plazo. En la Figura 7, se indica un diagrama causa-efecto, como referencia para la producción agrícola.

En “administración eficiente”, las temáticas investigativas pueden referirse al uso de la tecnología, medición de la eficiencia productiva, sistemas de control y mecanismos de financiamiento. En “cultivares mejorados” se tratarían temas como: selección de genotipos, interacción genotipo x ambiente, resistencia a plagas, calidad física del grano y calidad sensorial de la bebida.

La “aptitud agroecológica” conlleva a estudiar las características físicas, químicas y biológicas del suelo; los cambios en el comportamiento del clima, la captura de carbono; la fenología en función de las lluvias, temperatura, humedad y heliofanía, la dinámica de poblaciones insectiles, las curvas epidemiológicas de enfermedades, en la fisiografía se tratarían sobre la evolución de las zonas de vida<sup>2</sup> y las variaciones del entorno físico por factores antropogénicos, entre otros.

En las “buenas prácticas de producción” se tratarían temáticas sobre la fertilización química y orgánica, a la mecanización agrícola, al uso eficiente del agua de riego, a la preparación de semillas, a la crianza de plántulas, a los métodos de propagación, a la densidad poblacional, entre otras. En “buenas prácticas en poscosecha” se trataría sobre los

<sup>2</sup> La zona de vida se define como una unidad bioclimática natural en que se agrupan diferentes asociaciones de poblaciones vegetales correspondientes a determinados ámbitos de temperatura, precipitación y humedad.

métodos de cosecha (manual y mecanizada), de beneficio, de almacenamiento, de transporte y de clasificación.

El diagrama causa-efecto también orienta la conformación de los Grupos de Investigación. Al definir como objetivo, “incrementar la producción”, las temáticas y subtemas pueden abordarse por los especialistas (subtemas), en estrecha articulación a las partes (temas) y al todo (sistema productivo).

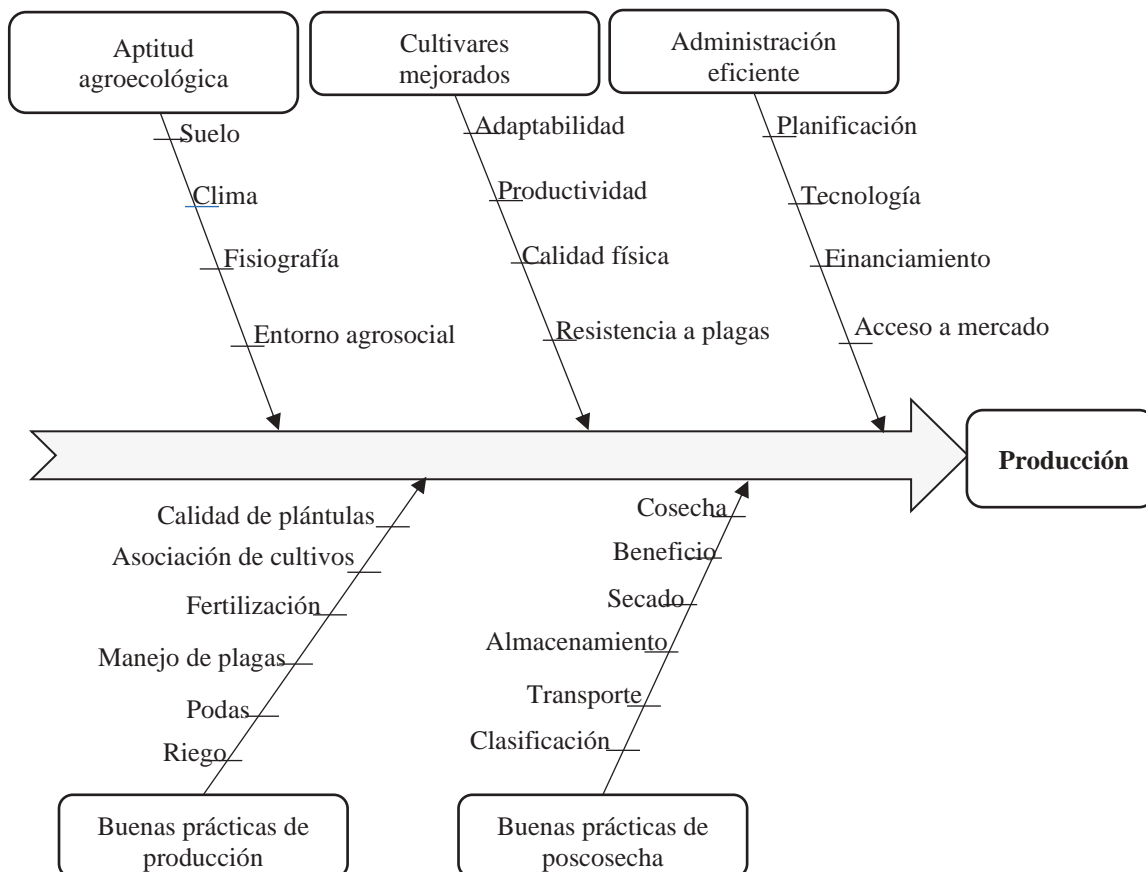


Figura 7. Diagrama causa - efecto de los factores determinantes de la producción agrícola

Los temas de investigación se tienen que priorizar con base en dos aspectos claves: (a) los factores limitantes asociados a las causas del problema (en distinta jerarquía) que se transforman en factores en estudio; y, (b) por la importancia socioeconómica de los territorios. Una provincia forma parte de una ecorregión y se conforma por cantones, parroquias y comunidades. La priorización de las localidades de intervención tiene que realizarse hasta un nivel técnicamente válido, que posibilite la extrapolación de los potenciales resultados a un “dominio de recomendación”.

La identificación de los factores y niveles a estudiar, tiene que realizarse con la acción participativa de los grupos focales que representen a las cadenas productivas, usando distintas técnicas andragógicas como: análisis de frecuencias, comparación pareada y representaciones infográficas. Un ejemplo del uso del análisis de frecuencias en la priorización de factores en estudio, se indica en el Cuadro 2.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 2. Análisis de frecuencias para la priorización de los factores causales de la baja productividad*

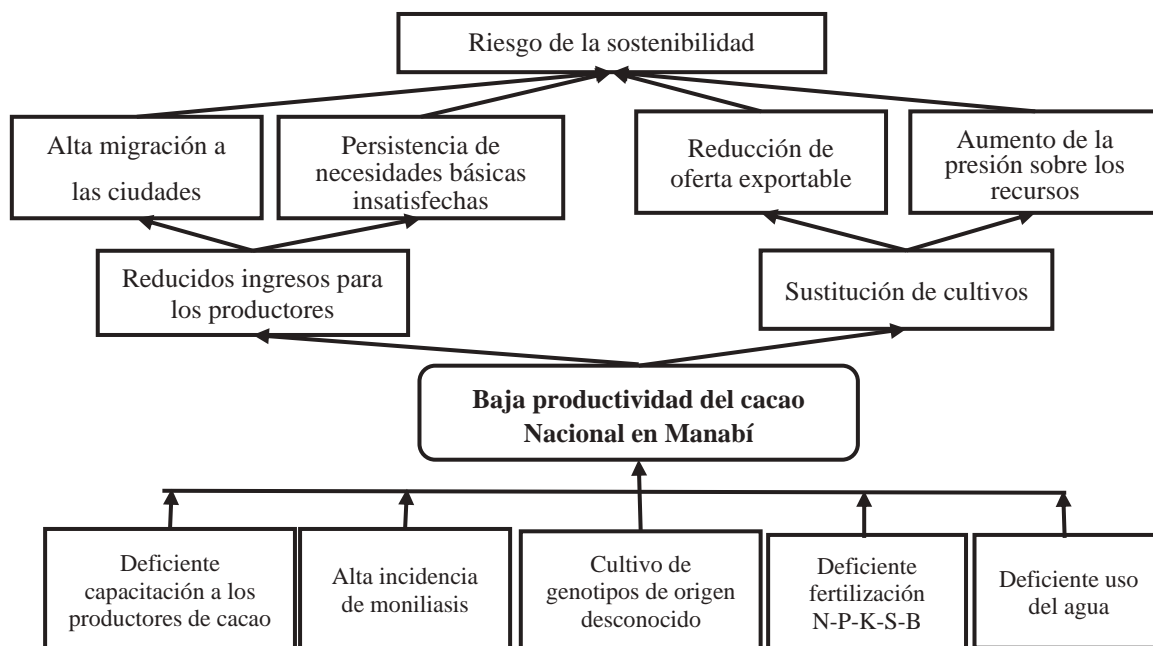
Causas de la baja productividad	Factores de estudio	Elección de prioridad	Frecuencia	Prioridad
Cultivares de bajo rendimiento	Genotipos	////////////////////	24	1
Densidades poblacionales inapropiadas	Densidad poblacional	////////////////////	22	2
Eficiente uso del agua de riego	Láminas de riego	////////////////////	21	3
Deficiencia de Boro en el suelo	Dosis de boro	////////////////////	17	4
Cultivares susceptibilidad a cierta plaga	Genotipos	//////////	14	5
Prevalencia de suelos ácidos	Enmiendas	//////////	14	5
Deficiencia de azufre en el suelo	Fuentes de azufre	////////	10	6

**Ejercicio 1:**

Elaborar un árbol de problemas sobre el cultivo de cacao Nacional en la provincia de Manabí

El problema central es la baja productividad del cacao Nacional en Manabí. Las causas del problema son: la deficiente capacitación a los productores, la alta incidencia de moniliasis, el cultivo de genotipos de origen desconocido, la deficiente fertilización con nitrógeno (N), fósforo (P<sub>2</sub>O<sub>5</sub>), potasio (K<sub>2</sub>O), azufre (S) y boro (B) (Figura 8).

En la referida figura se observa que los efectos se expresan en los reducidos ingresos para los productores, la sustitución de cultivos, la alta migración a las ciudades, la persistencia de las necesidades básicas insatisfechas, la reducción de la oferta exportable, el aumento de la presión sobre los recursos y, en general, la actividad cacaotera se coloca en situación de riesgo de su sostenibilidad.



*Figura 8. Árbol de problemas del cultivo de cacao Nacional en la provincia de Manabí*

**Ejercicio 2:**

Elaborar un árbol de objetivos sobre el cultivo de cacao Nacional en la provincia de Manabí

El objetivo general de la propuesta es incrementar la productividad del cacao Nacional que puede enmarcarse en el programa “Mejora de la productividad del cacao Nacional en Manabí”(Figura 9).

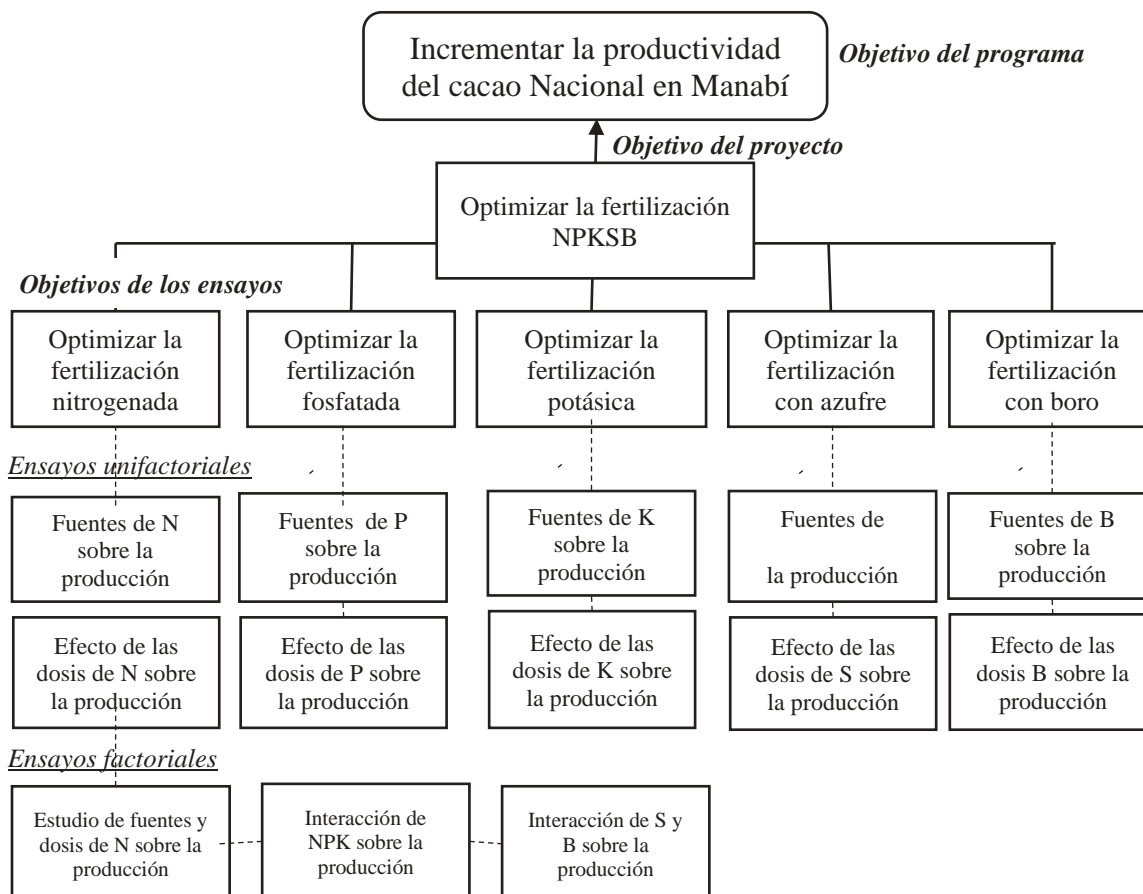


Figura 9. Identificación de objetivos de un programa de investigación en cacao Nacional en Manabí

Este programa podría estar conformado por los siguientes proyectos:

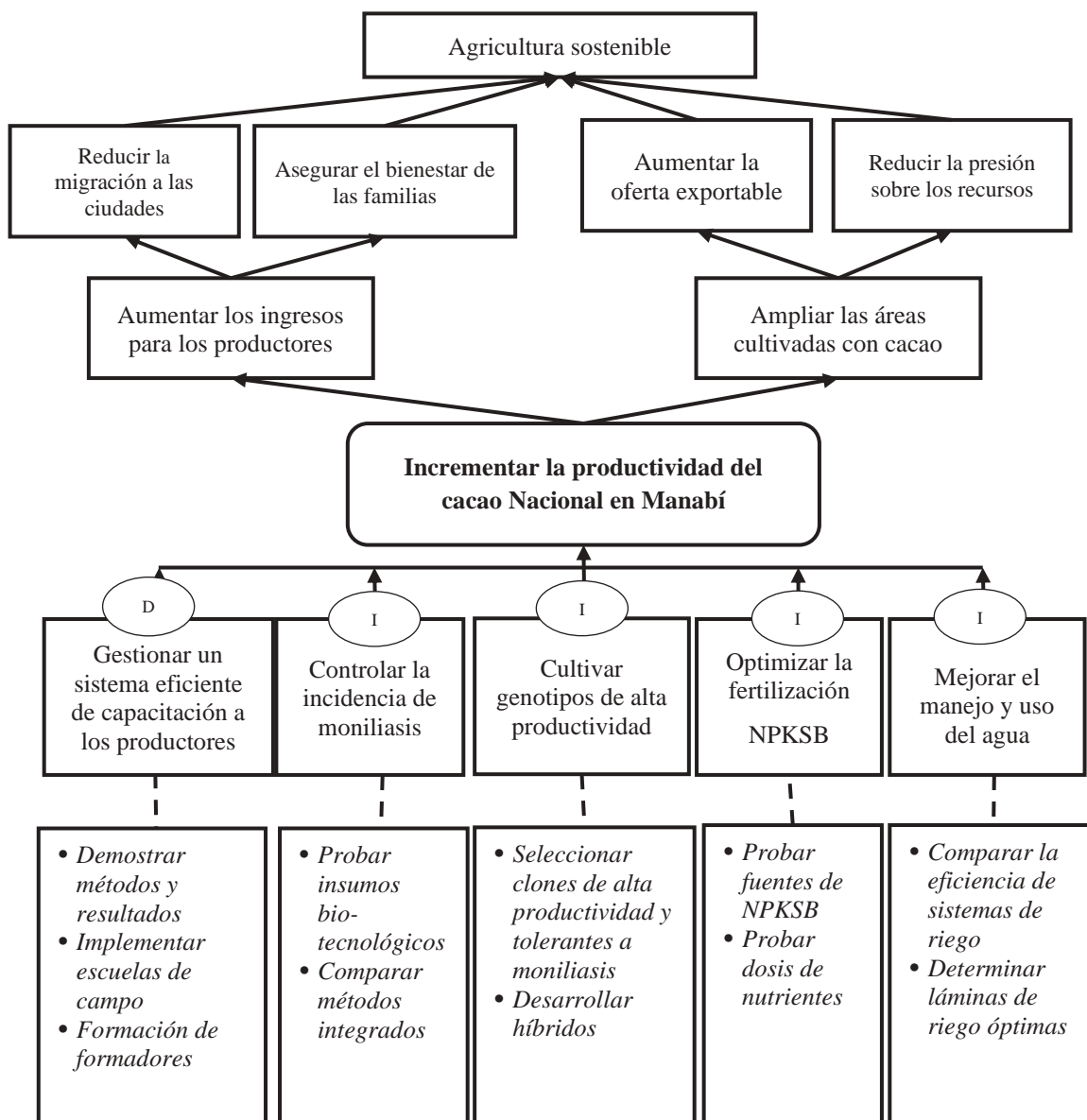
- Gestionar un sistema eficiente de capacitación
- Controlar la incidencia de moniliasis
- Cultivar genotipos de alta productividad
- Optimizar la fertilización con los nutrientes N-P-K-S-B
- Mejorar el manejo del suelo y agua

**Ejercicio 3:** Identificar la “idea del proyecto” que tiene el objetivo de optimizar la fertilización N-P-K-S-B para incrementar la productividad del cacao Nacional en Manabí

Para incrementar la productividad del cacao Nacional en Manabí, resulta imperativo la optimización de la fertilización con potasio (K), azufre (S) y boro (B) en cacaotales de tipo Nacional en la provincia de Manabí”. En estas circunstancias, hay que planear y ejecutar ensayos unifactoriales o multifactoriales en los distintos “dominios de recomendación”.



La “idea del proyecto” mencionado, se resume en la Figura 10. La priorización de los territorios se basa en criterios como: importancia del cultivo, la disponibilidad de recursos y los acuerdos colaborativos para la ejecución En Manabí, el cacao Nacional es muy importante, por el volumen de producción anual, en varios cantones de las zonas norte, centro y sur. ¿Dónde debe realizarse la investigación? ¿En una comunidad?, ¿en dos cantones?, ¿Quién financia?



Nota: D = Proyecto de Desarrollo y capacitación, I = proyecto de investigación

Figura 10. Construcción de la idea del proyecto de investigación alineado a un programa

Al priorizarse la zona centro de Manabí, por ejemplo, ¿Qué territorio se prioriza?, ¿Se investiga en los cantones Santa Ana y Portoviejo?, ¿Se prioriza una parroquia?. ¿Con qué criterios? La `priorización se extiende hasta la definición del recinto y la finca donde se debe hacer la investigación.



1.3.3.3 ENSAYO DE INVESTIGACIÓN

Un ensayo trata de manera rigurosa un problema específico y tiene como instrumento de planificación el “protocolo”, que a nivel de la investigación formativa son los proyectos de tesis y a nivel de la investigación generativa se trata de los perfiles (versión abreviada de la propuesta) o del protocolo del ensayo (parte de los proyectos y programas).

En la formulación de un protocolo de investigación, se considera fundamental cuidar las relaciones verbales y contenidos del título, objetivo general, objetivos específicos, hipótesis de investigación y resultados esperados. La redacción de la metodología debe detallar, en orden, la manera de lograr cada uno de los objetivos específicos. Por ejemplo, en el protocolo que trata como problema central la “baja productividad” en el cultivo de cacao Nacional, en Manabí, donde se destaca como causa la “deficiencia de nitrógeno en el suelo” el título podría ser: “Optimización de la fertilización nitrogenada para incrementar la productividad del cacao Nacional, en Manabí”.

Los conceptos deben estar interrelacionados, como en el ejemplo de la Figura 11. El término optimización se asocia al verbo optimizar (nivel predictivo). Como función se plantea: “la productividad (Y) depende de las dosis de nitrógeno (X)”. Esto se representa como el modelo:  $[Y = f(X)]$ . La fertilización nitrogenada es la acción requerida para solucionar el problema. Las dosis de nitrógeno constituyen los niveles del factor en estudio.

El problema “deficiencias de nitrógeno en el suelo” indica es prioridad probar fuentes nitrogenadas: orgánicas, químicas o mezclas; además, de diferentes dosis y formas de aplicación. La variable de respuesta es la producción, el objeto de estudio es cacao Nacional y el dominio de recomendación a donde se puede extrapolar los resultados, podría ser la zona cacaotera de la parroquia Río Chico, Portoviejo, Manabí.

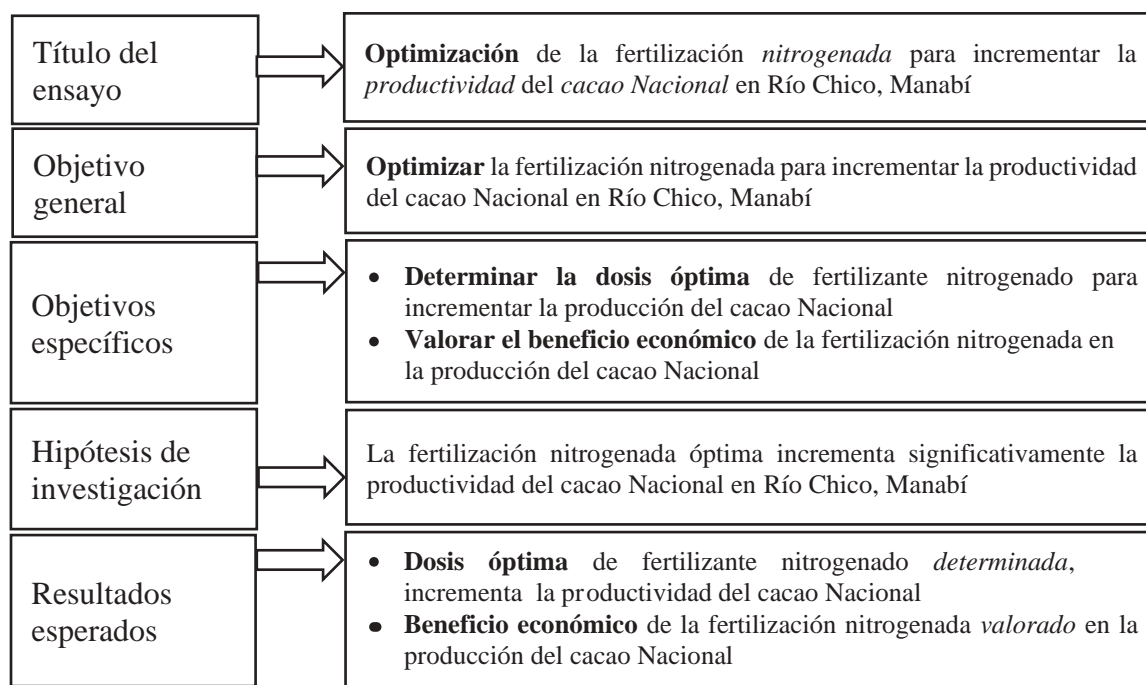


Figura 11. Relaciones entre título, objetivos, hipótesis y resultados esperados del proyecto

### 1.3.4 ESTRUCTURA DE LA PROPUESTA DE INVESTIGACIÓN

La propuesta de investigación tiene una estructura jerárquica, estrechamente ligada a la problemática (árbol de problemas) y los objetivos (árbol de objetivos), enmarcada en las líneas de la Secretaría Nacional de Ciencia, Tecnología e Innovación [SENESCYT], en los objetivos de desarrollo sostenible de la Agenda 2030 de la Organización de las Naciones Unidas [ONU] y en las políticas institucionales. La propuesta, por lo tanto, puede ser elaborada en los niveles de plan, de programa, de proyecto o de un protocolo para un estudio concreto.

La propuesta de investigación, en cualquier nivel, debe contener los siguientes componentes comunes: Introducción, marco teórico, marco metodológico, marco administrativo, referencias bibliográficas (Figura 12) y anexos.

**Introducción.**- Se detalla problemática, justificación, antecedentes investigativos, objetivos e hipótesis de investigación. ¿El problema identificado realmente no tiene alternativas de solución?, ¿Hay oportunidades de cambiar la situación problemática?.

**Marco teórico.**- Se presentan las teorías, los conceptos, los modelos relevantes, las variables y los métodos usados por otros equipos en temáticas similares. Se justifican las metodologías, los registros de variables y las transformaciones de datos, las técnicas de análisis que han usado otros investigadores, además se identifican las controversias existentes y las contribuciones de otros grupos de investigación. La literatura citada debe cumplir los principios de pertinencia y vigencia.

La pertinencia se refiere a la utilidad práctica de la información. Por ejemplo, al analizar la baja productividad del maíz en la provincia de Manabí, las citas y referencias prevalentemente deberán referirse a Manabí o aproximarse a sus condiciones agrosociales, para tener pertinencia. La vigencia está en función de la validez de la teoría. Por ejemplo, está vigente la teoría mendeliana (Mendel, 1865, pp. 3-47), así como el análisis de la varianza propuesto por Ronald Fisher (1926), mientras que las estadísticas de producción agrícola de hace dos años atrás, de cualquier territorio, ya no son vigentes.

**Marco metodológico.**- Es el indicativo del nivel investigativo, de las características del campo experimental (suelo, clima y fisiografía) o del escenario agrosocial (característica de la población, comunidad o muestra), el factor en estudio y niveles (lo contrario de factor limitante), el diseño de tratamientos, el diseño experimental, las variables experimentales, el registro de los datos (con sus escalas de medición), las técnicas del análisis estadístico y del análisis económico, así como el detalle del manejo del experimento.

**Marco administrativo** .- Es el conjunto de condiciones que aseguran el logro de los resultados. Entre otros elementos, comprende el cronograma de actividades, el presupuesto, el financiamiento, las alianzas, el plan de manejo ambiental (si fuese necesario) y las certificaciones pertinentes.

**Referencias bibliográficas.**- Las citas y las referencias bibliográficas deben estar articuladas, con base en las directrices de la institución, citando fuentes primarias y secundarias; así como la vigencia y la pertinencia de la literatura científica. La información debe ser local, regional, nacional e internacional.

**Anexos.**- Son elementos complementarios que contribuyen con información clave para la ejecución de la investigación como: Planos, formatos de registro de datos, modelos y/o fotos.

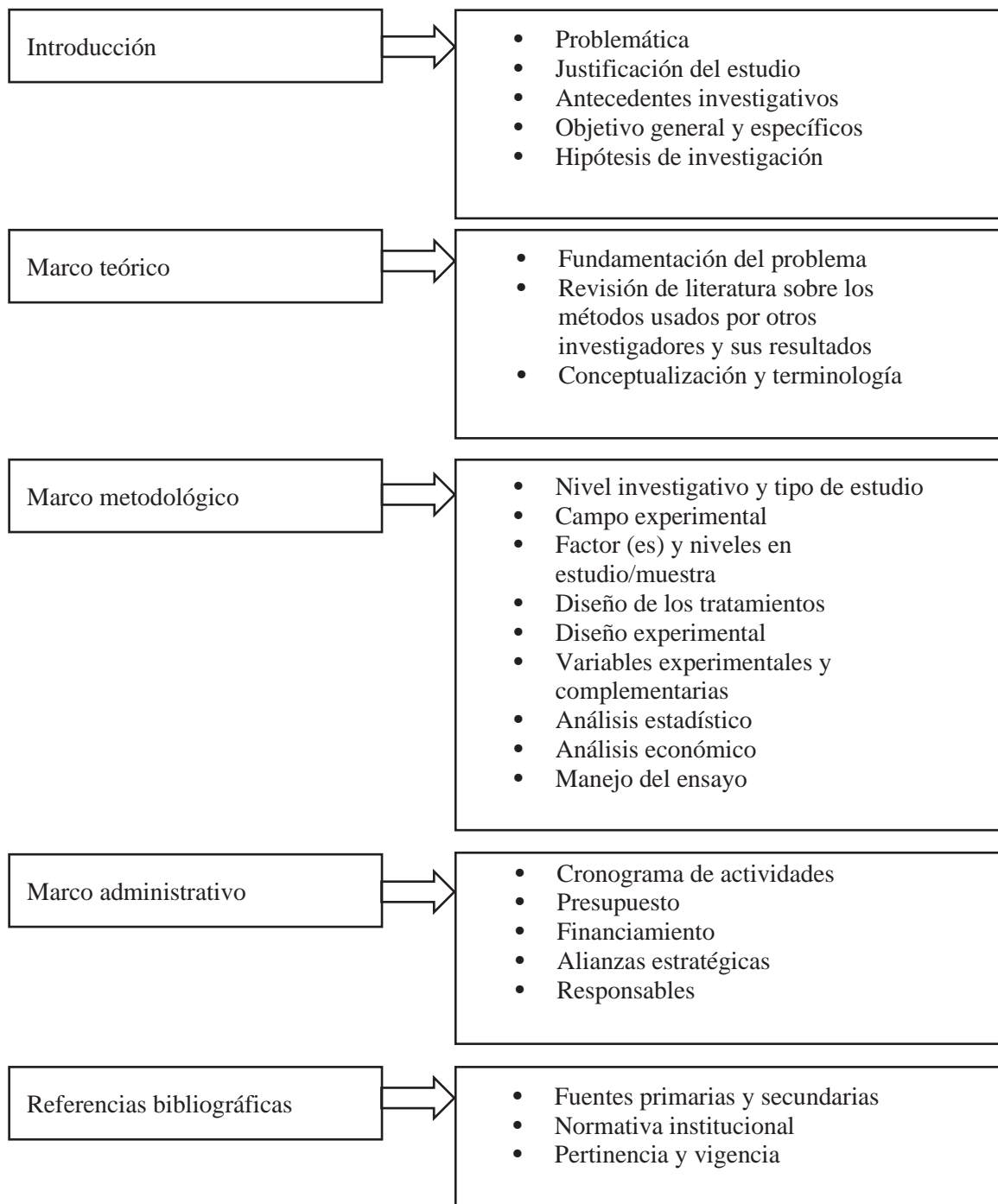


Figura 12. Estructura referencial de una propuesta de investigación agropecuaria

## CAPÍTULO II: POBLACIONES Y MUESTRAS

La población o universo, en estadística<sup>3</sup>, es el conjunto de individuos, elementos, objetos o eventos que tienen un valor representativo para la sociedad o un segmento de ella. Algunos ejemplos: habitantes del Ecuador, población afroamericana de Esmeraldas, población montuvia de Los Ríos, estudiantes de una carrera, alumnos del paralelo, insectos plaga del maíz, larvas por fruto en una guayaba, vacas lecheras en un establo, árboles de un bosque y número de peces en un estanque, entre otros.

### 2.1 CLASIFICACIÓN DE LAS POBLACIONES

Las poblaciones pueden clasificarse de distintas formas: reales y estadísticas, finitas e infinitas, accesible y no accesible, actual y futura.

#### 2.1.1 POBLACIÓN REAL Y POBLACIÓN ESTADÍSTICA

**Población real.**- Es el conjunto de elementos, individuos, objetos o sucesos que tienen un valor representativo para la sociedad o un segmento de ella. Los árboles son los elementos de un bosque (población real), los peces son elementos de un estanque piscícola, las plantas de maíz son elementos de un maizal, cada alumno es parte de una población estudiantil, una vaca es un elemento de un hato ganadero, una lata de conserva es parte de la producción diaria en una planta agroindustrial.

**Población estadística.**- Es el conjunto de valores medidos y asignados a todos los elementos de la población real, para cada variable de interés. Ejemplos: la altura (m) de todos los árboles del bosque (población real), el peso (g) de todos los peces de un estanque, el número de hojas de las plantas de maíz en un lote de cultivo, la edad en años de todos los alumnos de la universidad, la producción diaria de leche (litros) de las vacas de un hato, los pesos netos (g) de las latas producidas diariamente en una planta agroindustrial.

#### 2.1.2 POBLACIÓN FINITA Y POBLACIÓN INFINITA

**Población infinita.**- Se identifica como tal a las poblaciones muy grandes, donde los elementos, individuos u objetos son reales, pero identificarlos y medirlos implicaría un costo elevado. Ejemplo: un bosque de 10 o más hectáreas (la medición de altura de fuste en todos los árboles resultaría muy costoso).

**Población finita.**- Es el conjunto limitado de elementos, individuos u objetos con condiciones físicas, químicas, biológicas o de otra naturaleza comunes a un objeto de estudio. Ejemplos: las vacas lecheras de un establo para ordeño, las macetas con plantas ornamentales en una vivienda, las cajas *petri* de un experimento en laboratorio y los alumnos de un aula de clase.

#### 2.1.3 POBLACIÓN ACCESIBLE Y POBLACIÓN NO ACCESIBLE

**Población accesible.**- Es el conjunto de elementos, individuos u objetos de la población real que pueden observarse y medirse de manera directa. Ejemplos: plántulas de un vivero de palma africana, estudiantes de un paralelo, cajas *petri* en el mesón de laboratorio, jefes de familia de una comunidad.

---

<sup>3</sup> En el Anexo 1 se indica un glosario básico de estadística

**Población no accesible.**- Conjunto de elementos, individuos u objetos donde no existe la posibilidad de observarlos y medirlos directamente con facilidad. Ejemplos: Los cóndores en la región andina, los peces del río Guayas. La estadística ha desarrollado técnicas para el conteo y mediciones en estas poblaciones.

#### 2.1.4 POBLACIÓN ACTUAL Y POBLACIÓN FUTURA

**Población actual.**- Son los elementos o individuos que se pueden observar y medir, en el presente, en un momento y espacio concretos. Ejemplos: número de “pajuelas” importadas de una raza de ganado, número de semillas de un híbrido élite de maíz introducido y número de niños huérfanos en el orfanato de la ciudad.

**Población futura.**- Son los elementos o individuos (no objetos ni sucesos) que se contarían y medirían bajo determinadas circunstancias, proyectadas en un futuro. Por ejemplo: se introduce seis semillas de un híbrido élite de maíz desde Asia (población actual), que se compara con otros genotipos, donde una de semilla introducida podría dar origen a una variedad seleccionada y en cinco años podría haber miles de hectáreas con millones de semillas de maíz (población futura).

## 2.2 PARÁMETROS Y ESTADÍSTICOS

El parámetro es una medida concreta referida a una variable que caracteriza a una población a través del censo y el estadístico es una medida que caracteriza a una muestra representativa.

En las poblaciones infinitas, un censo resulta muy costoso y ocupa mucho tiempo, por lo tanto, se recurre al muestreo. En la muestra se realiza calculan los estadísticos o estadígrafos a los cuales se incorpora la probabilidad para estimar los parámetros de la población. En la estadística inferencial, las medidas de mayor importancia en las muestras son los estadísticos media ( $\bar{Y}$ ) y desviación estándar ( $S$ ) a partir de los cuales se estima los parámetros poblacionales: media ( $\mu$ ) y desviación estándar ( $\sigma$ ).

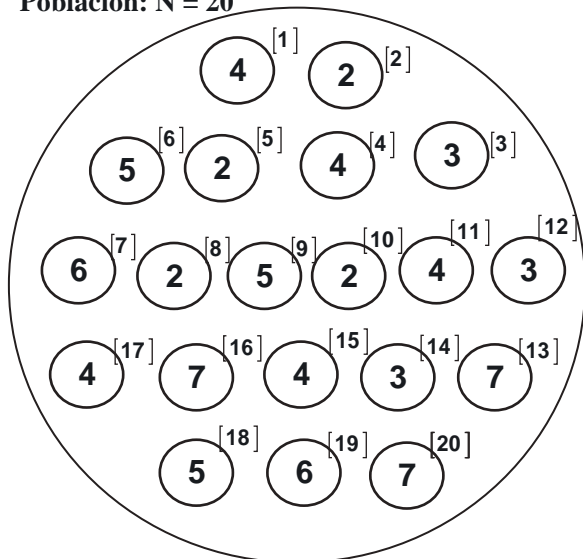
$$\begin{array}{l} \bar{Y} \text{ estima a } \rightarrow \mu_Y \\ S_Y \text{ estima a } \rightarrow \sigma_Y \end{array}$$

El teorema de límite central indica que si se seleccionan muestras al azar de una población determinada, conforme aumenta el tamaño de la muestra, la distribución se aproxima a la curva normal. En todos los casos, el objetivo del investigador es estimar los parámetros de la población, donde solo hay dos posibilidades: realizar el censo o el muestreo.

**Censo.**- Es el registro individual de los datos de interés, en todos los elementos, individuos u objetos de una población. Los datos de un censo, por ejemplo, del peso ( $g.planta^{-1}$ ) genera un parámetro que constituye un valor característico, único y constante. Los parámetros poblacionales más importantes son: media ( $\mu$ ) y desviación estándar ( $\sigma$ ).

**Muestreo.**- Es un procedimiento por medio del cual se toma una muestra representativa de la población. En el Gráfico 1, se indica la serie de datos de una población ( $N = 20$ ) y el cálculo de los parámetros media ( $\mu$ ) y desviación estándar ( $\sigma$ ). En el Gráfico 2, se exponen los resultados de los estadísticos calculados en dos muestras de tamaño  $n = 4$ .

Población: N = 20

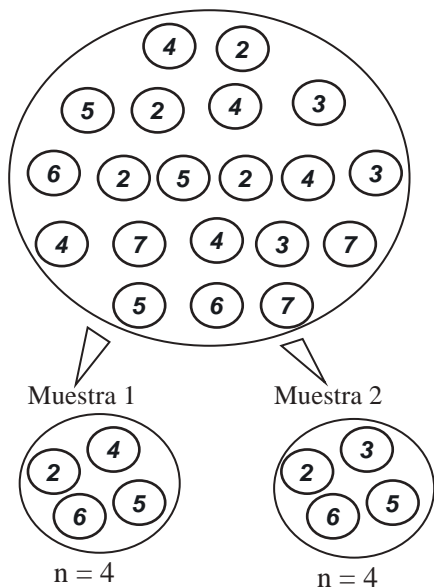


Población	Variable (Y)
1	4
2	2
3	3
4	4
5	2
6	5
7	6
8	2
9	5
10	2
11	3
12	4
13	7
14	3
15	4
16	7
17	4
18	5
19	6
20	7

Tamaño	N	20
Media	$\mu$	4,25
Desviación estándar	$\sigma$	1,67
Valor máximo	Máx.	7
Valor mínimo	Mín.	2

Gráfico 1. Población finita N = 20 y los parámetros media y desviación estándar

Población



	Muestra 1	Característica cuantitativa (Y)
	1	2
	2	4
	3	5
	4	6
Tamaño	n	4
Media	$\bar{Y}$	4,25
Desviación estándar	S	1,71
Valor máximo	Máximo	6
Valor mínimo	Mínimo	2

	Muestra 2	Característica cuantitativa (Y)
	1	3
	2	5
	3	6
	4	2
Tamaño	n	4
Media	$\bar{Y}$	4,00
Desviación estándar	S	1,83
Valor máximo	Máximo	6
Valor mínimo	Mínimo	2

Gráfico 2. Dos muestras aleatorias de n = 4 y los principales estadísticos

Las principales fórmulas para calcular parámetros y estadígrafos de las cuales pueden derivarse otras de interés, se indican en el Cuadro 3. Los parámetros se calculan registrando los datos en todos los elementos de la población (N), esta actividad llamada censo, mientras que los estadísticos se calculan en series de datos provenientes de muestras.

Cuadro 3. Fórmulas más usadas para el cálculo de parámetros y estadígrafos

Parámetros y estadísticos	Fórmulas para calcular parámetros	Fórmulas para calcular estadísticos	
Media simple	$\mu_s = \frac{(Y_{\text{máximo}} + Y_{\text{mínimo}})}{2}$	$\bar{Y}_s = \frac{(Y_{\text{máximo}} + Y_{\text{mínimo}})}{2}$	
Media aritmética	$\mu = \frac{\sum Y_i}{N}$	$\bar{Y} = \sum \frac{Y_i}{n}$	
Media ponderada	$\mu_p = \frac{\sum f_i Y_i}{\sum f_i}$	$\bar{Y}_p = \frac{\sum f_i Y_i}{\sum f_i}$	
Media geométrica	$\mu_G = \sqrt[n]{Y_1 Y_2 \dots Y_N}$	$\bar{Y}_G = \sqrt[n]{Y_1 Y_2 \dots Y_n}$	
Media armónica	$\mu_H = \frac{n}{\sum (\frac{1}{Y_i})}$	$\bar{Y}_H = \frac{n}{\sum (\frac{1}{Y_i})}$	
Varianza	$\sigma^2 = \frac{\sum Y^2 - \frac{(\sum Y)^2}{N}}{N}$	$S^2 = \frac{\sum Y^2 - \frac{(\sum Y)^2}{n}}{n - 1}$	$S^2 = \frac{\sum (Y - \bar{Y})^2}{n - 1}$
Desviación estándar	$\sigma = \sqrt{\frac{\sum Y^2 - \frac{(\sum Y)^2}{N}}{N}}$	$S = \sqrt{\frac{\sum Y^2 - \frac{(\sum Y)^2}{n}}{n - 1}}$	$S = \sqrt{S^2}$

### 2.3 TAMAÑO ÓPTIMO DE MUESTRA

La muestra es una parte representativa de la población. La determinación del tamaño óptimo permite un balance adecuado entre el costo del muestreo y la precisión obtenida. Las muestras de tamaño pequeño, debajo del óptimo, generan un incremento de la distancia entre los valores del estadístico y del parámetro (Aguilar, 2005, p. 336). La determinación del tamaño óptimo de muestra, en las poblaciones infinitas y finitas, se realiza en función de la variable de mayor interés (p.e.: frutos.árbol<sup>-1</sup> o producción en gramos.planta<sup>-1</sup>), no es recomendable hacerlo para todas las variables.

Una vez calculado el tamaño óptimo de muestra para la variable de mayor interés se aplica a todas las otras variables en estudio. En el muestreo, el incluir más sujetos de estudio, no es sinónimo de mejorar el estudio (Aguilar, 2005, p. 336). La determinación del tamaño óptimo de muestra, es clave para reducir los errores tipo I y tipo II (Badii et al., 2021, p. 3). Los factores que condicionan el tamaño de muestra son de tipo logístico o estadístico. En lo logístico hay que tomar en cuenta el costo de la recolección de datos y el tiempo requerido. En lo estadístico hay que considerar la probabilidad de equivocarse en la decisión (riesgo) y el error de muestreo que es el valor máximo que se puede admitir como variación de un estadístico.

**Hipótesis estadística.**- Se formula en términos de hipótesis nula (H<sub>0</sub>) e hipótesis alternativa (H<sub>1</sub>). La H<sub>0</sub> = representa nulidad de las diferencias. H<sub>0</sub>: μ<sub>2</sub> - μ<sub>1</sub> = 0, que se propone como “no hay diferencia estadística significativa” (ns), a un nivel de significación (α), dado. Si se probara, por ejemplo, tres variedades de tomate, H<sub>0</sub>: V<sub>1</sub> = V<sub>2</sub> = V<sub>3</sub>. En cambio, la hipótesis



alternativa,  $H_1$ : Al menos una de las variedades es estadísticamente diferente con un nivel de confianza concreto  $(1 - \alpha)$ .

Cabe destacar que la hipótesis de investigación, conceptualmente difiere de la hipótesis estadística, la primera es una conjetura del potencial logro y la segunda es formulación de los probables resultados de la prueba estadística, sea  $F$ ,  $t$ ,  $\chi^2$  u otra.

**Nivel de significación ( $\alpha$ ).**- Es la representación del riesgo de equivocarse en la decisión estadística y se representa por  $\alpha$ . En la investigación agropecuaria se aceptan los niveles de significación:  $\alpha = 0,05$  y  $0,01$ . En la investigación médica las probabilidades de equivocarse tienden a ser muy bajas como:  $\alpha = 0,00001$  (1 en 100.000 casos).

El nivel de significación  $\alpha$  corresponde al error tipo I definida como la probabilidad de rechazar  $H_0$  cuando en realidad es verdadera (falso positivo). Los valores críticos de  $Z$  asociados a la probabilidad de equivocarse en las decisiones, de mayor uso en la investigación agropecuaria, agroindustrial y ambiental, para las pruebas de una cola y de dos colas, se exponen en el Cuadro 4.

**Nivel de confianza ( $1 - \alpha$ ).**- Es el grado de confianza en que el valor verdadero del parámetro poblacional se encuentre dentro de la muestra. El nivel de confianza es complementario al nivel de significación:  $1 - \alpha$  (Aguilar, 2005, p. 337).

**Errores tipo II ( $\beta$ ).**- El error tipo II es la probabilidad de aceptar  $H_0$  cuando realmente es falsa (falso negativo).

**Poder estadístico.**- El poder estadístico, también conocido como potencia de la prueba, es la probabilidad de detectar diferencias significativas cuando realmente existen  $(1 - \beta)$ .

**Variabilidad.**- En función de la naturaleza de la población, las series de datos mostrarán mayor o menor dispersión, que se valoran a través de los estadígrafos como varianza, desviación estándar, rango, error típico, variación relativa y coeficiente de variación.

**Dato perdido.**- Al realizar muestreos o experimentos, puede haber valores perdidos, ya sea a causa de factores externos o porque resultaron atípicos (“fuera de tipo” u “outliers”). En un diseño completamente al azar (DCA), a perder una o más unidades experimentales, no es necesario calcular el dato faltante, se podría analizar el experimento como un DCA desbalanceado.

En los experimentos con diseño en bloques al azar (BA) o en los diseños en Cuadrado latino (CL), que son de tipo balanceado, hay que calcular el “dato de la parcela perdida”, que frecuentemente se identifica como “parcela perdida”. En este proceso debe considerarse que para calcular la varianza de un tratamiento se requiere por los menos de dos datos.

Cuadro 4. Valores críticos de  $Z$  asociados a los niveles de significación más usados en bioestadística

Nivel de significación		Nivel de confianza		Valores críticos de $Z$ asociados al nivel de significación	
$\alpha$	$\alpha$ (%)	$1 - \alpha$	$(100 - \alpha \%)$	Para dos colas	Para una cola
0,100	10	0,900	90	1,65	1,28
0,050	5	0,950	95	1,96	1,65
0,010	1	0,990	99	2,58	2,33
0,001	0,1	0,999	99,9	3,29	--



### 2.3.1 TAMAÑO ÓPTIMO EN POBLACIONES INFINITAS

En estudios cuantitativos, cuando se tienen poblaciones infinitas o finitas, se requiere hacer un muestreo inicial ( $n_0$ ) para conocer la varianza inicial  $S_0^2$ , decisión que la toma el investigador en base a su experiencia en muestreo. La fórmula para calcular el tamaño óptimo de muestra en poblaciones infinitas con datos cuantitativos (Aguilar, 2005, p. 336) es la siguientes:

$$n_{\text{óptimo}} = \frac{Z^2(S_0^2)}{d^2} \quad [1]$$

Dónde:

$n_{\text{óptimo}}$  = Tamaño óptimo de muestra

Z = Valor crítico de Z asociado al nivel de significación  $\alpha$

$\alpha$  = Riesgo o probabilidad de equivocarse en la decisión

$\alpha = 0,05 \sim 5 \%$  para pruebas de una cola y  $\frac{\alpha}{2} = 0,025 \sim 2,5 \%$  para pruebas de dos colas

Valor crítico de Z asociado a la prueba de dos colas: Cuando  $\frac{\alpha}{2} = 0,025 \sim 2,5 \%$   $\rightarrow Z = 1,96$

d = Error máximo aceptado por el investigador en este muestreo. Frecuentemente se fija como proporción de la media, aunque puede usarse otros criterios.

#### Ejercicio 4:

Calcular el tamaño óptimo de muestra, con el 95 % de confianza, para la variable peso de mazorcas de maíz. La muestra inicial,  $n_0 = 20$ , promedio  $\bar{Y} = 180$  g.mazorca<sup>-1</sup>, desviación estándar,  $S = 35$  g.mazorca<sup>-1</sup> y error máximo aceptable,  $d = \pm 10$  gramos.

#### Datos:

Variable: Peso de la mazorca (g)

Muestra inicial:  $n_0 = 20$  mazorcas

Desviación estándar:  $S = 35$  g.mazorca<sup>-1</sup>

Error:  $d = \pm 10$  g.mazorca<sup>-1</sup>

Riesgo:  $\alpha = 0,05$  para una cola y  $\frac{\alpha}{2} = 0,025$  para dos colas

Valor crítico de Z asociado a la prueba de dos colas:  $Z = 1,96$

#### Fórmula y análisis:

$$n_{\text{óptimo}} = \frac{Z^2(S_0^2)}{d^2} \rightarrow n_{\text{óptimo}} = \frac{(1,96)^2(35)^2}{(10)^2} = \frac{(3,84)(1225)}{100} = 47 \text{ mazorcas}$$

$$n_{\text{óptimo}} = 47 \text{ mazorcas}$$

#### Muestreo complementario:

El tamaño óptimo de muestra es 47 mazorcas. En el muestreo inicial se tomó 20 mazorcas, Para cumplir las condiciones de muestreo hace falta completar el muestreo

$$n_0 = 20 \text{ mazorcas}$$

$$n_{\text{óptimo}} = 47 \text{ mazorcas}$$

$$\text{Muestra complementaria} = n_{\text{óptima}} - n_{\text{inicial}} = 47 - 20 = 27 \text{ mazorcas}$$

#### Decisión:

El tamaño óptimo de muestra es 47 mazorcas. Hace falta tomar una muestra complementaria de 27 mazorcas, para cumplir las condiciones de riesgo y error.

**Ejercicio 5:**

Calcular el tamaño óptimo de muestra para la variable rendimiento en el cultivo de cacao (kg.hectárea<sup>-1</sup>) en base a una encuesta aplicada a 30 productores ( $n_0 = 30$ ). El rendimiento promedio fue  $\bar{Y} = 265$  kg.hectárea<sup>-1</sup> y una desviación estándar  $S = 42$  kg.hectárea<sup>-1</sup>. Se desea una confianza del 95 %, por tanto  $\alpha = 0,05 \rightarrow 5$  % de probabilidad de equivocarse. Se fija como un error máximo aceptable,  $d = \pm 15$  kg.

Variable: Rendimiento (kg)

Muestra inicial:  $n_0 = 30$  productores

Desviación estándar:  $S = 42$  kg.hectárea<sup>-1</sup>

Error:  $d = \pm 15$  kg.hectárea<sup>-1</sup>

Riesgo:  $\alpha = 0,05$  para una cola y  $\frac{\alpha}{2} = 0,025$  para dos colas

Valor crítico de Z asociado a la prueba de dos colas:  $Z = 1,96$

**Fórmula y análisis:**

$$n_{\text{óptimo}} = \frac{Z^2(S_0^2)}{d^2} \rightarrow n_{\text{óptimo}} = \frac{(1,96)^2 (42)^2}{(15)^2} = \frac{(3,84)(1764)}{(225)} = 30 \text{ productores}$$

**Muestreo complementario:**

En el muestreo inicial se tomó información del rendimiento promedio de 30 productores y tamaño óptimo de muestra es 30 productores, por lo tanto, no hace falta tomar una muestra complementaria. Las condiciones de riesgo y error prefijadas por el investigador están cumplidas.

**Decisión:**

El tamaño óptimo de muestra es 30 productores, con un 95% de confianza y un error de 15 kg.hectárea<sup>-1</sup>.

**2.3.2 TAMAÑO ÓPTIMO EN POBLACIONES FINITAS**

Una población finita (N) es aquella que tiene un número limitado de elementos, individuos u objetos. Por ejemplo: número de macetas en un invernadero, número de animales en un establo, número de cajas petri en un laboratorio, número de estudiantes en un aula de clase, o número familias de una pequeña comunidad.

El tamaño óptimo de muestra en poblaciones finitas se calcula usando la fórmula de poblaciones infinitas multiplicado por un coeficiente de “corrección por población finita”.

La fórmula para calcular el tamaño de muestra en poblaciones finitas es la siguiente:

$$n_{\text{óptimo}} = \frac{Z^2(S_0^2)}{d^2} \left(1 - \frac{n_0}{N}\right) \quad [2]$$

Dónde:

$n_{\text{óptimo}}$  = Tamaño óptimo de muestra

$Z = 1,96$  es el valor crítico asociado a la prueba de dos colas y  $\alpha = 0,05$

$S_0^2$  = Varianza del muestreo inicial

$d$  = Error máximo aceptado por el investigador

$n_0$  = Tamaño de la muestra inicial

$N$  = Tamaño de la población finita

**Ejercicio 6:**

Calcular el tamaño óptimo de muestra para la variable rendimiento en el cultivo de cacao, en base a la información de una encuesta a 20 productores ( $n_0 = 20$ ) de una comunidad rural donde hay 100 productores ( $N = 100$ ). El rendimiento promedio fue  $\bar{Y} = 265 \text{ kg.ha}^{-1}$  y la desviación estándar,  $S = 42 \text{ kg.ha}^{-1}$ . El investigador define una confianza en la decisión del 95 % ( $\alpha = 0,05$ ) y como error máximo aceptable,  $d = \pm 15 \text{ kg}$ .

Variable: Rendimiento (kg)

Muestra inicial:  $n_0 = 20$  productores

Desviación estándar:  $S = 42 \text{ kg.hectárea}^{-1}$

Error:  $d = \pm 15 \text{ kg.hectárea}^{-1}$

Riesgo:  $\alpha = 0,05$  para una cola y  $\frac{\alpha}{2} = 0,025$  para dos colas

Valor crítico de Z asociado a la prueba de dos colas:  $Z = 1,96$

$N = 100$  productores

**Fórmula y análisis:**

$$n_{\text{óptimo}} = \frac{Z^2(S_0^2)}{d^2} \left(1 - \frac{n_0}{N}\right) \rightarrow n_{\text{óptimo}} = \frac{(1,96)^2 (42)^2}{(15)^2} \left(1 - \frac{20}{100}\right) = \frac{(3,84)(1764)}{(225)} (1 - 0,20)$$

$$n_{\text{óptimo}} = 30 (0,8) = 24 \text{ productores}$$

**Muestreo complementario:**

En el muestreo inicial se tomó información de 20 productores, si el tamaño óptimo calculado es 24, significa que hace falta tomar una muestra complementaria de 4 productores.

**Decisión:**

El tamaño óptimo de muestra es 24 productores con un 95% de confianza y un error de  $15 \text{ kg.hectárea}^{-1}$ .

**2.3.3 TAMAÑO ÓPTIMO DE MUESTRA PARA CONTEOS**

El cálculo de tamaño óptimo de muestra es distinto en las poblaciones  $N$ , donde se conocen el total de unidades en estudio (Aguilar, 2005, p. 336), principalmente cuando se tratan de datos discretos, como: plantas sanas.parcela<sup>-1</sup>, frutos sanos.árbol<sup>-1</sup> u hogares con potencial de adquirir nuevos productos.

La fórmula usada en este caso, es la siguiente:

$$n_{\text{óptimo}} = \frac{N \cdot Z^2 \cdot p \cdot q}{d^2(N - 1) + Z^2 \cdot p \cdot q} \quad [3]$$

Dónde:

$d$  = Error máximo aceptado por el investigador

$N$  = Tamaño de la población

$p$  = Probabilidad de respuestas afirmativas ( $p = 0,5$ )

$q$  = Probabilidad complementaria ( $1 - p = 0,5$ )

Los niveles de significación y los valores de probabilidad pueden variar en función de la naturaleza de la investigación.

**Ejercicio 7:**

Determinar el número de hogares a ser encuestados (tamaño de muestra) para tener una opinión sobre un mercado potencial de 10000 hogares (N = 10000) donde se quiere introducir un producto alimenticio en forma de conserva. El investigador prefija el nivel de significación en 5% y como error máximo aceptable ± 5 % alrededor de la media (d = 0,05)

**Análisis:**

$$n_{\text{óptimo}} = \frac{N \cdot Z_{\alpha/2}^2 \cdot p \cdot q}{d^2(N-1) + Z_{\alpha/2}^2 \cdot p \cdot q} \rightarrow n_{\text{óptimo}} = \frac{10000 (1,96^2)(0,05)(0,05)}{(0,05)^2(10000-1) + 1,96^2 \cdot (0,05)(0,05)}$$

$$n_{\text{óptimo}} = \frac{10000(3,8416)(0,5)(0,5)}{0,0025(9999) + 3,8416 (0,05)(0,05)} = 370 \text{ hogares}$$

**Decisión:**

El tamaño óptimo de muestra es 370 hogares con una confianza del 95 %.

**2.4 MÉTODOS DE MUESTREO**

Los métodos de muestreo pueden ser no probabilísticos o probabilísticos. Los dos métodos son aplicables en la investigación científica, dependiendo de las circunstancias y objetivos del estudio (Gulland, 1966).

**2.4.1 MUESTREO NO PROBABILÍSTICO**

El muestreo no probabilístico de uso en la investigación se clasifica: por conveniencia, por juicio del experto y por cuotas. Otros métodos no probabilísticos referidos en la literatura son de poco interés práctico.

**Muestreo por conveniencia.**- Puede aplicarse por su menor costo o menor tiempo, es una muestra sesgada, no representativa y solo aplicable a estudios exploratorios. Por ejemplo: Si se trata de evaluar la productividad en el cultivo de arroz, se direcciona a tomar la información en las cinco fincas de mayor producción y en otras cinco de menor producción.

**Muestreo por juicio de experto.**- Las observaciones a individuos, objetos o elementos que son de particular interés de un especialista, pueden estar direccionadas, según su buen juicio. Por ejemplo: Se requiere información de fincas que producen cacao del clon EETP 800. El experto recorre un transecto donde se cultiva cacao y hace sus paradas para observar los cacaotales de interés y muestrear. Este método de muestreo es de tipo exploratorio.

**Muestreo por cuotas.**- Se basa en la segmentación del espacio físico, según factores de interés, o estratos, pero sin aleatorización. Se pueden controlar algunas características, pero la representatividad de la muestra, no es segura.

**2.4.2 MUESTREO PROBABILÍSTICO**

Una muestra, para ser una parte representativa de la población biológica o no biológica, finita o infinita, debe cumplir las condiciones de aleatoriedad, independencia y tamaño óptimo.

**Aleatoriedad.**- Todos los elementos, individuos u objetos de una población tienen la misma oportunidad de ser considerados en la muestra, ese es el principio de la aleatoriedad.

**Independencia.**- Se refiere al principio de evitar los sesgos de cualquier tipo, motivados o no, por parte del equipo de investigación.

**Tamaño óptimo de muestra.**- Es el número de individuos, objetos o elementos necesarios para conformar una muestra representativa de la población.

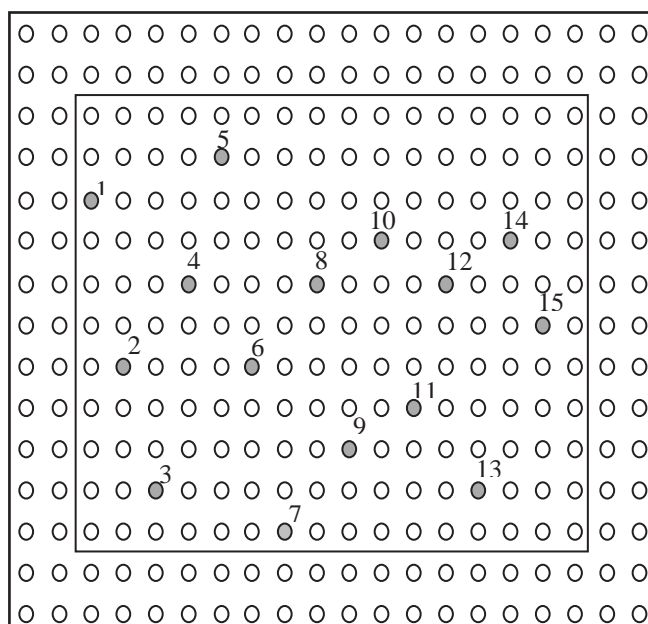
Los métodos de muestreo probabilístico de mayor uso en la investigación del campo agropecuario, forestal, ambiental y agroindustrial son: aleatorio simple, sistemático, estratificado por conglomerados.

#### 2.4.2.1 MUESTREO ALEATORIO SIMPLE

El método de muestreo aleatorio simple es un procedimiento que da, a cada individuo o elemento de la población, la misma oportunidad de conformar la muestra. Por ejemplo, en un cultivo de maíz, de cualquier área, cada planta forma parte de la población y basado en su posición puede asignarse un número.

La uniformidad del lote lo determina la topografía del terreno (p.e.: plano), la fecha de siembra (11-01-2018), el genotipo es único (p.e.: H-603) y las prácticas agrícolas no diferenciadas. La variable de respuesta dependerá del objetivo, como: producción de grano ( $g.planta^{-1}$ ), número de hojas, altura de planta (cm), peso en vivo ( $kg.animal^{-1}$ ).

En la Figura 13, se expone un diagrama de la población ( $N = 200$ ) y del muestreo aleatorio simple ( $n = 15$ ). En cada elemento de la muestra se registrarán los datos de las variables pertinentes, según los objetivos del estudio.



Nota: Población ( $N$ ) = 200 y muestra ( $n$ ) = 15

Figura 13. Diagrama de un muestreo aleatorio simple

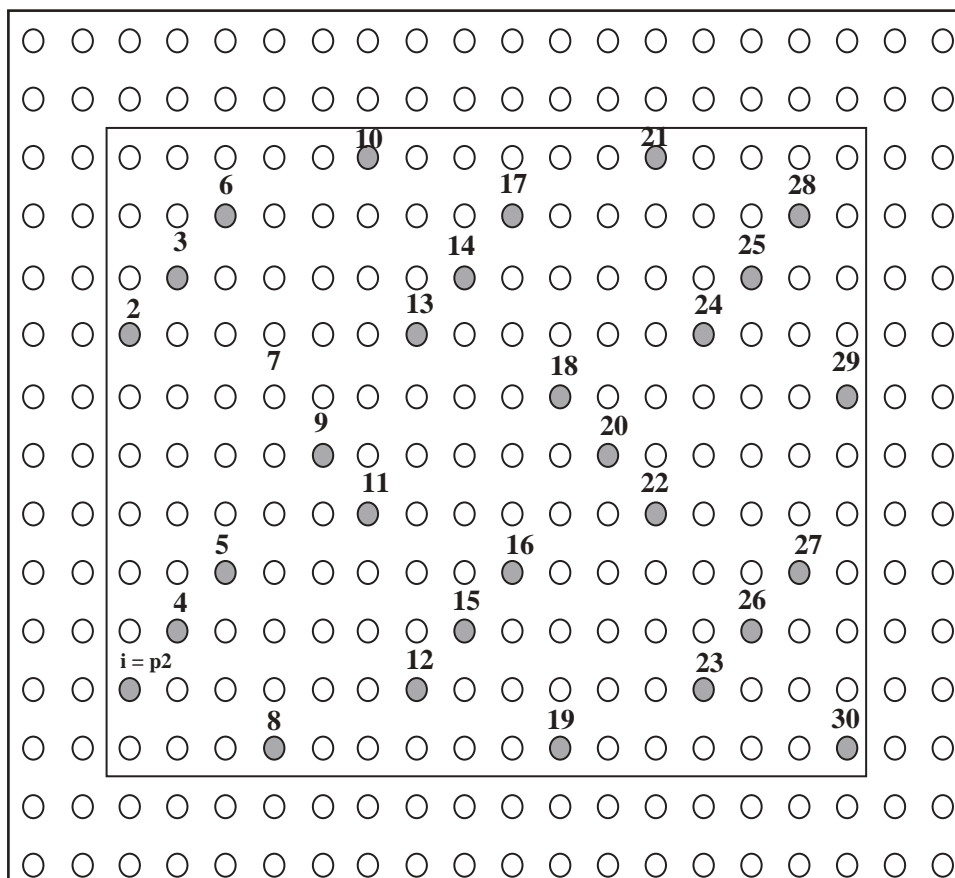
#### 2.4.2.2 MUESTREO SISTEMÁTICO

En toda parcela de investigación agrícola debe haber un borde o parcela circundante del área útil, que no se considera en el muestreo para evitar errores a causa de factores no controlables (Badii et al., 2004, p.37). Esta área útil es el efectivo espacio de muestreo.

El muestreo sistemático es un método aleatorio donde se fija el marco muestral (excluyendo la parcela de borde), basado en el tamaño óptimo de muestra, distribuyendo el muestreo en forma proporcional ( $k = constante$ ), a partir de un punto inicial (i), al azar. El punto inicial (i), corresponde al individuo u objeto donde se toma aleatoriamente el primer

dato (por ejemplo, en la planta 2 dentro de la parcela útil,  $i = 2$ ). A partir de este punto  $i$ , se aplica la constante  $k$  (por ejemplo:  $k = 6$ ) y se toman los datos que corresponde a la muestra de tamaño óptimo predefinida (por ejemplo:  $n = 30$ ). Este caso se indica en la Figura 14.

El valor  $K$  dependerá de la naturaleza de la población y del estudio. Como ejemplos:  $K =$  cada 6 plantas, planta después de cada 30 pasos, dato cada 10 minutos, encuesta cada 10 personas, cada semana, cada mes o trimestrales, según sea el objetivo del estudio.



Nota:  $i =$  planta 2 (dato 1),  $K =$  cada 6 plantas,  $n = 30$  y  $N = 200$

Figura 14. Muestreo sistemático en un campo experimental uniforme

### 2.4.2.3 MUESTREO ESTRATIFICADO

Una población puede mostrar diferentes niveles de heterogeneidad: En la población humana se distinguen dos estratos o subpoblaciones: hombres y mujeres. En este caso, al estudiar la estatura o peso, por ejemplo, evidentemente que habrá una diferenciación. En el caso de la topografía de un terreno puede haber estratos por el grado de pendiente. En un hato ganadero habría estratos por sexo, por raza o por edad. En un bosque, habría estratos por especies forestales, por la gradiente del terreno o por edad.

Una población  $N$  con subpoblaciones  $N_i$  indica que tiene estratos. Si las subpoblaciones  $N_i$  fueran de igual tamaño, la afijación es simple, pues a cada estrato le correspondería un mismo valor. Si las subpoblaciones  $N_i$  tienen distinto tamaño, hay que calcular las proporciones pertinentes a cada estrato y se conoce como afijación proporcional. Cuando se evidencian limitaciones por el costo, la accesibilidad o el tiempo en la toma de datos se debe armonizar entre lo posible y lo deseable, para procurar una afijación óptima.

El muestreo por estratos con afijación proporcional, se calcula usando la fórmula:

$$\text{Muestra por estrato} = (n_{\text{óptimo}}) * \left[ \frac{N_i}{N} \right] \quad [4]$$

Dónde:

$n_{\text{óptimo}}$  = Tamaño óptimo de la muestra para todo el lote (población N)

$N_i$  = Población de un estrato concreto

N = Tamaño de la población

La población humana, en la familia, aula de clase, carrera universitaria, parroquia, cantón, provincia, país o región, tiene dos subpoblaciones: hombres y mujeres. Para estimar los parámetros debe considerarse a los elementos de cada subpoblación como un estrato, por lo tanto, el registro de datos y análisis se realizan por separado. Además, la población humana puede categorizarse por etnias, edades, nacionalidades, creencias religiosas, escolaridad, condiciones de vida, estado de salud u otras, que dependerá del objetivo de estudio. En las poblaciones vegetales y animales, sistemas productivos o poblaciones no biológicas se usan los criterios pertinentes, según las circunstancias.

### Ejercicio 8:

Representar gráficamente a los estudiantes de bioestadística que conforman una población finita ( $N = 20$ ) de los cuales 8 son hombres ( $N_H = 8$ ) y 12 son mujeres ( $N_M = 12$ ).

En la figura 15, se expone la población finita con dos subpoblaciones o estratos.

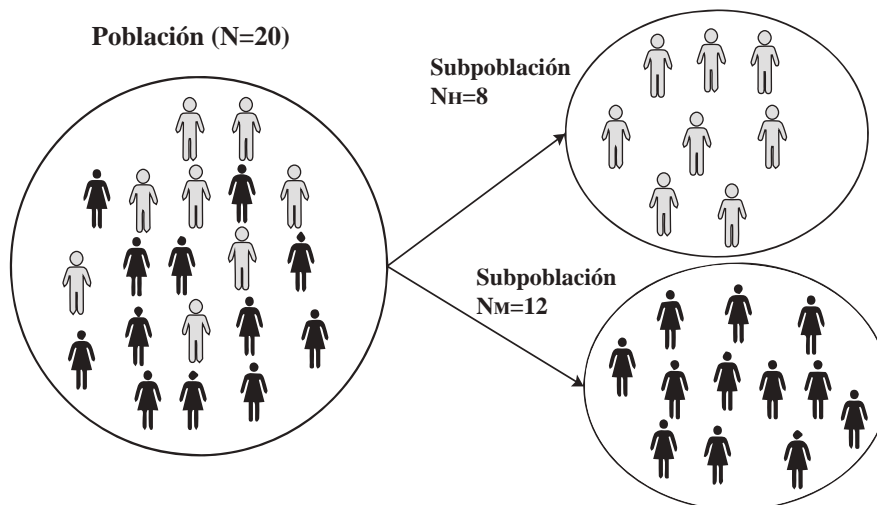


Figura 15. Población humana y subpoblaciones (estratos) por género

### Ejercicio 9:

Calcular los tamaños de muestra para cada uno de los tres estratos de una plantación de 180 árboles frutales ( $N = 180$ ), donde se calculó previamente su tamaño óptimo de muestra en 45 árboles ( $n_{\text{óptimo}} = 45$ ). Los tres estratos por la pendiente del terreno (subpoblaciones) se conforman por distinto número de árboles:  $N_1 = 95$ ,  $N_2 = 50$  y  $N_3 = 35$ .

### Representación gráfica:

En la Figura 16, se observa un diagrama de un lote de cultivo de un frutal perenne, en un terreno con tres estratos.



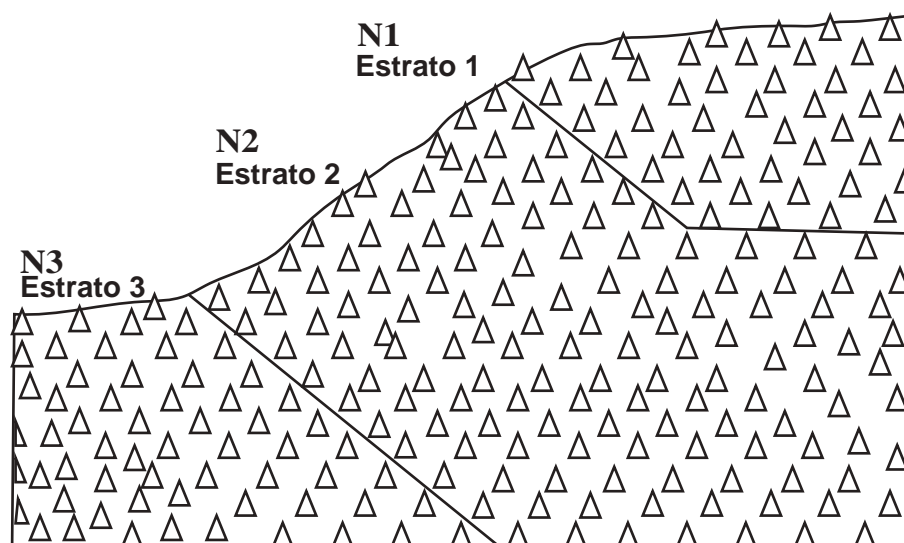


Figura 16. Estratificación de un terreno cultivado en base al grado de pendiente

**Análisis:**

La relación entre población N y muestras por estratos ( $n_i$ ) se indica en Cuadro 5.

Cuadro 5. Cálculo de las proporciones de las submuestras en un muestreo estratificado

Estratos por topografía del terreno	Población de árboles frutales	Proporción: $\frac{N_i}{N}$	Muestra por estrato: $n_{\text{óptimo}} * \left[\frac{N_i}{N}\right]$
Terreno ondulado	$N_1 = 95$	0,53	$n_1 = 45 * 0,53 = 24$
Terreno de ladera	$N_2 = 50$	0,28	$n_2 = 45 * 0,28 = 12$
Terreno plano bajo	$N_3 = 35$	0,19	$n_3 = 45 * 0,19 = 9$
Total	$N = 180$	1,00	$n_{\text{óptimo}} = 45 \text{ árboles}$

**Decisión:**

El tamaño óptimo de muestra es 45 árboles que se distribuye en tres estratos, según la topografía del terreno: en el ondulado 24, en la ladera 12 y en el plano 9 árboles.

2.4.2.4 MUESTREO POR CONGLOMERADOS

Es una técnica que aprovecha la existencia de grupos (conglomerados) en la población que contiene las características a estudiar. En este caso, se seleccionan únicamente una parte de estos conglomerados, en forma aleatoria, y a partir del muestreo dentro de los grupos seleccionados se puede deducir la información de interés del total de la población (Badii et al., 2004, p. 37).

En cada conglomerado hay unidades de estudio que representan la variabilidad de la población N. Posteriormente, dentro de cada conglomerado se aplica un muestreo aleatorio, sistemático o estratificado, según sea el caso, hecho que lo convierte en un muestreo polietápico. En un primer momento se sortean los conglomerados y luego se continúa con el muestreo aleatorio, en cada uno de ellos.



**Ejercicio 10:**

Con el método de muestreo por conglomerados, elegir aleatoriamente tres comunidades rurales (50%) de un conjunto de seis que tiene un territorio de interés, para aplicar una encuesta de diagnóstico agrosocial.

El diagrama de las comunidades rurales y la selección de tres conglomerados mediante sorteo se indica en la Figura 17. Dentro de cada conglomerado se aplicará un muestreo aleatorio.

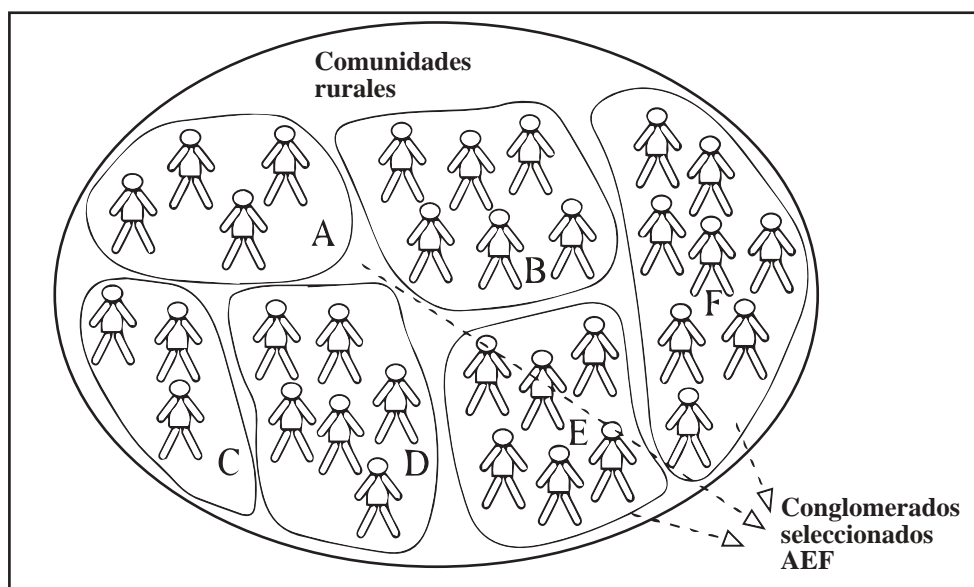


Figura 17. Muestreo por conglomerados en seis comunidades rurales de un territorio

## 2.5 TÉCNICAS DE MUESTREO

En la planificación del muestreo, antes de definir el método y tamaño de muestra, hay que conocer la forma de dispersión de la población (estudios exploratorios).

Los organismos vivos en el campo, se dispersan de diferentes formas (Figura 18):

- Errática (p.e.: algunas malezas)
- Aleatoria (p.e.: muchas plagas y patógenos)
- Uniforme (p.e.: cultivos, bosques e industria)
- Agregada (p.e.: avispas).

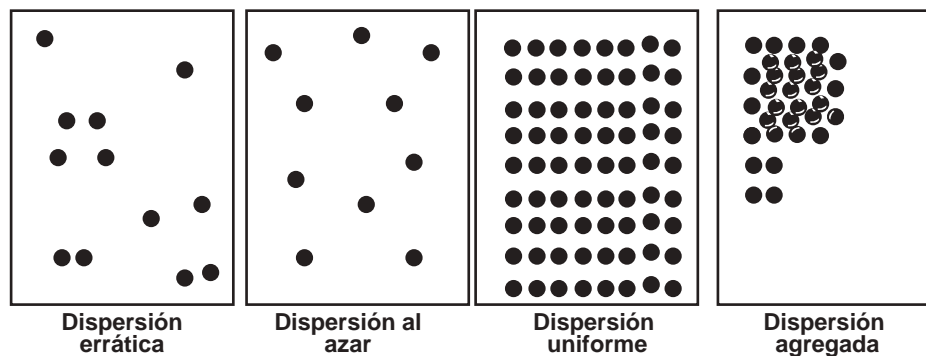


Figura 18. Distintas formas de dispersión de las poblaciones de organismos vivos

La distribución de los árboles en un agroecosistema (como un cafetal), simple (monocultivo) o compleja (policultivo), puede ser mapeada (Figura 19). En base de la medición del radio de la copa (medición con una cinta métrica del tronco hacia la gotera), se puede mediar el área de cobertura, individual o de todos los árboles.

Las técnicas de muestreo en diagonal, zigzag o circulares se usan para la toma de muestras de suelo, muestreo en cultivos y bosques establecidos de una sola especie. Para el muestreo de plagas se usan las técnicas circulares y guarda griega (Figura 20).

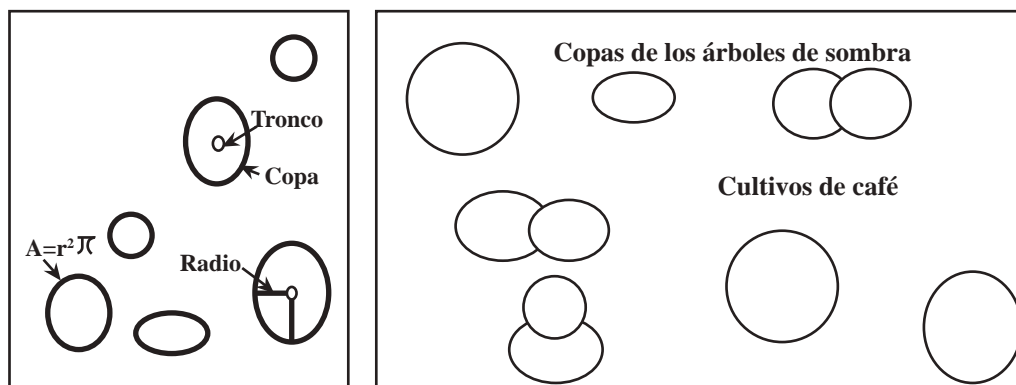


Figura 19. Mapeo de la cobertura arbórea en sistemas cafetaleros

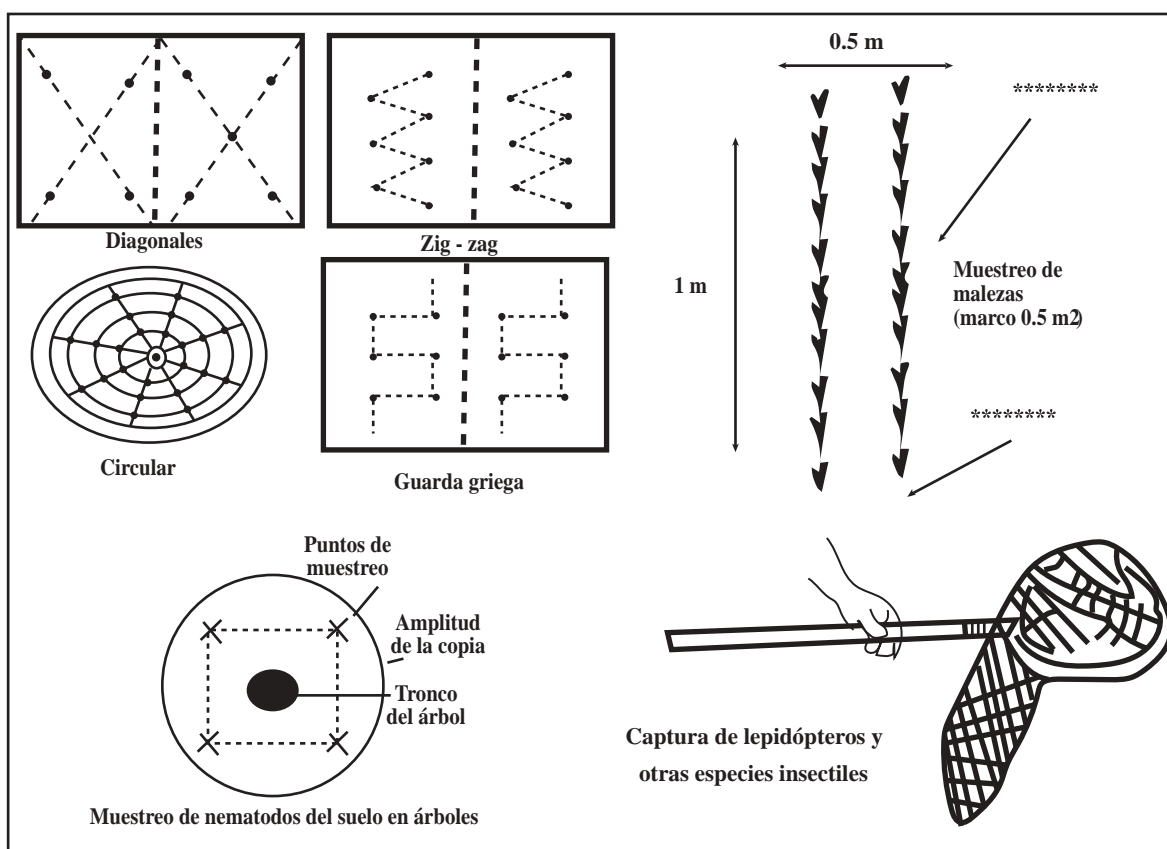


Figura 20. Técnicas de muestreo usadas en la investigación agrícola

## 2.6 VARIACIÓN

La variación es una propiedad inherente a toda característica medible en las poblaciones y en las muestras de organismos vivos o inertes. Un rasgo morfológico, un atributo sensorial o una característica física, química o biológica que cambia y se modifica por causas conocidas o desconocidas mostrando diferencias en sus magnitudes.

### 2.6.1 VARIABILIDAD BIOLÓGICA

Los organismos vivos se organizan en cinco reinos: vegetal, animal, fungi (hongos), monera (bacterias) y protista (protozoos). Los seres vivos pueden estudiarse a nivel de las células, tejidos, órganos, sistemas, individuos o como parte de comunidades biológicas y de ecosistemas, que en conjunto conforman la biósfera

Un ecosistema está compuesto por la biota de un área específica, que se interrelacionan con la abiota. Los cultivos vegetales se consideran ecosistemas artificiales. La variación es una propiedad intrínseca de las poblaciones y muestras, tanto de la biota como de la abiota, incluido los procesos industriales. La variación biológica es la tendencia de las poblaciones e individuos a diferenciarse. No hay dos individuos iguales, aun cuando sean gemelos idénticos, clones o híbridos. En las poblaciones vegetales y animales, los caracteres morfológicos, en conjunto, se conoce como fenotipo.

La variación fenotípica (VF) está en función del genotipo (G), del ambiente (A) y de la interacción genotipo x ambiente (IGA). El fenotipo, por lo tanto, es el conjunto de características morfo-fisiológicas de las poblaciones comunes, que son tangibles observables, medibles y comparables experimentalmente.

$$VF = f(G + A + IGA)$$

El factor genético (G) está relacionado con los genes, los cuales tienen acción directa sobre la expresión del fenotipo. El factor ambiente (A) se refiere al entorno biofísico (condiciones físicas, químicas y biológicas) donde se desarrollan las poblaciones de interés económico. Está claro que las plantas cultivadas no escogen el lugar donde van a crecer sino es el agricultor en función de sus intereses. La interacción genotipo x ambiente (IGA) se refiere al comportamiento diferencial de los cultivos en los distintos territorios.

En las poblaciones de organismos vivos, frecuentemente se encuentran cambios cualitativos en el fenotipo, heredables, que son el reflejo de las modificaciones en la secuencia del ADN (ácido desoxirribonucleico). Estos cambios cromosómicos se originan en forma natural por “errores” durante la división celular o por la exposición a agentes mutagénicos físicos (radiación ionizante) o químicos (ácido nitroso y otros).

Una mutación natural es la variedad Pache de café (plantas de porte muy bajo), que es un mutante de la variedad típica (porte alto). En la actualidad, se han desarrollado nuevas variedades de cultivos en base al uso de mutagénicos, que han enriquecido la variabilidad biológica.

### 2.6.2 VARIABILIDAD EN PROCESOS

Los procesos administrativos e industriales presentan variaciones que se estudian usando los mismos métodos y técnicas aplicadas en las ciencias de la vida. En procesos simples como la atención al cliente se estudia la eficiencia y en los procesos industriales, se

estudian los factores: materia prima (MO), ambiente físico u organizacional (A) y tiempos-movimientos de manipulación (IMA).

En una biblioteca, por ejemplo, al analizar un conjunto de libros, se puede estudiar el tipo de papel, la durabilidad; el tipo de pasta, la frecuencia de consulta, la temperatura y humedad interna y grado de deterioro de los libros, con los objetivos de determinar la forma de conservación o contabilizar el número de usuarios/semana.

En la elaboración de un producto agroindustrial, por ejemplo, se consideran los componentes identificados como 5M: materias primas, maquinaria, métodos, mano de obra y medio ambiente. En la materia prima se distingue el origen, el proveedor, la procedencia y los atributos y defectos. En la maquinaria se mide el nivel de complejidad, la fuente de energía y las habilidades para la operación. Los métodos se refieren a la operación, muestreos, control de calidad, análisis estadístico y toma de decisiones. En mano de obra se valora la escolaridad, la capacitación continua y las habilidades de los operadores. En el medio ambiente se trata del ámbito interno (bienestar, limpieza, manejo de desechos, señalética y servicios) y del externo (contaminación, alteración del paisaje, entre otros).

### 2.6.3 ESCALAS DE MEDICIÓN

La medición es la asignación de datos basado en protocolos específicos usando instrumentos de medición. Las propiedades del sistema numérico asociadas con las escalas de medición son: identidad, magnitud, intervalo y cero absoluto. Se identifican las escalas nominal, ordinal, de intervalo constante sin cero real y de intervalo constante con cero real que también se conoce como escala de razón o de proporción (Cuadro 6).

Los datos nominales y ordinales se transforman a índices, porcentajes o proporciones (Badii et al. 2021, p. 1). En estudios agrosociales, ambientales y de mercado y agroindustriales se usan escalas de Likert que son medidas ordinales para averiguar niveles de acuerdo, importancia jerárquica o de preferencias (Matas, 2018, p. 44). A diferencia de la simple pregunta dicotómica de “sí” o “no”, la escala de Likert permite diferenciar las respuestas en varios niveles, asumiendo que la fuerza e intensidad de las respuestas es lineal y que las actitudes pueden ser medidas con instrumentos cualitativos.

En la Figura 21, se exponen varias mediciones individuales de las variables de interés, en una población específica; por ejemplo: estudiantes del curso de bioestadística.

Cuadro 6. Tipos de dato, escalas de medición y tipos de variable, relaciones y ejemplos

Tipos de datos	Escala de medición	Tipos de variable	Relaciones	Ejemplos
Cualitativos	Nominal	Clasificatoria	Equivalencia	<ul style="list-style-type: none"> <li>Estado sanitario:</li> <li>Sano - enfermo (dicotómica)</li> <li>Nacionalidades (politómica)</li> </ul>
	Ordinal	Jerárquica	Equivalencia y jerarquía	<ul style="list-style-type: none"> <li>Vigor vegetal: De 1 a 5</li> <li>Calidad sensorial: De 1 a 10</li> <li>Satisfacción: De 1 a 10</li> </ul>
Cuantitativos	Intervalo constante sin cero real (intervalo)	Cuantitativa continua	Equivalencia, jerarquía y proporción conocida con signo $\pm$	<ul style="list-style-type: none"> <li>Temperatura: - 5°C</li> <li>Hora de ingreso: 08H00</li> <li>Saldo en la cuenta: -\$100</li> </ul>
	Intervalo constante con cero real (razón)	Cuantitativa continua	Equivalencia, jerarquía y proporción conocida solo con signo positivo	<ul style="list-style-type: none"> <li>Peso del fruto: 25 g</li> <li>Proteína en grano: 18 %</li> <li>Altura de planta: 200 cm</li> <li>pH del suelo: 6,50</li> </ul>

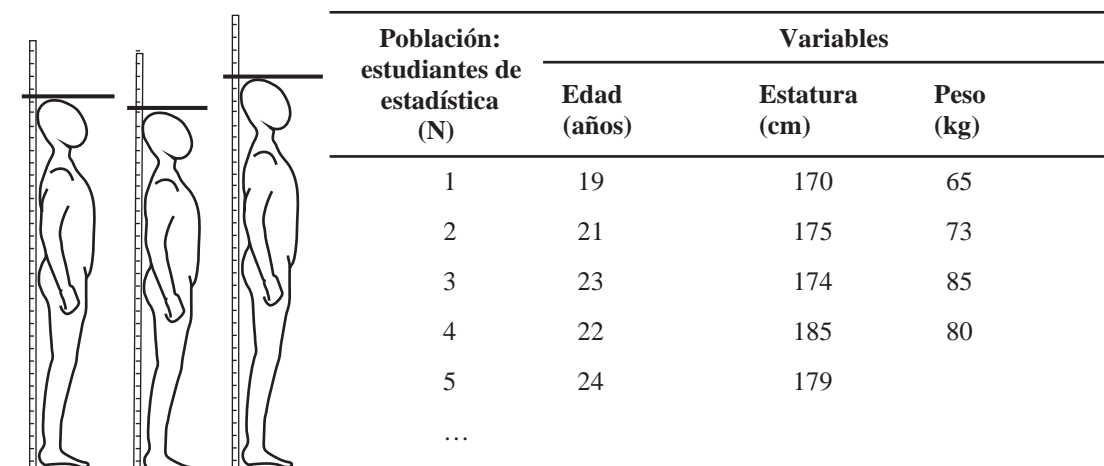


Figura 21. Mediciones de variables de interés en poblaciones finitas

Una aplicación de la escala Likert podría ser: ¿Qué tan importante es para usted un café caliente en el desayuno? Conteste la pregunta marcando un número en la escala ordinal indicada en la Figura 22.

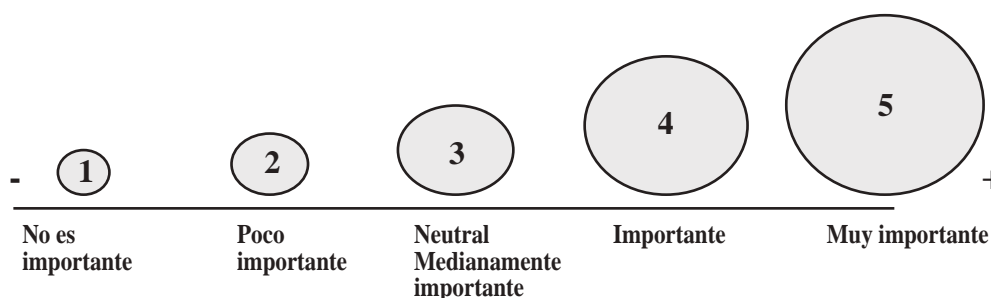


Figura 22. Escala Likert aplicada para valorar la importancia del café en el desayuno

En los procesos investigativos se han adoptado escalas ordinales específicas para la recolección de datos, algunas de las cuales se han generalizado y forman parte de protocolos internacionales, como los estándares de evaluación sensorial (SCA) y la satisfacción de vida.

**Evaluación sensorial de café.**- Se valoran 10 atributos organolépticos, cada uno usando la escala ordinal de 1 a 10, por tanto, la calidad sensorial de los cafés se califica sobre 100 puntos, según los estándares de Specialty Coffee Association-SCA (Millet, 2022).

**Satisfacción de vida.**- Según Oliver et al. (2018, p. 33), la satisfacción con el nivel de vida puede valorarse en base a cinco autoconceptos: académico, social, emocional, familiar y físico, usando una escala ordinal de 1 a 5 (1 = total desacuerdo y 5 = total acuerdo).

## 2.7 DATOS Y VARIABLES

Una característica común de las poblaciones y muestras es la variación. La altura de los árboles de un bosque mostrará diferencias, así como el peso de 10 mazorcas de maíz, tomadas al azar. Una población o una muestra contiene características distintivas, por ejemplo en maíz: días a la floración, días a la cosecha, altura de la planta, peso del grano, contenido de proteína, tolerancia a la sequía, entre otras. Cada característica que es medible se denomina variable. Las variables experimentales se definen en función de los objetivos del estudio y pueden clasificarse en cualitativos y cuantitativos (García, 2011, p. 22).

2.7.1 TIPOS DE DATOS

En estadística, hay diferentes tipos de datos o de variables que se utilizan para realizar análisis que se agrupan en cualitativos y cuantitativos (Lind et al., 2012, p. 9). Los datos cualitativos se obtienen mediante mediciones con las escalas nominal (clasificación) y ordinal (orden jerárquico). Los datos cuantitativos se obtienen mediante conteos (datos discretos) o usando escalas de intervalo constante.

En la Figura 23, se expone la relación entre los tipos de datos y las escalas de medición. En el Cuadro 7, se enlista un conjunto de variables de referencia, para distintas áreas del conocimiento, con sus escalas de medición y unidades de medida. Las variables experimentales se definirán en función de los objetivos de la investigación.

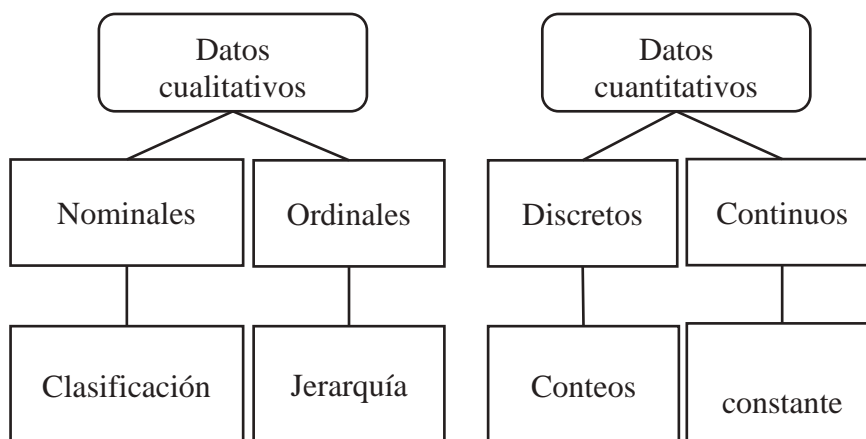


Figura 23. Tipos de datos cuantitativos y cualitativos

Cuadro 7. Áreas del conocimiento, elemento de estudio, variables, escala de medición, tipos de datos y unidad de medida en estudios agrosociales y ambientales

Área del conocimiento	Elemento de estudio	Variables Experimentales	Escala de medición	Tipo de dato	Unidad de medida
Agronomía	Planta	Altura de la planta	Razón	Continuo	cm
		Área foliar	Razón	Continuo	cm <sup>2</sup>
		Número de nudos	Conteo	Discreto	número
		Estado sanitario	Ordinal	Jerárquico	0 a 5
	Fruto	Peso del fruto	Razón	Continuo	g
		Color del fruto	Nominal	Categorico	colores
Zootecnia	Producción animal	Peso en pie	Razón	Continuo	kg.animal <sup>-1</sup>
		Estado de salud	Conteo	Jerárquico	0 a 5
Agroindustria	Calidad	Contenido de proteína	Razón	Continuo	g.100 g <sup>-1</sup>
		Calidad sensorial	Ordinal	Jerárquico	1 a 10
		Temperatura de conservación	Intervalo	Continuo	0 a 5
Social	Bienestar	Satisfacción en el empleo	Ordinal	Jerárquico	1 a 10
		Ingresos mensuales	Razón	Continuo	\$.mes <sup>-1</sup>
Ambiente	Recursos naturales	Erosión del suelo	Ordinal	Jerárquico	1 a 5
		Lombrices del suelo	Conteo	Discreto	número
		Temperatura atmosférica	Intervalo	Continuo	°C

## 2.7.2 TRANSFORMACIONES DE DATOS

La transformación es un procedimiento orientado a ajustar la serie de datos originales de una variable de respuesta al cumplimiento del supuesto de distribución normal o facilitar el análisis, la interpretación y la comprensión de los resultados.

### 2.7.2.1 TRANSFORMACIÓN DE DATOS ORDINALES A PORCENTAJES

Los datos cualitativos se transforman a por ciento mediante el uso de dos fórmulas:

Modelo clásico:

$$Y_i(\%) = \frac{Y_i}{n} (100) \quad [5]$$

Dónde:

$Y_i(\%)$  = Datos transformados a porcentaje

$Y_i$  = Datos originales de la variable

$n$  = Número de observaciones

Modelo de Townsend & Heuberger

$$Y_i(\%) = \frac{\sum(a.n)}{A.N} (100) \quad [6]$$

Dónde:

$I(\%)$  = Índice porcentual (por ejemplo, estado sanitario)

$a_i$  = Calificación de las observaciones

$n$  = Número de observaciones por cada categoría

$A$  = Valor más alto de la escala

$N$  = Número total de las observaciones en la muestra

#### Ejercicio 11:

Transformar los datos del estado sanitario de 200 plantas, medidas en escala ordinal de 0 a 3 a valores porcentuales.

**Datos:**

Los datos en escala ordinal y transformados a por ciento, se exponen en el Cuadro 8.

Cuadro 8. Transformación de datos ordinales a porcentajes

	Estado sanitario de plantas de pimiento (a)				Total de plantas (n)
	Sanas	Poco enfermas	Medianamente enfermas	Muy enfermas	
	0	1	2	3	
Plantas observadas por grado escalar (n)	4	11	15	170	200
Estado sanitario: $Y_i(\%) = \frac{Y_i}{n} (100)$	2,0	5,5	7,5	85,0	100

$$Y_i(\%) = \frac{\sum(a.n)}{A.N} (100) \quad \rightarrow \quad Y_i(\%) = \frac{(0*4+1*11+2*15+3*170)}{3(200)} (100) = \frac{551}{600} (100) = 92 \%$$

El 92 % de las plantas están enfermas, que corresponde a la incidencia.



**Interpretación:**

El 85 % están muy enfermas y 13 % en menor intensidad. La incidencia es del 92 %. Los dos procedimientos son adecuados para interpretar el estado sanitario de las plantas.

**Decisión:**

La incidencia de la enfermedad es de 92 %.

2.7.2.2 TRANSFORMACIÓN DE DATOS A INCREMENTOS PORCENTUALES

La transformación de datos originales a porcentajes en función del testigo tiene un propósito comparativo como valores incrementales. En este caso, el testigo equivale a 100%. El incremento porcentual de la variable de respuesta se calcula con la fórmula siguiente:

$$\text{Incremento (\%)} = \left[ \frac{T_i - T_0}{T_0} \right] (100) \quad [7]$$

Dónde:

Incremento % = Cambio porcentual respecto del testigo

T<sub>i</sub> = Dato original del tratamiento

T<sub>0</sub> = Dato original del testigo

**Ejercicio 12:**

Transformar a porcentaje los promedios de producción (g.parcela<sup>-1</sup>), considerando el promedio del testigo como 100 % ¿Cuál es el incremento en los tratamientos?

**Datos y análisis:**

Los datos originales y los transformados a incremento, se exponen en el Cuadro 9.

Cuadro 9. Transformación de datos cuantitativos a porcentaje en función del testigo

Tratamientos	Producción (g.parcela <sup>-1</sup> )	Porcentaje	Incremento (%): $\left[ \frac{T_i - T_0}{T_0} \right] (100)$
T0 (Testigo)	589	100	-
T1	1267	215	115
T2	1489	253	153

**Decisión:**

Si el 100 % es la producción de T<sub>0</sub>, los incrementos en T<sub>1</sub> y T<sub>2</sub> son de 115 % y de 153 %, respectivamente.

2.7.2.3 TRANSFORMACIÓN A VALORES ESCALAR AL MÁXIMO

La transformación de una serie de datos cuantitativos a “Escalar Al Máximo” (EAM), facilita la interpretación de los efectos en función de la mejor respuesta.

La fórmula usada para realizar esta transformación es:

$$\text{EAM(\%)} = \frac{T_i}{T_{\text{Máximo}}} (100) \quad [8]$$

Dónde:

EAM = Valor escalar al máximo

T<sub>i</sub> = Valor original obtenido en un tratamiento cualquiera

T<sub>Máximo</sub> = Valor original más alto obtenido con el mejor tratamiento

**Ejercicio 13:**

Transformar los datos promedios de peso de frutos/parcela de los tratamientos a valores Escalar Al Máximo (EAM).

**Datos:**

Los datos cuantitativos originales y los transformados a los valores Escalar al Máximo (EAM), se exponen en el Cuadro 10.

*Cuadro 10. Transformación de datos cuantitativos a valor Escalar al Máximo*

Tratamientos	Promedio peso del fruto (g)	EAM	EAM (%)
Testigo (T0)	589	0,396	39,6
T1	1267	0,851	85,1
T2	1489	1,00	100

**Análisis:**

$$EAM(\%) = \frac{T_i}{T_{\text{Máximo}}} (100) \quad \rightarrow EAM(T0) = \frac{589}{1489} (100) = 39,6 \%$$

$$EAM(T1) = \frac{1267}{1489} (100) = 85,1\% \quad \rightarrow EAM(T2) = \frac{1489}{1489} (100) = 100 \%$$

**Decisión:**

El testigo (T0) registró un peso promedio de los frutos.parcela<sup>-1</sup>, equivalente al 39,6 % comparado con el peso del tratamiento T2 (100 %), que resultó más alto.

2.7.2.4 TRANSFORMACIONES A RANGOS

La transformación de una serie de datos cuantitativos o cualitativos a rangos se basa en la asignación de puntajes, de menor a mayor (ranking), según la posición del dato original, en orden jerárquico. Esta transformación se aplica cuando se constatan datos provenientes de mediciones con cualquier escala.

Los datos transformados a rangos se usan en varias pruebas no paramétricas como: Correlación de Spearman, análisis de varianza de Kruskal-Wallis y análisis de varianza por rangos de Friedman.

En el programa Excel, la transformación de una serie de datos a rangos, se realiza usando la función:

$$=JERARQUIA.MEDIA(\text{número}, \text{referencia}, [\text{orden}])$$

**Ejercicio 14:**

Transformar los datos cuantitativos a rangos cuando no hay ligas (A) y cuando hay ligas (B). La variable de respuesta es del tipo “mayor es mejor”.

**Datos y transformación a rangos:**

Los datos cuantitativos originales en escala ordinal y los transformados a rangos, se exponen en el Cuadro 11.

Cuadro 11. Transformación de datos originales a rangos, con y sin valores repetidos.

	Observaciones	Datos originales	Datos transformados a rangos
(A) Datos originales sin valores repetidos	A1	58	2
	A2	100	5
	A3	88	4
	A4	50	1
	A5	70	3
(B) Datos originales con valores repetidos	B1	55	1
	B2	70	2,5
	B3	70	2,5
	B4	110	4

**Decisión:**

En el grupo A, el mejor resultado corresponde a N2 = 100 unidades (posición 5). En el grupo B, la mejor respuesta se observa en L4 = 110 unidades (posición 4).

2.7.2.5 TRANSFORMACIONES MULTIPLICATIVAS Y DIVISORIAS

Las series de datos con valores muy pequeños o muy altos pueden transformarse sea mediante la multiplicación o la división, por constantes conocidas para facilitar el análisis. Al realizar las transformaciones multiplicativas y divisorias, hay que considerar el sistema internacional de unidades (Pérez, 2015, pp. 49-74). Ejemplos: la transformación de miligramos a gramos, de gramos a kilos o de kilos a toneladas.

**Ejercicio 15:**

Realizar transformaciones multiplicativas (A) y divisorias (B) en las series de datos indicadas.

**Datos y análisis:**

Ejemplos de datos originales y transformados por constantes de multiplicación o de división (Cuadro 12).

Cuadro 12. Transformación de datos originales con factores multiplicativos y divisorios

A: Transformaciones multiplicativas			B: Transformaciones divisorias		
n	Datos originales en gramos	Transformación a miligramos	n	Datos originales en kilos	Transformación a toneladas métricas
n1	0,0010	1,0	n1	111.111	111,1
n2	0,0054	5,4	n2	528.956	529.0
n3	0,0067	6,7	n3	236.422	236,4
n4	0,0078	7,8	n4	453	0,453

**Decisión:**

La transformación multiplicativa de gramos (g) a miligramos (mg), así como la transformación divisoria de kg a toneladas, facilita la comprensión de las magnitudes.

### 2.7.2.6 TRANSFORMACIONES LOGARÍTMICAS

Los valores originales (X) muy elevados para poder analizarlos estadísticamente deben transformarse en logaritmos de X, sean de base 10 (log) o de base  $e = 2,71828$ , llamados naturales. Estas transformaciones se usan en análisis de conteos y en estudios dinámicos de tiempos, movimientos y transectos, comprimen los valores altos y expanden los bajos.

En edafología, para medir la acidez del suelo se usa la escala de pH, que varía de 1 a 14. El pH es un modelo logarítmico de los cationes de hidrógeno.  $pH = \log_{10}[H^+]$ . Esto significa que un pH = 5 (100.000 cationes  $H^+$ ) y un pH = 6 (1000.000 cationes  $H^+$ ) tienen una diferencia de 10 veces el número de cationes  $H^+$ .

Algunas consideraciones para su uso en el análisis matemático:

El logaritmo de 1, en cualquier base, es 0.

El logaritmo de un producto es igual a la suma de los logaritmos de sus factores:

$$\log M.N = \log M + \log N$$

El logaritmo de un cociente es igual a la resta de los logaritmos de sus factores:

$$\log \left[ \frac{M}{N} \right] = \log M - \log N$$

#### Ejercicio 16:

Transformar los datos originales de unidades formadoras de colonias (X) a valores logarítmicos y de notación científica.

#### Datos y análisis:

Los datos cuantitativos originales, transformados a valores logarítmicos y con notación científica, se exponen en el Cuadro 13.

Cuadro 13. Transformación de datos cuantitativos a valores logarítmicos y con notación científica

Unidades formadoras de colonias (UFC)	Valores originales X	log X	ln (X)	Notación científica
U1	100.000	5,00	11,51	1,00 E + 05
U2	1.000.000	6,00	13,82	1,00 E + 06
U3	10.000.000	7,00	16,12	1,00 E + 07
U4	100.000.000	8,00	18,42	1,00 E + 08
U5	58.325.456	7,77	17,88	5,833 E + 07
U6	444.678	5,648	13,01	4,447 E + 05

Nota: log = logaritmo de base 10 y ln = logaritmo de base  $e = 2,71828$

#### Decisión:

Los valores de X al transformarse a log, ln y notación científica, reducen su magnitud, tornándose manejables y de fácil comprensión.

2.7.2.7 TRANSFORMACIONES A RADICALES O POTENCIAS

Las transformaciones a radicales o potencias se usa cuando la distribución original de datos tiende a ser asimétrica. Las aplicadas más frecuentemente en estadística son:

$Y = \sqrt{X}$  : Cuando en valores porcentuales muy altos o muy bajos

$Y = \sqrt{X + 1}$  : Cuando los datos en porcentaje fluctúan de 0 a 20 % o de 80 a 100 %.

$Y = \sqrt{X + 0,5}$ : Cuando se trata de conteos y hay valores de 0

$Y = \frac{1}{X}$  : Se usa para analizar tiempos

$Y = X^2$  : Se usa cuando hay asimetría negativa.

**Ejercicio 17:**

Transformar los datos porcentuales de incidencia de minador de la hoja del cafeto a valores radicales o potencias e interpretar los resultados.

**Datos y análisis:**

Los datos porcentuales pueden transformarse con distintos métodos, principalmente como los expuestos en el Cuadro 14.

Cuadro 14. Transformación de datos porcentuales a radicales y potencias

Incidencia de minador (%) X	$\sqrt{X}$	$\sqrt{X + 0,5}$	$\sqrt{X + 1}$	$\frac{1}{X}$	$X^2$
0	0	0,707	1,0		0
12	3,46	3,536	3,6	0,083	144
18	4,24	4,301	4,4	0,056	324
80	8,94	8,972	9,0	0,0125	6400
90	9,49	9,513	9,5	0,0111	8100
100	10,00	10,025	10,0	0,0100	10000

**Decisión:**

La transformación de los datos originales a radicales o potencias genera nuevas series de datos que posibilitan el ajuste de las distribuciones asimétricas.

2.7.2.8 TIPIFICACIÓN DE DATOS

La tipificación es una potente transformación de datos originales a desviaciones típicas usando el estadístico Z, lo que permite comparar la posición relativa de los datos respecto de su media y ajustar los rasgos de asimetría de la curva normal (Martínez et al, 2020). La fórmula para la transformación de los datos originales al valor Z, es la siguiente:

$$Z = \frac{(x - \bar{X})}{s} \quad [9]$$

Dónde:

Z = Valores tipificados a partir de la transformación de los datos originales

X = Un dato original de la serie de observaciones de la muestra

$\bar{X}$  = Media de los datos de la variable original

S = Desviación estándar de la serie de datos de la variable original

**Ejercicio 18:**

Tipificar los valores de X de la serie de datos adjunta e interpretar.

**Datos y análisis:**

Los datos cuantitativos originales y los tipificados, se exponen en el Cuadro 15.

Cuadro 15. Transformación de datos a los valores Z y cálculo de las funciones densidad y distribución

$n_i$	Producción (X: kg.parcela <sup>-1</sup> )	$\bar{X}$	S	$Z = \frac{(X - \bar{X})}{S}$	Función densidad f(x)	Función distribución F(X)
1	0,45	10,041	0,282	-2,099	0,044	0,018
2	0,45	10,041	0,282	-2,099	0,044	0,018
3	0,62	10,041	0,282	-1,496	0,130	0,067
4	0,67	10,041	0,282	-1,318	0,167	0,094
5	0,69	10,041	0,282	-1,247	0,183	0,106
6	0,71	10,041	0,282	-1,176	0,200	0,120
7	0,71	10,041	0,282	-1,176	0,200	0,120
8	0,78	10,041	0,282	-0,927	0,260	0,177
9	0,79	10,041	0,282	-0,892	0,268	0,186

**Distribución de las funciones:**

Las funciones de distribución F(X) y de densidad f(x) (Gráfico 3), a partir de los estadísticos  $\bar{X}$  y S, posibilita visualizar la forma de campana o el alejamiento de ella.

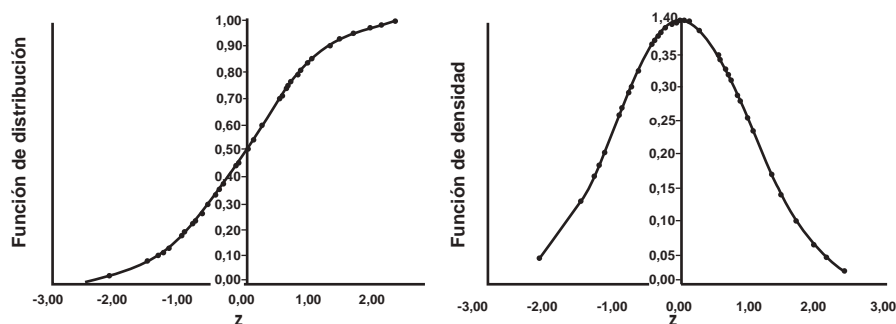


Gráfico 3. Funciones de distribución y de densidad para datos cuantitativos tipificados

**Decisión:**

Cuando la media de la muestra  $\bar{X} = 10,041$  el valor  $Z = 0$  y el área bajo la curva es igual a 1,0. Si el parámetro  $\mu = 0$ , el 50 % de los datos están al lado derecho y el otro 50 % al lado izquierdo de la curva.

**2.7.3 CÁLCULO DE DATOS PERDIDOS**

En un experimento puede haber datos faltantes a causa de una “parcela perdida” (no hay dato) o los datos pueden ser “fuera de tipo”. En los experimentos con diseño en bloques al azar, cuando se constata una “parcela perdida”, hay que calcular el “dato perdido. Se sugiere no asignar solo la media del tratamiento, porque también hay efecto de los bloques, por lo tanto, resultaría impreciso.

En los experimentos con diseños completamente al azar, no hace falta calcular “parcela perdida”, habría que analizar como un diseño DCA con desigual número de observaciones.

2.7.3.1 ESTIMACIÓN CON LA FÓRMULA DE YATES

Para calcular los datos perdidos, Robles (1986, p. 222) propone usar el método basado en la fórmula de Yates:

$$Y_{pp} = \frac{t(T) + r(R) - G}{(t - 1)(r - 1)} \quad [10]$$

Dónde:

$Y_{pp}$  = Valor estimado de la “parcela perdida” en un experimento de bloques al azar

t = Número de tratamientos

r = Número de repeticiones

T = Valor de la suma de los tratamientos sin corregir

R = Valor de la suma de los bloques o repeticiones sin corregir

G = Suma total en el experimento sin corregir

**Ejercicio 19:**

En un experimento de 5 tratamientos y 4 repeticiones, falta el dato de la parcela III-T3 de la variable producción (Y), en kg.parcela<sup>-1</sup>. Calcular el dato faltante con la fórmula de Yates.

**Datos:**

Los datos del ensayo con “parcela perdida” en la repetición III y tratamiento T3, se exponen en el Cuadro 16.

Cuadro 16. Datos de un ensayo de cinco tratamientos y cuatro repeticiones con una “parcela perdida”

Repeticiones	T1	T2	T3	T4	T5	$\Sigma$ Repeticiones sin corregir	$\bar{Y}$ Repeticiones sin corregir
I	80	68	70	60	102	380	76
II	95	70	80	55	90	390	78
III	99	73	Parcela perdida	85	111	368	92
IV	87	75	77	80	96	415	83
$\Sigma$ Tratamientos sin corregir	361	286	227	280	399	1553	311
$\bar{Y}$ Tratamientos sin corregir	90	72	76	70	100	388	82

Nota: Tratamientos t = 5 y Repeticiones r = 4. T3 = suma del tratamiento 3 sin corregir = 227, RIII = suma de repetición III sin corregir = 368 y G = suma total sin corregir = 1553

**Análisis:**

$$Y_{pp} = \frac{t(T) + r(R) - G}{(t - 1)(r - 1)} \quad \rightarrow \quad Y_{pp} = \frac{5(227) + 4(368) - 1553}{(5 - 1)(4 - 1)} = 88 \text{ kg}$$

**Decisión:**

El valor estimado de la parcela perdida III-T3 es 88 kg.



### 2.7.3.2 CÁLCULO DE PARCELA PERDIDA CON LA MEDIA GEOMÉTRICA

El “dato perdido” puede calcularse como la media geométrica a partir de los promedios “sin corregir” de repetición y tratamiento pertinentes. Las medias originales generan una nueva media. La fórmula de la media geométrica es la siguiente:

$$Y_{pp} = \sqrt{\bar{T}(\bar{R})} \quad [11]$$

Dónde:

$Y_{pp}$  = Valor estimado de la parcela perdida en un ensayo en bloques al azar

$\bar{T}$  = Valor promedio del tratamiento sin corrección

$\bar{R}$  = Valor promedio de la repetición sin corrección

#### Ejercicio 20:

En un experimento de 5 tratamientos y 4 repeticiones, falta el dato RIII-T3 de la variable producción (Y), en kg.parcela<sup>-1</sup>. Calcular el dato faltante, usando la media geométrica.

#### Datos:

$\bar{T}_3 = 76$  kg.parcela<sup>-1</sup>: promedio del tratamiento T3 sin corrección

$\bar{R}_{III} = 92$  kg.parcela<sup>-1</sup>: promedio del bloque R3 sin corrección

#### Análisis:

$$Y_{pp} = \sqrt{\bar{T}(\bar{R})} \rightarrow Y_{pp} = \sqrt{76(92)} = \sqrt{6992} = 84 \text{ kg.parcela}^{-1}$$

#### Decisión:

El valor estimado del dato faltante es 84 kg.parcela<sup>-1</sup>.

## 2.8 VALIDACIÓN DE LA CALIDAD DE LOS DATOS

La adecuada calidad de los datos es una condición esencial para iniciar el análisis y asegurar la validez de los resultados. Esta acción involucra la exploración de la variación en la serie de datos mediante un dispersograma, la identificación de los puntos “fuera de tipo” y, cuando sea pertinente, el cálculo del “dato perdido”.

Un dato atípico es una observación muy grande o muy pequeña que tiene un efecto desproporcionado sobre la media o la desviación estándar, situación que conduce a decisiones equivocadas. Hay varias herramientas para identificar datos atípicos, como: Dispersograma, límites de control y rango intercuartílico.

### 2.8.1 DIAGRAMA DE DISPERSIÓN

El dispersograma es una herramienta de análisis gráfico usado para relacionar las variables y valorar el comportamiento de una serie de datos cuantitativos. El dispersograma se realiza con los datos originales o con los residuos. En el Gráfico 4, se indican los datos de altura de 32 clones de café robusta, a razón de 10 plantas por clon, donde se observa un grupo de puntos, muy arriba de la “nube de datos”. Se verificó que corresponden a la altura de un clon muy alto<sup>4</sup> y no son “fuera de tipo”.

<sup>4</sup> Este genotipo corresponde a un grupo de clones seleccionados por el agricultor Ramiro Piñas, de Sucumbíos

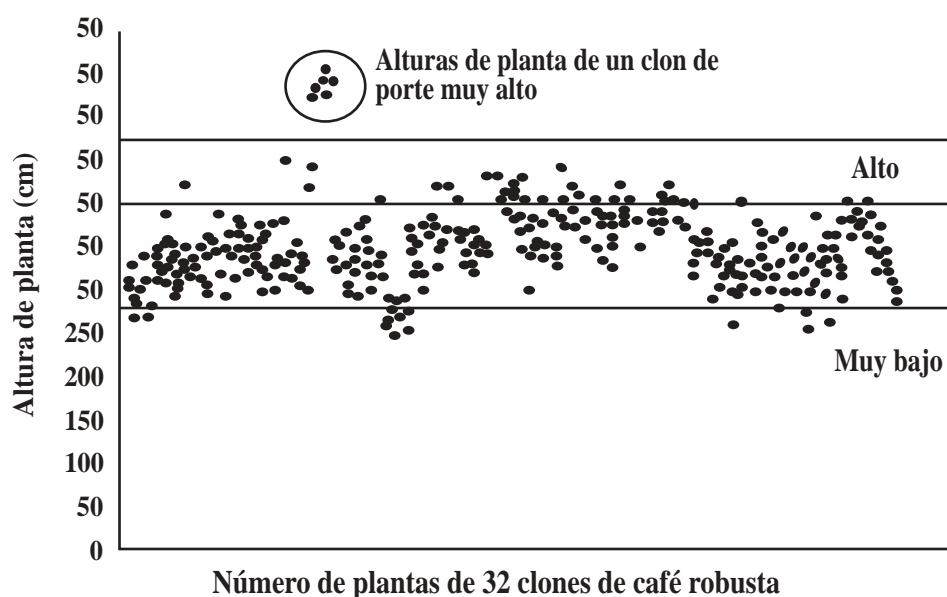


Gráfico 4. Dispersograma de la altura de planta de 32 clones de café robusta

El dispersograma de los residuos (diferencias entre los valores observados y la media general), permite valorar los potenciales sesgos, a cualquiera de los lados de la distribución.

La fórmula para calcular los residuos es la siguiente

$$\text{Residuo} = X_i - \bar{X} \quad [12]$$

Dónde:

$X_i$  = Valor observado en un punto

$\bar{X}$  = Valor promedio de a serie de datos

El procedimiento para analizar los residuos en un dispersograma, es el siguiente:

- Calcular las diferencias entre valores observados y la media general
- Graficar un dispersograma con los residuos y determinar si hay un alejamiento de los puntos por encima y por debajo de cero. Si uno o más puntos se alejan de la “nube de datos” se convierten en sospechosos de ser “atípicos”
- Explicar el origen del desvío, en caso de no haber argumento se considera como un dato “fuera de tipo”
- Recalcular los estadígrafos luego de la eliminación de los datos atípicos

En el Gráfico 5 se indica un dispersograma con los valores de los residuos, donde los puntos se ubican alrededor del cero, variando proporcionalmente en el rango de - 60 a + 60.

Los residuos con valores  $> \pm 60$  se alejan de la franja aceptable de datos y es notoria la tendencia a ser atípicos. En estas circunstancias, debe realizarse una “limpieza” de los datos atípicos y reprocesamientos de los análisis.

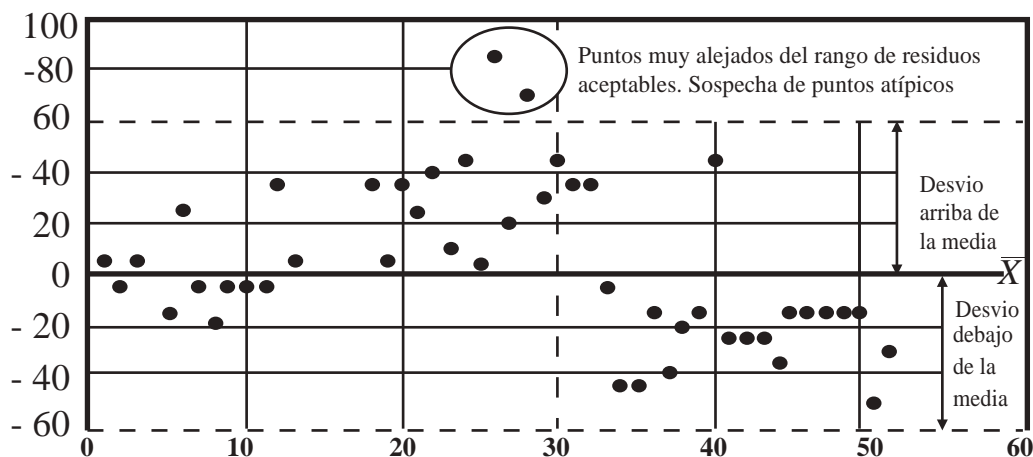


Gráfico 5. Dispersograma de residuos para identificar "puntos atípicos"

### 2.8.2 LÍMITES DE CONTROL

La indicación de los límites de control superior (LSC) e inferior (LIC) se fundamentan en dos principios: la normativa legal vigente<sup>5</sup>, y en la política empresarial<sup>6</sup>. Durante el monitoreo que realizan los responsables del control de calidad, en cualquier momento, pueden verificar que el proceso está bajo control (dentro de los límites) o fuera de los límites inferior y superior de control. En el momento de detectar un dato atípico, tiene que decidirse una inmediata intervención para volver al proceso normal. Por ejemplo: En una planta de semillas, se aceptan defectos físicos entre los límites: LIC = 0 % y LSC = 4 %, siendo la media LC = 2 %. Si en un monitoreo, en cualquier momento, se llegase a observar un dato con más del 4 % de defectos, el proceso está fuera control (Gráfico 6). Habría que parar la operación para tomar las medidas correctivas.

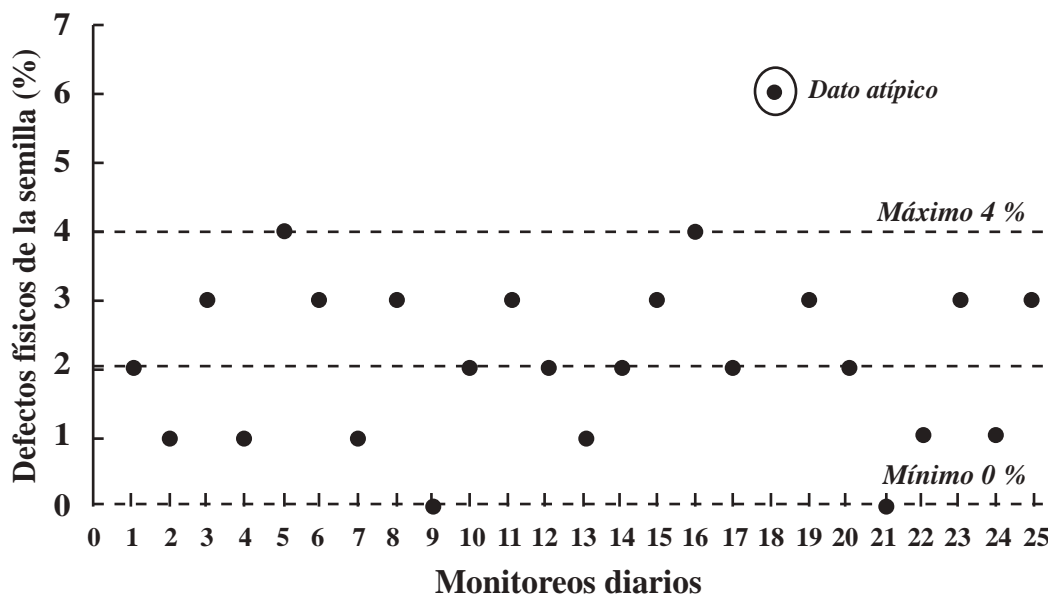


Gráfico 6. Monitoreo de defectos físicos del grano en una planta de semillas

<sup>5</sup> Cada producto tiene normas que regulan su aplicación. Como: los límites máximos de residuos (LMR) fijado por el Codex alimentarius, los límites de ocratoxina A en cacao, café y otros.

<sup>6</sup> La política empresarial se basa en el cumplimiento de normas legales y los valores propios de ella como la ética, bioética, satisfacción al cliente, entre otras.

### 2.8.3 CONTROL DE DATOS ATÍPICOS CON LA RELACIÓN MEDIA Y DESVIACIÓN TÍPICA

Este método de determinación de “datos fuera de tipo” (DFT) u “Outlier”, se basa en la siguiente relación:

$$DFT = \frac{(Y_i - \bar{Y})}{s} > 3 \quad [13]$$

Dónde:

DFT = Dato atípico, “dato fuera de tipo” u “Outlier”

$Y_i - \bar{Y}$  = Diferencia entre el dato observado y la media general

S = Desviación estándar de la muestra

#### Ejercicio 21:

Determinar si el dato  $Y_i = 115$  es atípico en la serie de datos de altura de planta (cm), donde  $\bar{Y} = 80$  cm y  $S = 10$  cm.

#### Regla de decisión:

Si  $\frac{(Y_i - \bar{Y})}{s} > 3$ : Dato atípico

Si  $\frac{(Y_i - \bar{Y})}{s} \geq 3$ : Dato típico de la serie (correcto)

#### Análisis:

$$\frac{(Y_i - \bar{Y})}{s} = \frac{(115 - 80)}{10} = 3,5 > 3,0. \text{ El dato } Y_i = 115 \text{ cm es un dato atípico.}$$

#### Decisión:

$Y_i = 115$  cm es un dato fuera de tipo (dato atípico).

### 2.8.4 DETECCIÓN DE DATOS ATÍPICOS CON EL RANGO INTERCUARTÍLICO

El rango intercuartílico (RI) se calcula a partir de la diferencia entre los cuartiles Q3 y Q1, representados en el diagrama de caja. El RI posibilita valorar la distribución e identificar la presencia de datos atípicos (Lind et al., 2012, p. 118), localizados en la parte superior o inferior de la serie de datos.

La regla de decisión se fundamenta en las fórmulas:

$$RI = Q3 - Q1 \quad [14]$$

$$D_{sa} > Q3 + 1,5(RI) \quad [15]$$

$$D_{ia} < Q1 - 1,5(RI) \quad [16]$$

Dónde:

RI = Rango intercuartílico

Q3 = Valor límite del cuartil 3

Q1 = Valor límite del cuartil 1

$D_{sa}$  = Dato superior aceptado como límite de la serie

$D_{ia}$  = Dato inferior aceptado como límite de la serie

### Ejercicio 22:

Determinar si el valor máximo observado de producción  $X_i = 2.739 \text{ g.planta}^{-1}$ , en una serie de datos, corresponde a un dato atípico, en base al análisis del diagrama de caja indicado.

#### Análisis:

En el Gráfico 7, se detalla el diagrama de caja y los valores claves del análisis.

Rango intercuartílico (RI) =  $Q3 - Q1 = 1315 - 629 = 686 \text{ g.planta}^{-1}$

$D_{sa}$  = Dato superior aceptado como límite de la serie. Si  $D_{sa} > Q3 + 1,5(RI)$  el dato es atípico.

$D_{sa} = 1315 + 1,5(686) = 1.315 + 1.029 = 2.344 \text{ g.planta}^{-1}$ . El dato 2739 > 2.344 es atípico.

$D_{ia}$  = Dato inferior aceptado como límite de la serie. Si  $D_{ia} < Q1 - 1,5(RI)$  el dato es atípico.

$D_i = 629 - 1,5(686) = 629 - 1.029 = -400 \sim 0 \text{ g.planta}^{-1}$ . El dato 139 corresponde a la serie.

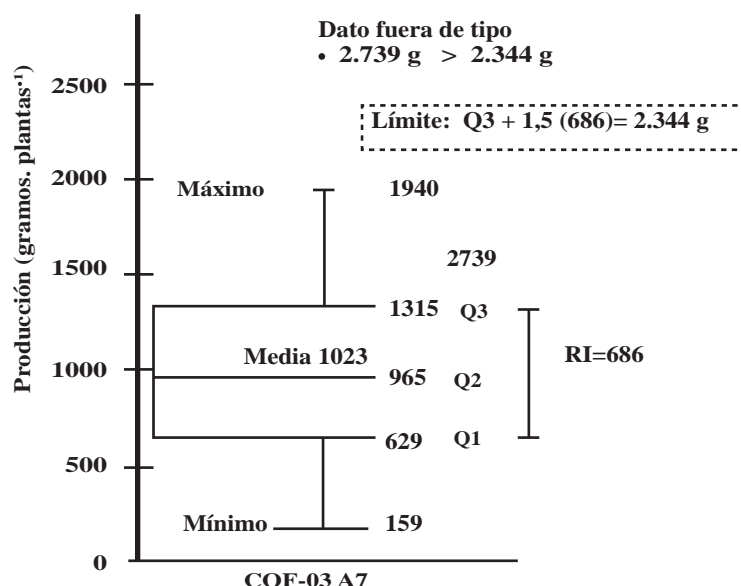


Gráfico 7. Identificación de datos atípicos basado en los cuartiles y el rango intercuartílico

#### Decisión:

El dato  $2739 \text{ g.planta}^{-1}$  es atípico. El valor máximo aceptable es  $2344 \text{ g.planta}^{-1}$ .

## 2.9 PRUEBAS ESTADÍSTICAS

En la investigación científica se efectúan las pruebas estadísticas para determinar con objetividad si se acepta o rechaza una hipótesis con base al cumplimiento de condiciones de las series de datos y la fijación de los niveles de riesgo y error, para la toma de decisiones.

La prueba estadística, según la hipótesis, puede ser de cola derecha, de cola izquierda o de dos colas. El término colas hace referencia a las prolongaciones de la curva normal hacia sus extremos derecho o izquierdo, construida con la serie de datos de la muestra o del experimento (Gráfico 8).

Con las pruebas de hipótesis se comparan tratamientos, localidades u otros factores en estudio y se toman decisiones: aceptar  $H_0$  (equivale a rechazar  $H_1$ ) o aceptar  $H_1$  (equivale a rechazar  $H_0$ ). Supóngase que se compara la producción de tres variedades de maíz, en el análisis de varianza (prueba F), las hipótesis estadísticas serían:

$H_0$ :  $V_1 = V_2 = V_3$ . La producción de las tres variedades es estadísticamente igual.

$H_1$ : En producción, al menos una variedad es significativamente diferente.

En el caso de aceptar  $H_1$ , para identificar la variedad con la producción significativamente mayor que las otras, se debe aplicar una de las pruebas *post hoc* (significa “después de esto”), como: Tukey, Duncan y Diferencia Mínima significativa (DMS).

### 2.9. 1. PRUEBAS DE UNA COLA Y DE DOS COLAS

En la prueba de cola derecha, el investigador orienta su interés hacia una respuesta “mayor es mejor”. Por ejemplo: “Determinar si la calidad sensorial de los cafés de Manabí, en la escala SCA  $\geq 80$  puntos, para considerarlos como especiales”.

La prueba de cola izquierda expresa el interés del investigador por el criterio “menor es mejor”. Por ejemplo: “Determinar si la incidencia de una plaga agrícola  $< 8\%$ . La prueba de dos colas expresa la no preferencia del investigador por un resultado. Por ejemplo: Comparar los rendimientos de los tratamientos T1 y T2 con el 95 % de confianza”.

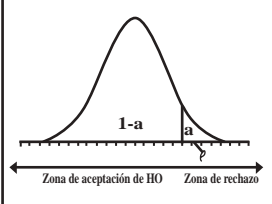
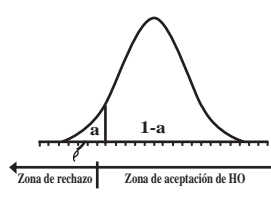
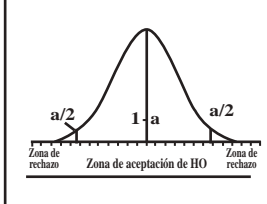
<b>Hipótesis estadísticas</b>	$H_0: \bar{X} \leq A$ $H_1: \bar{X} > A$	$H_0: \bar{X} \geq A$ $H_1: \bar{X} < A$	$H_0: \bar{X}_1 = \bar{X}_2$ $H_1: \bar{X}_1 \neq \bar{X}_2$
<b>Tipo de prueba</b>	Una cola (unilateral)	Una cola (unilateral)	Dos colas (bilateral)
<b>Región crítica</b>	Derecha	Izquierda	Ambos lados
<b>Gráfico</b>			
<b>Ejemplo:</b>	¿El café de Manabí tiene más de 80 puntos SCA?	¿La incidencia de roya no supera el umbral de 5%?	¿Las medias de rendimiento son estadísticamente iguales?

Gráfico 8. Hipótesis, pruebas, zona de rechazo de hipótesis nula y distribución normal

### 2.9.2 ERRORES EN LAS PRUEBAS ESTADÍSTICAS

En la estadística se constatan dos tipos de errores. En el Cuadro 17 se indica las relaciones de los errores de los tipos I y II con las decisiones estadísticas. Para controlar los errores en las pruebas estadísticas y que los resultados tengan validez hay que comprobar el cumplimiento de los supuestos exigidos para pruebas paramétricas.

Cuadro 17. Tipos de error estadístico en las pruebas de hipótesis

		Realidad	
		H <sub>0</sub> verdadera (A = B)	H <sub>0</sub> falsa (A ≠ B)
Decisión estadística sobre H <sub>0</sub>	Aceptar H <sub>0</sub>	Nivel de confianza (1 - α) Decisión correcta	β Error tipo II <i>Falso negativo</i>
	Rechazar H <sub>0</sub>	α Error tipo I: <i>Falso positivo</i>	Potencia de la prueba: (1 - β) <u>Decisión correcta</u>

Nota: Error tipo I es rechazar H<sub>0</sub> cuando en realidad es verdadera. El error tipo II es aceptar H<sub>0</sub> cuando es falsa

### 2.9.3. POTENCIA DE LA PRUEBA ESTADÍSTICA

El poder estadístico o potencia de la prueba es la probabilidad de que la hipótesis alternativa (H<sub>1</sub>) sea aceptada cuando es verdadera. Esta depende del tamaño de muestra y determina la fiabilidad de los resultados (Quesada y Figuerola, 2010, p. 24).

Los factores a considerar en el análisis de la potencia de la prueba estadísticas son:

- La magnitud del efecto considerado aceptable en una H<sub>0</sub>. ¿Hasta cuanto se considera estadísticamente igual?
- La variabilidad de los datos. ¿Hay alta variación razonable en los datos?
- El tamaño de la muestra es clave. En muestras de grandes se incrementan los grados de libertad y tiende a elevar la potencia estadística.
- El nivel de significación α se asocia a la probabilidad de cometer error tipo I.

La fórmula usada para calcular μ para T<sub>A</sub> es la siguiente:

$$\mu = \bar{Y} + Z_{\alpha} \frac{\sigma}{\sqrt{n}} \quad [17]$$

La fórmula que se usa para calcular Z considerando μ de T<sub>A</sub> y el promedio de T<sub>B</sub>, es:

$$Z = \frac{\mu_{TA} - \bar{Y}_{TB}}{\frac{\sigma}{\sqrt{n}}} \quad [18]$$

#### Ejercicio 23:

Calcular el poder estadístico, con una confianza del 95% para una prueba de dos colas, para un ensayo comparativo de dos tratamientos. La muestra n = 25, en TA el promedio fue  $\bar{Y}_A = 50 \text{ kg.parcela}^{-1}$  con una varianza  $S^2 = 4 \text{ kg}^2$  y en TB se obtuvo una media  $\bar{Y}_B = 51,5 \text{ kg.parcela}^{-1}$  con una varianza  $S^2 = 3 \text{ kg}^2$

#### Datos:

n = 25;  $\bar{Y}_A = 50 \text{ kg.parcela}^{-1}$ ;  $S^2 = 4,0$ ; S = 2,0 y α = 0,05

$Z_{\alpha} = 1,96$  (valor asociado α = 0,05 en pruebas de dos colas)

$\bar{Y}_B = 51,5 \text{ kg.parcela}^{-1}$

$S^2 = 3$

S = 1,732



**Fórmulas:**

$$\mu = \bar{Y} + Z_{\alpha} \frac{\sigma}{\sqrt{n}} \quad Z = \frac{\mu_{TA} - \bar{Y}_{TB}}{\frac{\sigma}{\sqrt{n}}}$$

**Hipótesis:**

H<sub>0</sub>: TA = TB

H<sub>1</sub>: TA ≠ TB

**Análisis:**

**Paso 1:** Estimar  $\mu$  del TA considerando el nivel de significación  $\alpha = 0,05$

$$\mu = \bar{Y} + Z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

$$\mu \text{ de } TA_{\alpha} = 50 + 1,96 \frac{2}{\sqrt{25}} = 50 + 1,96(0,4) = 50,784 \text{ kg}$$

**Paso 2:** Calcular Z para la diferencia de promedio  $TA_{\alpha} = 50,784$  y el promedio del contraste, en este caso el promedio de TB = 51,5 kg.parcela<sup>-1</sup>.

$$Z = \frac{\mu_{TA} - \bar{Y}_{TB}}{\frac{\sigma}{\sqrt{n}}} \rightarrow Z = \frac{50,784 - 51,5}{\frac{2}{\sqrt{25}}} = -1,79$$

**Paso 3:** Obtener el valor  $p$  asociado a  $Z = -1,79$ . El valor  $p$  de  $Z$  se obtiene en la tabla de distribución normal estandarizada o usando Excel.

En el programa Excel, el valor  $p$  de  $Z$  se obtiene con la función: =DISTR.NORM.ESTAND(Z).

Cuando  $Z = -1,79 \rightarrow$  el valor  $p$  de  $Z = 0,037 \sim 3,7 \%$ .

El valor  $p = 0,037$  corresponde a  $\beta$  o probabilidad de cometer error tipo II: “aceptar H<sub>0</sub> cuando en realidad es falsa”.

**Paso 4:** Calculo de la potencia de prueba.

$$\text{Potencia} = 1 - \beta \rightarrow 1 - 0,037 = 0,963 \sim 96,3 \%$$

El valor  $1 - \beta = 0,963$  corresponde a la potencia de la prueba estadística que equivale a la probabilidad de rechazar H<sub>0</sub> cuando realmente es falsa.

**Decisión:**

La potencia de la prueba es 96,3 % para una prueba de dos colas.

## 2.10 DISTRIBUCIONES DE PROBABILIDADES

La distribución de probabilidad es una característica intrínseca de la variable aleatoria dentro de la población, se asemeja a la distribución de frecuencias, describiendo la forma, situación ligada al tamaño de muestra, homogeneidad de la población y tipo de variable.

Una distribución teórica tiene dos rasgos: La función densidad y la función distribución. La función densidad indica es la distribución de los valores de la serie de datos dentro de una región concreta del área bajo la curva ( $p < 1$ ) en un punto específico X. La función distribución corresponde a la posición de dicho valor X en la ordenada.

## 2.10.1 DISTRIBUCIONES CONTINUAS

Las distribuciones normal, uniforme y geométrica son propias de las variables continuas.

### 2.10.1.1 DISTRIBUCIÓN NORMAL

La distribución normal también conocida como campana de Gauss es clave en la inferencia estadística. Se caracteriza por la forma simétrica de la curva y las dos colas que corresponden a los valores extremos, derecha e izquierda, extendidas de manera indefinida (Martínez y Martínez, 2010). La distribución normal se especifica por dos parámetros determinantes de la función densidad:  $N(\mu, \sigma)$ . Al analizar una serie, la distribución de frecuencias, absolutas o relativas, los histogramas con sus polígonos de frecuencias, configuran la forma de distribución.

Para valorar si la distribución tiende a la normal, se debe considerar lo siguiente:

- La media, mediana y moda tienden a coincidir en el mismo punto
- La curva de campana es simétrica, por lo tanto, el 50% de los datos se ubican hacia el lado derecho de la media y el otro 50% al lado izquierdo.
- Si el Coeficientes de asimetría ( $A$ )  $> \text{abs } |0,5|$  hay sesgo, a la derecha (+) o a la izquierda (-), evidenciándose un alejamiento de la curva normal.
- Si el Coeficientes de Curtosis ( $K$ )  $> \text{abs } |0,5|$  la curva es muy puntiaguda (+) o muy aplanada (-), habiendo un alejamiento de la normalidad.

### DISTRIBUCIÓN NORMAL ESTANDARIZADA

La transformación de datos originales a valores  $Z$ , se conoce como normalización o tipificación o estandarización<sup>7</sup>, que se adaptan al modelo  $N(\mu, \sigma)$  o su equivalente gaussiano  $N(0, 1)$ . La curva normal estándar se elabora con los datos tipificados, siendo:  $\mu = 0$  y  $\sigma = 1$ .

La función densidad para la variable  $X$  en  $N(\mu, \sigma)$ , se calcula con la fórmula:

$$f(x) = f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad [19]$$

Dónde:

$-\infty < x < \infty$

$\pi = 3,14159$

$e = 2,71828$

La fórmula para tipificar las variables cuantitativas es la siguiente:

$$Z = \frac{(X - \bar{X})}{S}$$

Dónde:

$Z$  = Estadístico medido en desviaciones estándar

$X$  = Valor individual de la variable cuantitativa,  $X_i$

$\bar{X}$  = Promedio

El área bajo la curva equivale a 1,0 ~ 100 %, por lo tanto, cualquier punto de interés puede calcularse en base a la distancia de ese punto hasta la media de la curva (Gráfico 9).

<sup>7</sup> En el Anexo 3 se indica la forma de obtener los valores  $Z$  en la tabla de distribución normal estándar

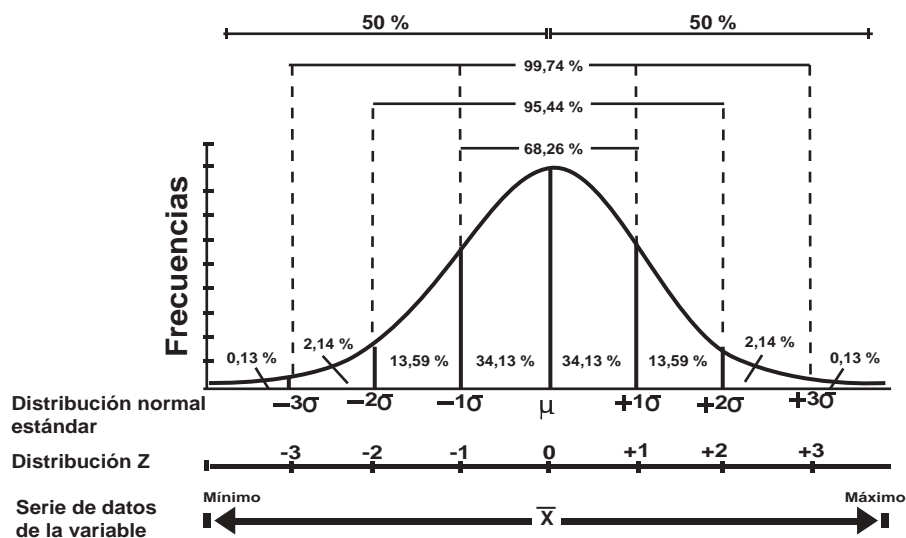


Gráfico 9. Propiedades de la distribución normal estandarizada

La prueba Z se realiza en cuatro pasos:

- Paso 1.- Definir gráficamente el área de interés bajo la curva normal
- Paso 2.- Determinar el valor Z
- Paso 3.- Buscar el valor  $p$  de Z en la tabla de probabilidades
- Paso 4.- Cuando se requiere calcular la probabilidad de segmentos dentro de la curva, hay que sumar o restar las áreas parciales pertinentes.

En Excel, el valor  $p$  de Z se obtiene con la función:

$$=DISTR.NORM.ESTAND(Z)$$

### Ejercicio 24:

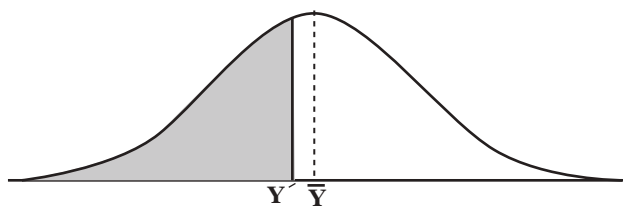
Determinar probabilidades de cosechar  $\leq 300 \text{ kg} \cdot \text{ha}^{-1}$  a partir de un diagnóstico productivo del cultivo de cacao, en un territorio específico. La muestra fue  $n = 50$  fincas, la media de rendimiento  $\bar{Y} = 360 \text{ kg} \cdot \text{ha}^{-1}$  con una desviación estándar  $S = 66 \text{ kg}$ .

#### Datos:

$$n = 50 \text{ fincas}, \bar{Y} = 360 \text{ kg ha}^{-1}, S = 66 \text{ kg y } Y = 300 \text{ kg} \cdot \text{ha}^{-1}$$

#### Análisis:

**Paso 1.-** Definir gráficamente el área de interés bajo la curva normal



**Paso 2.-** Determinar el valor Z

$$Z = \frac{(Y - \bar{Y})}{s} = \frac{300 - 360}{66} = -0,9091$$

**Paso 3.-** Buscar el valor  $p$  de  $Z$  en la tabla de probabilidades.

$$p(Z) = 0,182 \sim 18,2 \%$$

**Análisis e interpretación:**

En la próxima cosecha de cacao, la probabilidad de alcanzar rendimientos  $\leq 300$  kg.ha<sup>-1</sup> es del 18,2 %. Complementariamente, la probabilidad de obtener rendimientos  $> 300$  kg.ha<sup>-1</sup> es del 82 %.

**Decisión:**

En 9 de las 50 fincas es probable que se obtengan rendimientos  $\leq 300$  kg.ha<sup>-1</sup>.

### Ejercicio 25:

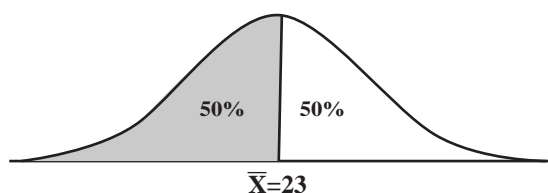
La información histórica de la temperatura de una ciudad M indica que esta variable climática sigue una distribución normal. La temperatura media del mes de febrero es  $\mu = 23^\circ$  C con una desviación típica  $S = 2^\circ$  C.

**Datos:**

$$\mu = 23^\circ\text{C}, X = 23^\circ\text{C} \text{ y } S = 2^\circ\text{C}$$

**¿Cuántos días del mes de febrero tendrían más de  $23^\circ$  C?**

**Paso 1.-** Definir gráficamente el área de interés bajo la curva normal



En febrero, los datos de las temperaturas medias  $\geq 23^\circ\text{C}$  será el 50% (promedio histórico), que en la curva normal corresponde a  $\mu = 0$ .

**Paso 2.-** Determinar el valor  $Z$

$$Z = \frac{(X - \bar{X})}{s} = \frac{23 - 23}{2} = 0$$

**Paso 3.-** Buscar el valor  $p$  ( $Z$ ) en la tabla de probabilidades

En la tabla de distribución normal estándar a  $Z = 0$  le corresponde un  $p$  ( $Z$ ) = 0,50  $\sim$  50%

**Análisis e interpretación:**

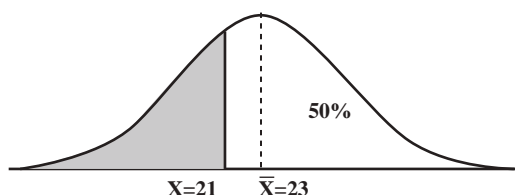
Se conoce que el mes de febrero tiene 28 días y la probabilidad de que la temperatura media sea mayor que  $23^\circ$  C es  $p = 0,50$ , el número de días ( $N$ ) con temperaturas  $\geq 23^\circ$  C se calcula como:  $p(N) = 0,50(28) = 14$  días.

**Decisión:**

Probablemente 14 días de febrero próximo tendrán temperaturas mayores que  $23^\circ\text{C}$ .

**¿Cuántos días del mes de febrero tendrían menos de 21° C?**

**Paso 1.-** Definir gráficamente el área de interés bajo la curva normal



**Paso 2.-** Determinar el valor Z

$$Z = \frac{(X - \bar{X})}{s} = \frac{21 - 23}{2} = -1$$

**Paso 3.-** Buscar el valor  $p$  (Z) en la tabla de probabilidades específica.

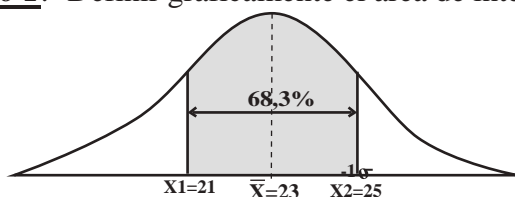
$$p(Z) = 0,159 \sim 15,9 \%$$

**Decisión:**

Dado que febrero tiene 28 días, el número de días que probablemente tengan temperatura  $\leq 21^\circ\text{C}$  se calcula como:  $p(N) = 0,159(28) = 4$  días.

**¿Cuántos días del mes de febrero tendrían de 21° a 25° C?**

**Paso 1.-** Definir gráficamente el área de interés bajo la curva normal



**Paso 2.-** Determinar el valor Z

$$Z_1 = \frac{(X_1 - \bar{X})}{s} = \frac{21 - 23}{2} = -1$$

$$Z_2 = \frac{(X_2 - \bar{X})}{s} = \frac{25 - 23}{2} = +1$$

**Paso 3.-** Obtener valores  $p$  de Z en la tabla de distribución normal estandarizada o usando la función de Excel pertinente. Los valores de Z se indican a continuación:

Puntos y segmento	$p$ de Z	Área bajo la curva (%)
X1: 21° C	0,159	15,9
X2: 25° C	0,841	84,1
X2 - X1	0,683	68,3

**Paso 4.-** Calcular el área bajo la curva, entre los puntos  $X_1 = 21^\circ\text{C}$  y  $X_2 = 25^\circ\text{C}$   $X_2 = 0,841$  y  $X_1 = 0,159$ . La diferencia  $X_2 - X_1 = 0,841 - 0,159 = 0,683 \sim 68,3 \%$ .

**Análisis e interpretación:**

Dado que febrero tiene 28 días, las temperaturas medias entre 21 y 25° C, ocurrirán en  $p(N) = 0,683(28) = 19$  días, asumiendo que la temperatura sigue una distribución normal.

**Decisión:**

Es probable que en 19 días del mes de febrero próximo, la temperatura varíe de 21° a 25° C.

### 2.10.1.2 DISTRIBUCIÓN UNIFORME

La distribución uniforme o rectangular describe una variable aleatoria con probabilidad constante sobre un intervalo  $a$  y  $b$ . La probabilidad de un suceso depende de la amplitud del intervalo y no de su posición.

La función densidad está dada por:

$$f(x; a, b) = f(x; a, b) = \frac{1}{b - a}$$

Siendo:  $a < x < b$

La función  $Y = f(X)$  sigue una distribución uniforme en el intervalo  $(0, 1)$ . Las fórmulas relacionadas con la distribución uniforme son las siguientes:

$$\mu = \frac{\text{Mínimo} + \text{Máximo}}{2} \quad [20]$$

$$\sigma^2 = \frac{(b - a)^2}{12} \quad [21]$$

$$p(X = x) = \frac{x - a}{b - a} \quad [22]$$

#### **Ejercicio 26:**

Calcular la probabilidad de que un valor  $X$  sea mayor que 1000 en una variable distribuida uniformemente entre 380 y 1200 unidades.

#### **Datos y análisis:**

Mínimo =  $a = 380$

Máximo =  $b = 1200$

$X = 1000$

$$\mu = \frac{\text{Mínimo} + \text{Máximo}}{2} = \frac{380 + 1.200}{2} = 790$$

$$\sigma^2 = \frac{(b - a)^2}{12} = \frac{(1200 - 380)^2}{12} = 56.033$$

$$p(X = 1000) = \frac{1000 - 380}{1200 - 380} = 0,7561 \text{ lado izquierdo de } X$$

Esto significa que el área debajo de  $X = 1000$  es de  $0,756 \sim 75,6 \%$ . El área, arriba de  $X = 1000$ , corresponde a la diferencia:  $1 - 0,756 = 0,244 \sim 24,4 \%$  (lado derecho de  $X$ ).

#### **Decisión:**

El  $24,4 \%$  de los datos, en la serie uniformemente distribuida, probablemente sea mayor que 1000 unidades.

### 2.10.2 DISTRIBUCIONES DISCRETAS

Las distribuciones discretas (aplicadas a conteos), más comunes son: uniforme discreta, binomial y Poisson.

#### 2.10.2.1 DISTRIBUCIÓN UNIFORME DISCRETA

La distribución uniforme discreta describe el comportamiento de poblaciones finitas, desde  $a$  hasta  $b$ , con valores consecutivos que tienen la misma probabilidad  $k$ .

Las fórmulas requeridas en la distribución uniforme discreta son:

$$\mu = \frac{(a+b)}{2} \quad [23]$$

$$K = \frac{1}{b} \quad [24]$$

$$p(X) = k(X_i) \quad [25]$$

Dónde:

$p(X)$  = probabilidad en  $X$

$K$  = probabilidad de ocurrencia

$a$  = valor mínimo

$b$  = valor máximo

Serie de datos:  $a, a + 1, a + 2, \dots, b$ .  $a < b$

### Ejercicio 27:

Un examen consta de 50 preguntas. Un alumno solo estudió las primeras 35. ¿Cuál es la probabilidad de que el examen cubra las preguntas estudiadas?

**Datos:**

$a = 1$  y  $b = 50$

$$\mu = \frac{(a+b)}{2} \rightarrow \mu = \frac{(1+50)}{2} = 25,5$$

**Análisis:**

$X = 35$  preguntas

$$K = \frac{1}{b} \rightarrow K = \frac{1}{50} \rightarrow 0,02$$

$$p(X = 35) = k(X_i) = 0,02(35) = 0,7 = 70 \%$$

**Decisión:**

El 70% del examen probablemente tendrá correspondencia con las preguntas estudiadas por el alumno.

### 2.10.2.2 DISTRIBUCIÓN BINOMIAL

Esta distribución es una descripción del número de éxitos al realizar  $n$  experimentos Bernoulli, independientes entre sí, para una variable discreta. Ésta se caracteriza por ser dicotómica (mutuamente excluyente), como: vivo o muerto, sano o enfermo, cara o sello, aprueba o no aprueba.

Una variable discreta con  $X_i$  de éxitos en  $n$  pruebas independientes, con la misma probabilidad de ocurrencia  $p$  y de no ocurrencia  $q = 1 - p$ , sigue una distribución binomial:  $B(n, p)$ . La distribución binomial se aplica en poblaciones finitas con reemplazo o con procesos estables. Ejemplos: piezas defectuosas, flores fecundadas, insectos muertos o frutos enfermos.

Las propiedades de la distribución binomial son:

Media:  $\mu = np$

Varianza:  $\sigma^2 = npq$



Desviación típica:  $\sigma = \sqrt{npq}$

Coefficiente de sesgo:  $\partial_3 = \frac{q-p}{\sqrt{npq}}$

Los valores  $p(X)$  se obtiene de la tabla binomial.

Un valor  $\partial_3 > abs|0,50|$  indica que la distribución está sesgada hacia la derecha (+) o hacia la izquierda (-). Si el  $\partial_3 < abs|0,50|$  indica que la distribución es simétrica.

Si  $n$  es grande en  $X \sim B(n, p)$  y los valores  $p$  y  $q$  “no son próximos a cero”, se tiene una media  $np$  y una varianza  $npq$  con una fuerte aproximación a la distribución normal:  $N(0,1)$ , por lo tanto, se puede calcular  $p$  a través de  $Z$  que se calcula con la fórmula siguiente:

$$Z = \frac{x - np}{\sqrt{npq}} \quad [26]$$

La prueba binomial se orienta a determinar el número de “éxitos” en una secuencia de  $n$  ensayos con una probabilidad de ocurrencia  $0 < p < 1$ .

La fórmula para calcular la probabilidad binomial es la siguiente:

$$p(X) = \frac{n!(p)^x (q)^{n-x}}{X!(n-X)!} \quad [27]$$

En Excel, la  $p$  binomial se calcula con la función:

=DISTR.BINOM(núm\_éxito;ensayos;prob\_éxito;acumulado)

### Ejercicio 28:

Analizar la germinación de 10 semillas de una especie forestal usando la prueba binomial. Según la ficha técnica de la empresa, la probabilidad de que germine después de seis meses de almacenado es  $p = 0,25$  (Di Rienzo et al., 2009, p. 82).

#### Datos:

$n = 10$  semillas y  $p_{teórica} = 0,25$

**Calcular la media de semillas germinadas conforme a  $p = 0,25$  (ficha técnica)**

$\mu = n p_t \rightarrow \mu = 10 (0,25) = 2,5$  semillas

Las semillas que germinarían, según la ficha técnica, es 2,5 semillas ~ de 2 a 3.

**Calcular la probabilidad de que germinen 7 de las 10 semillas**

$$p(X) = \frac{n!(p)^x (q)^{n-x}}{X!(n-X)!} \quad \rightarrow \quad p(X = 7) = \frac{10!(0,25)^7 (0,75)^{10-7}}{7!(10-7)!} = 0,0031 \sim 0,31\%$$

La probabilidad de que germinen 7 de las 10 semillas es 0,31 %.

**Calcular la probabilidad de que germinen 3 de 10 semillas**

$$p(X = 3) = \frac{10!(0,25)^3 (0,75)^{10-3}}{3!(10-3)!} = 0,2503$$

La probabilidad de que germinen tres de las 10 semillas es  $p = 0,2503 \sim 25\%$ .

**Calcular la coincidencia entre probabilidades observada ( $p$  binomial) y esperada ( $p$  ficha técnica) cuando germinaron 3 de 10 semillas**

La probabilidad de que  $p_{\text{binomial}} = 0,2503$  coincida con el  $p_{\text{teórico}} = 0,25$  de la ficha técnica, se calcula en base a su relación:

$$p = \frac{p(X = 3)}{p_t} = \frac{0,2503}{0,25} = 1 \sim 100 \%$$

**Decisión:**

La probabilidad de que germinen 3 de 10 semillas es 100 %.

**Ejercicio 29:**

Determinar la compatibilidad de un clon de cacao en base al resultado de 30 autopolinizaciones, donde se constató 12 fecundaciones positivas, después de 30 días de la polinización. Un clon se considera autocompatible cuando la autofecundación  $\geq 30$  % (Quinaluisa et al., 2021).

**Datos:**

$$p_{\text{teórica}} = 0,30 \sim 30 \%$$

$$q = 1 - 0,30 = 0,70$$

Autopolinizaciones:  $n = 30$

Autofecundaciones:  $X = 12$  éxitos

**Calcular la media esperada de autofecundaciones en 30 autopolinizaciones:**

$$\mu = np = 30(0,30) = 9 \text{ autofecundaciones (teóricas) en 30 autopolinizaciones (realizadas).}$$

Las 12 autofecundaciones observadas  $> 9$  autofecundaciones teóricas esperadas

**Calcular la autocompatibilidad cuando hay 12 fecundaciones de 30 autopolinizaciones.**

$$p(X = 12) = \frac{30!(0,3)^{12} (0,70)^{30-12}}{12!(30-12)!} = 0,0749 \sim 7,49 \%. \text{ Es la probabilidad de equivocarse.}$$

La probabilidad complementaria corresponde al nivel de confianza:

$$1 - p = 1 - 0,0749 = 0,9251 \sim 92,51 \%$$

**Decisión:**

El clon de cacao es autocompatible con el 92,51 % de confianza.

**Ejercicio 30:**

En la Comunidad “El Manzano”, el 10% de los agricultores produce manzanas verdes. Se aplica una encuesta a 5 productores de manzanas verdes y se plantean algunas cuestiones, que requieren como respuestas:

¿Cuál es la variable aleatoria?: Los productores de manzana

¿Cuál es el tipo de distribución?: Unos producen manzanas verdes y otras variedades distintas (Binomial)

¿Qué valores puede tomar la variable?: Del 0 al 100 %

¿Cuál es la probabilidad de que ninguno produzca manzanas verdes?

Como hay 5 productores de manzanas verdes, que equivales al 10%, entonces habría 50 productores en total.

$$p(PMV = 0) = \frac{PMV}{PMV + POM} = \frac{0}{50} = 0$$

¿Cuál es la probabilidad de que por lo menos haya 4 productores de manzanas verdes entre los 50?

$$p(PMV = 4) = \frac{PMV}{PMV + POM} = \frac{4}{50} = 0,08 \sim 8\%$$

**Decisión:**

La probabilidad de que al menos haya cuatro productores de manzana verde es 8 %.

2.10.2.3 DISTRIBUCIÓN GEOMÉTRICA

Esta distribución trata del número de intentos a realizar en un experimento hasta obtener el primer éxito. Se conoce como geométrica debido a que su forma coincide con el k-ésimo término de la progresión geométrica. La probabilidad de éxito se mantiene constante para todas las repeticiones. Las fórmulas usadas en esta distribución son las siguientes:

$$\mu = \frac{1}{p} \quad [28]$$

$$\sigma^2 = \frac{(1-p)}{p^2} \quad [29]$$

$$p(X = x) = p(1 - p)^{x-1} \quad [30]$$

Dónde:

p = probabilidad de ocurrencia de un suceso

$\sigma^2$  = Varianza

$\mu$  = Promedio

X = Suceso

x = Intentos

**Ejercicio 31:**

En una plantación de rosas, se monitorea la presencia de escamas en un momento determinado. Se conoce que la probabilidad de encontrar escamas es 8 % (p = 0,08). Cálculo de la probabilidad de encontrar escamas en el segundo intento.

**Datos y análisis:**

$$\mu = \frac{1}{p} = \frac{1}{0,08} = 12 \text{ plantas}$$

$$\sigma^2 = \frac{(1-p)}{p^2} = \frac{(0,92)}{0,0064} = 144$$

$$p(X = x) = p(1 - p)^{x-1}$$

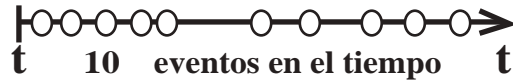
$$p(X = 2) = 0,08(0,92)^{2-1} = 0,0736 \sim 7,4 \%$$

**Decisión:**

La probabilidad de encontrar escamas en el segundo intento es 7,4 %.

2.10.2.4 DISTRIBUCIÓN DE POISSON

La distribución de Poisson expresa la probabilidad de que ocurra un determinado número de eventos, a partir de una frecuencia observada, en un tiempo y espacio específico. La ocurrencia de un evento en el tiempo se esquemática como:



Se usa para calcular la probabilidad de ocurrencia de sucesos raros o impredecibles en variables discretas, cuando la muestra  $n$  es grande y la probabilidad de éxitos  $p$  es pequeña. Ejemplos: insectos muertos.día<sup>-1</sup>, mutaciones observadas al irradiar semillas o peces > 500 g.

Cabe indicar que cuando  $\lambda \geq 10$  la distribución Poisson se aproxima a la normal.

Las condiciones que se debe cumplir para su uso son:  $p < 0,10$  y  $np < 10$ . Por lo tanto,  $p(X = k)$  es la probabilidad de ocurrencia cuando la variable discreta  $X$  toma un valor finito  $k$  que es el número de éxitos por unidad, lambda  $\lambda$  que es la ocurrencia media por unidad (tiempo, volumen, área) y la constante  $e = 2,7118$ .

Las fórmulas básicas usadas en la distribución de Poisson son:

$$\lambda = np = \mu \tag{31}$$

$$\sigma^2 = \lambda \tag{32}$$

$$p(x = k) = e^{-\lambda} \left(\frac{\lambda^k}{k!}\right) \tag{33}$$

Dónde:

$p$  = Probabilidad de éxitos

$n$  = Tamaño de muestra

$\lambda$  = Ocurrencia media por unidad de tiempo, volumen o área

$e = 2,7118$

**Ejercicio 32:**

Calcular el número probable de animales enfermos en una población bovina de 50.000 reses, cuando la enfermedad tiene una probabilidad de ocurrencia:  $p = \frac{1}{10.000}$

**Datos:**

Población (N) = 50.000 animales.

$$p = \frac{1}{10.000} = 0,0001$$

$np = \lambda = 50.000(0,0001) = 5$  animales probablemente estarían enfermos en N= 50.000.

**Calcular la probabilidad de encontrar 3 animales enfermos**

$$(X = k) = e^{-\lambda} \left(\frac{\lambda^k}{k!}\right) \quad \rightarrow \quad (X = 3) = 2,7118^{-5} \left(\frac{5^3}{3!}\right) = 0,1421 \approx 14 \%$$

**Decisión:**

La probabilidad de encontrar tres animales enfermos entre N = 50.000 es del 14 %.



## CAPÍTULO III: INVESTIGACIÓN EXPLORATORIA

La investigación exploratoria se usa cuando el problema no está definido con claridad, en escenarios temporales y espaciales concretos. Se constata la naturaleza de la población pero no se puede muestrear, se desconoce las “causas del problema”, por lo tanto no hay fundamento para formular una hipótesis (Supo y Caverro, 2014, p.41). Ejemplos: (a) una empresa quiere saber cómo reaccionan los consumidores ante la nueva presentación de su producto, (b) un agricultor detecta en su cultivo de naranja una planta con frutos atípicos y desconoce las perspectivas de aprovechamiento; y, (c) un técnico observa una planta marchita en un lote de cultivo y desconoce el agente causal.

En los estudios exploratorios hay flexibilidad para adaptarse a las circunstancias y sienta las bases de estudios descriptivos, relacionales e inclusive experimentales, en una fase temprana. Los diagnósticos agrosociales, la tipificación de los sistemas de producción y la caracterización fenotípica de genotipos tienen un propósito exploratorio, tratan de identificar factores y variables y a partir de esa información, contextualizar un problema.

La investigación exploratoria tiene la finalidad de aclarar la naturaleza del problema y determinar las prioridades de investigación, por eso se conocen también como estudios preliminares. Estos estudios pueden ser a nivel de expertos o a nivel de la población (Supo, 2013, p. 7). A nivel de expertos, no necesariamente investigadores, pueden informar acerca de eventos o hechos de interés de potencial aprovechamiento. A nivel de la población se apoya en la observación directa, clasificación usando escalas nominal u ordinal y conteos; acompañada de revisión de literatura y recopilación de información primaria y secundaria.

La fuente primaria proporciona información nueva y original, resultado de una investigación, mientras que la fuente secundaria es la información elaborada por terceros o mediante la reorganización desde las fuentes primarias. Las encuestas se usan para recopilar información primaria.

Con el avance de la tecnología, las encuestas se pueden realizar en línea o a través de teléfonos móviles. Las entrevistas permiten obtener información detallada sobre temas específicos. Las consultas a expertos y a grupos focales es una técnica usada para valorar la magnitud de la problemática de interés. En la investigación cualitativa, un aspecto clave, luego de definir los objetivos, es la elaboración de los instrumentos de recolección de datos, como las encuestas, que deben validarse con pruebas de confiabilidad<sup>8</sup>. El coeficiente alfa de Cronbach se usa para probar la confiabilidad en matrices de datos provenientes de mediciones ordinales y el Coeficiente de confiabilidad de Rulón se usa para matrices de datos dicotómicos.

Para asegurar la confiabilidad de los instrumentos, Supo (2013) propone seguir 10 pasos: revisar la literatura, explorar conceptos, enlistar temas, formular preguntas, seleccionar jueces, aplicar prueba piloto, evaluar la consistencia, disminuir las preguntas redundantes, reducir las dimensiones e identificar criterios (p. 5).

### 3.1 ELABORACIÓN DE INSTRUMENTOS

Para preparar una encuesta, en los distintos ámbitos del conocimiento, tiene que haber claridad en el objetivo, en las variables requeridas y en la escala de medición. En una indagación sobre los efectos de los pesticidas, las variables deberán referirse al uso de esos

---

<sup>8</sup> En el programa SPSS v22, la fiabilidad de una matriz de datos se determina directamente con la función Analizar, escala y “análisis de fiabilidad”

productos y las escalas de medición tienen que estar en función del grado de profundidad requerida.

Por ejemplo: Variable ordinal efecto sobre la salud de los trabajadores que aplican.

No aplica (0), Bajo (1), Medio bajo (2), Medio (3), Medio alto (4), Muy alto (5).

La generación de variables ordinales, por lo tanto, tienen que estar en armonía con el interés de información. Por ejemplo: generar seis variables ordinales sobre el uso de insecticidas en el cultivo de tomate:

Descripción de la variable	Código
¿Aplica el insecticida Radiant ( <i>spinetoram</i> ) en su cultivo de tomate?	V1
¿Aplica otros insecticidas para controlar <i>negrita</i> en tomate?	V2
¿Al comprar insecticidas, busca el más económico?	V3
¿Al adquirir insecticidas, averigua si tiene efecto positivo en la producción?	V4
¿Conoce la persistencia del insecticida en el producto de consumo?	V5
¿En qué medida el agroquímico usado afecta a la salud de los trabajadores?	V6

En una muestra  $n = 10$  agricultores, propietarios de fincas, se aplica la encuesta (instrumento) de seis preguntas que generará seis respuestas. Estas tienen que organizarse en una tabla de doble entrada para el correspondiente análisis. En el Cuadro 18, se expone, a modo de ejemplo, las respuestas de 10 agricultores en relación a las seis preguntas, usando la escala ordinal de 1 a 5 (ejemplo hipotético).

Cuadro 18. Datos de una encuesta de seis ítems a productores de tomate, usando la escala ordinal de 1 a 5, referida al uso de insecticidas y cálculos básicos

Agricultores (n)	V1	V2	V3	V4	V5	V6	$\Sigma$
1	4	4	4	5	5	2	24
2	3	3	2	4	4	1	17
3	4	4	4	4	5	2	23
4	5	5	5	5	4	3	27
5	4	4	4	5	5	3	25
6	2	3	4	5	3	2	19
7	4	5	4	5	5	2	25
8	2	3	4	4	4	2	19
9	4	3	4	5	5	3	24
10	2	3	4	3	5	2	19
$\Sigma$	34	37	39	45	45	22	<u>222</u>
Varianza de los ítems	1,04	0,61	0,49	0,45	0,45	0,36	$\rightarrow \Sigma S^2 = 3,40$

### 3.2 CONFIABILIDAD DEL INSTRUMENTO

Una condición fundamental en estudios cualitativos es la validación del instrumento de medición, como son los formatos de encuesta y las escalas ordinales o nominales que deben tener el aval de expertos. Los instrumentos de medición, como balanzas, termómetros, peachímetros y otros, con los que se miden datos cuantitativos, tienen que estar calibrados para poder confiar en los datos generados.



La confiabilidad es el grado en que el uso repetido del instrumento produce resultados consistentes. La validez se refiere al grado de exactitud en que el instrumento mide lo que debe medir. La objetividad es el grado de insensibilidad a los sesgos del instrumento. La confiabilidad se calcula con el Coeficiente alfa de Cronbach para escalas ordinales y con el Coeficiente de confiabilidad Rulón para escalas nominales dicotómicas (Supo, 2013, p. 37).

### 3.2.1 COEFICIENTE DE CONFIABILIDAD ALFA DE CRONBACH

Un instrumento de medición debe tener consistencia interna, que significa una aproximación a la validación del constructo, basado en la cuantificación de la correlación entre los ítems de la encuesta. El Coeficiente  $\alpha$  de Cronbach estima la consistencia interna en datos ordinales, cuando hay de tres a veinte ítems (Oviedo y Campo, 2005, p. 572).

La fórmula para calcular el coeficiente  $\alpha$  de Cronbach es la siguiente:

$$\alpha = \frac{k}{k-1} \left[ 1 - \frac{\sum S_i^2}{S_T^2} \right] \quad [34]$$

Dónde:

$\alpha$  = Coeficiente alfa de Cronbach

K = Número de ítems (preguntas)

$S_i^2$  = Varianza de los ítems individuales (pregunta)

$S_T^2$  = Varianza de la suma de las respuestas por sujeto

La regla de decisión recomendada por Corral (2009, p. 244) es:

Coeficiente de 0,81 a 1,00	Muy alta confiabilidad
Coeficiente de 0,61 a 0,80	Alta confiabilidad
Coeficiente de 0,41 a 0,60	Moderada confiabilidad
Coeficiente de 0,21 a 0,40	Baja confiabilidad
Coeficiente de 0,01 a 0,20	Muy baja confiabilidad

#### Ejercicio 33:

Determinar la confiabilidad del instrumento usando el Coeficiente  $\alpha$  de Cronbach, referente al impacto del uso de insecticidas, para K = 6 ítems, con varianza de la suma de las respuestas de 10,36 y la varianza de los ítems individuales de 3,40.

#### Datos y análisis:

K = 6 ítems o preguntas

$$\sum S_i^2 = 3,40$$

$$S_T^2 = \frac{(24)^2 + (17)^2 + \dots + (19)^2 - FC}{N} = 10,36$$

#### Cálculo del Coeficiente $\alpha$ de Cronbach:

$$\alpha = \frac{k}{k-1} \left[ 1 - \frac{\sum S_i^2}{S_T^2} \right]$$

$$\alpha = \frac{6}{5} \left[ 1 - \frac{3,40}{10,36} \right] = 0,806$$

#### Decisión:

La confiabilidad del instrumento es muy alta con un  $\alpha$  Cronbach de 0,806.

### 3.2.2 COEFICIENTE DE CONFIABILIDAD DE RULÓN

El Coeficiente de confiabilidad de Rulón se basa en la comparación de dos mitades de datos de la matriz de encuestas dicotómicas. Se basa en que la diferencia entre mitades tiene sólo al error aleatorio y no necesariamente varianzas iguales (Corral, 2009, p. 240).

La fórmula para calcular el Coeficiente de confiabilidad de Rulón es:

$$r_{cm} = 1 - \frac{S_d^2}{S_T^2} \quad [35]$$

Dónde:

$r_{cm}$  = Coeficiente de confiabilidad de las mitades

$S_d^2$  = Varianza de las diferencias de los puntajes de los grupos pares e impares

$S_T^2$  = Varianza total de las respuestas afirmativas

$n$  = Encuestados

$V$  = Preguntas o ítems

$$S_T^2 = \frac{\sum T^2 - \frac{(\sum T)^2}{n-1}}{n} \quad [36]$$

$$S_d^2 = \frac{\sum d^2 - \frac{(\sum d)^2}{n-1}}{n} \quad [37]$$

La regla de decisión propuesta por Corral (2009, p. 244), es la siguiente:

0,81 - 1,00	Confiabilidad muy alta
0,61 - 0,80	Confiabilidad alta
0,41 - 0,60	Confiabilidad moderada
0,21 - 0,40	Confiabilidad baja
0,01 - 0,20	Confiabilidad muy baja

#### Ejercicio 34:

Determinar el Coeficiente de confiabilidad de Rulón con los resultados de una encuesta de seis preguntas dicotómicas: 1 = afirmativas y 0 = negativas, aplicada a 12 agricultores.

#### Datos:

La matriz de una encuesta de seis preguntas dicotómicas (ítems) a 12 agricultores (ítems), se exponen en el Cuadro 19.

Cuadro 19. Matriz de resultados de una encuesta de seis variables con respuestas dicotómicas

Encuestados	V1	V2	V3	V4	V5	V6	Total respuestas afirmativas	
1	1	0	0	1	1	1	4	Varianza del total de respuestas afirmativas:  $S_T^2 = 0,629$
2	1	0	1	1	1	1	5	
3	0	1	1	1	1	1	5	
4	1	1	1	0	1	1	5	
5	1	1	1	0	1	0	4	
6	0	1	1	0	1	1	4	
7	1	1	1	1	0	0	4	
8	1	1	1	1	1	0	5	
9	0	1	1	1	0	0	3	
10	1	0	0	1	1	1	4	
11	1	0	1	0	1	0	3	
12	1	0	1	1	0	0	3	

**Separación de pares e impares:**

Mitad impar	V1	V2	V3	V4	V5	V6	Total respuestas afirmativas	Mitad par	V1	V2	V3	V4	V5	V6	Total respuestas afirmativas
1	1	0	0	1	1	1	4	2	1	0	1	1	1	1	5
3	0	1	1	1	1	1	5	4	1	1	1	0	1	1	5
5	1	1	1	0	1	0	4	6	0	1	1	0	1	1	4
7	1	1	1	1	0	0	4	8	1	1	1	1	1	0	5
9	0	1	1	1	0	0	3	10	1	0	0	1	1	1	4
11	1	0	1	0	1	0	3	12	1	0	1	1	0	0	3

**Cálculos de las varianzas:**

Mitad impar: Total de respuestas afirmativas	Mitad par: Total de respuestas afirmativas	Diferencias (impar – par)
4	5	-1
5	5	0
4	4	0
4	5	-1
3	4	-1
3	3	0
Varianza de las diferencias: $S_d^2$		0,300
Varianza total: $S_T^2$		0,629

**Cálculo del Coeficiente de confiabilidad de Rulón:**

$$r_{cm} = 1 - \frac{S_d^2}{S_T^2} \rightarrow r_{cm} = 1 - \frac{0,30}{0,629} = 0,523$$

**Decisión:**

El Coeficiente de Rulón,  $r_{cm} = 0,523$  indica que el instrumento tiene una moderada confiabilidad.

### 3.3 DIAGNÓSTICO

Un diagnóstico se orienta a explorar un hecho, indagando una situación, no necesariamente problemática, en procura de identificar las oportunidades de mejora continua o de las causas probables de un problema que requiera investigación descriptiva. Cuando se indaga en multiambientes se refiere a estudios trasversales y si se proyecta conocer la evolución de un hecho en el tiempo, se trata de estudios longitudinales.

La observación de los síntomas de deficiencias de elementos minerales en los cultivos, la valoración de los síntomas de enfermedades fungosas en las hojas o frutos, el diagnóstico de las pudriciones de las raíces, son estudios exploratorios. Las observaciones de los síntomas de enfermedades foliares (exploratorio), por ejemplo, conllevan a hacer presunciones de agentes causales. Mientras no se realice la identificación taxonómica y las pruebas pertinentes, quedan solo como presuntas causas.

Generalmente, los estudios exploratorios se distinguen de los descriptivos porque no tienen una hipótesis de investigación, se desconocen los factores causales de un hecho y solo se enfocan a tratar de comprenderlo, sentando las bases para estudios en niveles superiores. El diagnóstico de la situación de salud de una población humana, vegetal o animal tiene el propósito de indagar y conocer una enfermedad basado en los signos y síntomas.

### 3.3.1 PROBABILIDAD DE SUCESOS SEGÚN LA LEY DE LAPLACE

La ley de Laplace permite calcular la probabilidad ( $p$ ) de un suceso cuando los resultados posibles tienen la misma posibilidad de aparecer en  $n$  casos (Montes, 2007, p. 4). Esta regla se sintetiza en la fórmula para calcular  $p$  de ocurrencia de un evento  $A$ . El valor  $p$  se interpreta como proporción o porcentaje (Badii et al., 2021, p. 2).

La fórmula general para calcular  $p$  es:

$$p(A) = \frac{A}{n} \quad [38]$$

Donde:

$p(A)$  = probabilidad de ocurrencia de eventos  $A$

$A$  = Casos probables

$n$  = Casos posibles

Para proyectar el valor  $p$  hacia una población ( $N$ ), se usa la fórmula:

$$p(X = A) = \frac{A}{n}(N) \quad [39]$$

Dónde:

$p$  = Proporción de ocurrencia de  $A$

$A$  = Eventos esperados

$n$  = Total de eventos

$p(X = A)$  = Proporción esperada del evento  $A$  en una población conocida

#### **Ejercicio 35:**

Supóngase que un investigador recorre un lote de 650 árboles de una especie forestal e inspecciona aleatoriamente 50 árboles ( $n = 50$ ), observándose un hecho: hay seis árboles con daños en el tronco a causa de insectos no identificados. ¿Cuántos árboles podrían tener la incidencia de insectos afectando al tronco en toda la población?

#### **Datos y análisis:**

$N = 650$ ,  $n = 50$ ,  $A = 6$

$p = \frac{A}{n} = (\text{Árboles con frutos de color atípico})/(\text{Árboles de la muestra})$

$p = \frac{6}{50} = 0,12 \approx 12\%$

#### **Interpretación:**

Si  $p = 12\%$ , la incidencia de la plaga en  $N = 650$  se calcula como:

$p(A) = \frac{A}{n}(N) = \frac{6}{50}(650) = 0,12(650) = 78$  árboles

#### **Decisión:**

Es probable que 78 de los 650 árboles estén afectados por la plaga.

**Ejercicio 36:**

Un experto en fruticultura visita por casualidad un lote de una especie frutal, de población  $N = 6500$  árboles. Al inspeccionar una muestra de  $n = 150$  árboles, observó seis árboles con frutos de coloración atípica ( $A = 6$ ). ¿Cuál es la proporción de árboles que tendrían frutos de coloración atípica? ¿Cuántos árboles en la población tendrían frutos con coloración atípica?

**Datos:**

$$n = 150 \text{ y } A = 6$$

**Análisis:**

¿Cuál es la proporción teórica de árboles con frutos de coloración atípica?

$$p = \frac{A}{n} = \frac{\text{Árboles con frutos con coloración atípica}}{\text{Árboles de la muestra}} = \frac{6}{150} = 0,04 \sim 4,0\%$$

¿Cuántos árboles con frutos de coloración atípica es probable encontrar en  $N = 6500$ ?

$$p(X) = \frac{A}{n}(N) = \frac{6}{150}(6500) = 0,04(6500) = 260 \text{ árboles}$$

**Decisión:**

En  $N = 6500$  árboles habrá 260 árboles con frutos de coloración atípica.

3.3.2 SENSIBILIDAD Y ESPECIFICIDAD

La Sensibilidad y Especificidad son términos usados para referirse a la exactitud de los resultados de las pruebas de diagnóstico sanitario<sup>9</sup>. La Sensibilidad es la probabilidad de clasificar correctamente a un individuo enfermo (Martínez et al., 2020). La Especificidad es la probabilidad de que el resultado de la prueba sea negativo si realmente no hay enfermedad.

La fórmula para calcular la sensibilidad es la siguiente:

$$SEN = \frac{VP}{TEC} (100) \quad [40]$$

Dónde:

SEN = Sensibilidad en por ciento

VP = Verdaderos positivos (individuos enfermos) concordantes entre valoración visual y el resultado de laboratorio

TEC = Total de individuos enfermos confirmados en el laboratorio

La fórmula para calcular la especificidad es la siguiente:

$$ESP = \frac{VN}{TSC} (100) \quad [41]$$

Dónde:

ESP = Especificidad en por ciento

VN = Verdaderos negativos (individuos sanos) concordantes entre valoración visual y el resultado de laboratorio

TSC = Total de individuos sanos confirmados en el laboratorio

<sup>9</sup> En medicina, el término positivo, en un diagnóstico, significa que el individuo está enfermo. Se sugiere adaptar este concepto en las otras ciencias que estudian poblaciones vegetales y animales.

**Ejercicio 37:**

En una muestra aleatoria de 25 plantas se diagnosticó en el campo, frutos con antracnosis. El especialista visualmente indicó que 15 plantas tienen frutos enfermos y 10 plantas tienen frutos sanos. Se tomó un fruto por planta para verificar en el laboratorio y el resultado fue: 16 frutos enfermos y 9 frutos efectivamente sanos. Calcular la sensibilidad y especificidad.

**Datos y análisis:**

Los datos de la prueba de diagnóstico relacionados con el diagnóstico visual en campo realizado por el experto, se expone en el Cuadro 20.

*Cuadro 20. Resultados del diagnóstico en el laboratorio contrastado con el diagnóstico visual en campo*

		Prueba de diagnóstico en laboratorio		Total
		Enfermas	Sanas	
Diagnóstico visual en campo	Enfermas	11	4	15
	Sanas	5	5	10
Total		16	9	25

$$SEN = \frac{VP}{TEC} (100) = \frac{11}{16} (100) = 69 \%$$

$$ESP = \frac{VN}{TSC} (100) = \frac{5}{9} (100) = 56 \%$$

**Interpretación:**

La sensibilidad como la probabilidad de encontrar individuos enfermos, basado en la valoración visual de los síntomas en campo y verificado en el laboratorio fue SEN = 69 %. La especificidad que es la probabilidad de verificar que hay individuos sanos en base a pruebas de laboratorio fue ESP = 56 %.

**Decisión:**

La sensibilidad fue 69 % y la especificidad 56 %.

**3.3.3 FRECUENCIAS DE HECHOS NOVEDOSOS**

En los estudios exploratorios se buscan hechos novedosos, en variables cualitativas o cuantitativas. Los análisis de suelos, encuestas agro-sociales o diagnósticos agroturísticos, agroambientales, fitosanitarios y zoonosanitarios, son estudios de este tipo.

**Ejercicio 38:**

Un investigador observa un lote de cultivo de naranja, detectando plantas con frutos atípicos, indaga y confirma que el hecho es recurrente. Explorar los atributos diferenciadores en los frutos, en una zona de interés agrícola.

**Variables y escalas de medición:**

Se definen los atributos diferenciadores de los frutos y las escalas de medición. Por ejemplo: definición de variables morfológicas que se mide con escalas ordinales. Análisis de los puntajes parciales y totales de las calificaciones.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

Si las variables son del tipo “mayor es mejor”, el puntaje más alto posibilita identificar el árbol de interés genético. Una opción segura para analizar los datos es el uso de valores escalares al máximo (EAM) (Cuadro 21). Se complementa con la elaboración de un croquis de la parcela (Figura 24) y recopilación de la información biológica y geográfica.

*Cuadro 21. Matriz de puntajes de atributos de interés económico en árboles de naranja*

Variables	Escala ordinal	Árbol 1	Árbol 2	Árbol 3
Forma del fruto	1 Desuniforme	--	1	--
	2 Uniforme	2	--	2
Maduración del fruto	1 Desuniforme	--	--	--
	2 Uniforme	2	2	2
Tamaño del fruto	1 Pequeño	--	--	--
	2 Mediano	2	--	--
	3 Grande	--	3	3
Color del fruto	1 Poco atractivo	--	1	--
	2 Atractivo	2	--	2
Estado general del fruto	1 Inadecuado	--	--	--
	2 Medianamente adecuado	2	2	--
	3 Adecuado	--	--	3
Suma de puntajes:		10	9	12
Escarlar Al Máximo (%):		83	75	100

Hileras	Columnas									
	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	●	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	●	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0
10	0	0	0	●	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0

Nota: Árboles 1(3-3), 2(4-10) y 3(6-5)

*Figura 24. Identificación de los árboles de interés genético a nivel de campo*

**Decisión:**

El árbol 3, ubicado en la posición 6-5 tiende a diferenciarse por el fruto. Se recomienda introducirlo a un banco de germoplasma.

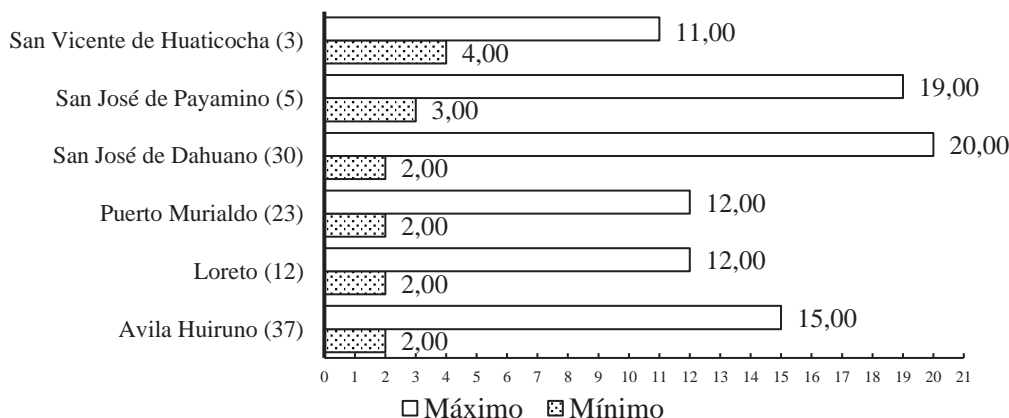


**Ejercicio 39:**

En base al análisis de suelos de 110 fincas del cantón Loreto, Orellana, explorar la situación del contenido de azufre e indicar si se justifica un proyecto de investigación en esta temática.

**Datos:**

Los resultados del análisis químico del azufre (S) se expresan en forma cuantitativa (ppm) y cualitativa (bajo, medio y alto). El patrón de azufre (S) indica que: si  $S < 10$  ppm el contenido es bajo, si  $S$  varía 10 - 19 ppm el contenido es medio y si  $S > 19$  ppm es alto. Los valores máximos y mínimos de S, desglosado por parroquias, indican que prevalecen los suelos con niveles bajos de S, excepto en San José de Dahuano (Gráfico 10).



Nota: Patrón S: bajo < 10 ppm, medio 10 - 19 ppm y alto > 19 ppm

Gráfico 10. Contenidos máximos y mínimos de azufre en suelos de 110 fincas del cantón Loreto

**Decisión:**

En Loreto, en las fincas cafetaleras, prevalecen suelos con contenidos bajos y medios de azufre. Se justifica un estudio sobre esta temática.

**Ejercicio 40:**

Representar las frecuencias absolutas de los contenidos de macro y micronutrientes de 150 fincas cafetaleras de las provincias Orellana y Sucumbíos.

**Datos:**

Los datos analizados corresponden a 150 muestras de suelos dedicados al cultivo de café robusta, de las provincias Orellana y Sucumbíos.

**Resultados:**

Los elementos nitrógeno (N), azufre (S), fósforo ( $P_2O_5$ ), boro (B), magnesio (Mg), zinc (Zn), calcio (Ca) y potasio ( $K_2O$ ) se encuentran en niveles entre bajo y medio (colores rojo y amarillo, respectivamente) (Gráfico 11).

Esto significa que se requiere de la adición de esos nutrimentos para elevar la productividad de los cafetales. El hierro (Fe) está en niveles altos. El manganeso (Mn) está en niveles altos en 127 de las 150 muestras.

En análisis de cada uno de los elementos conlleva a la posibilidad de realizar ensayos monofactoriales, bifactoriales o multifactoriales que, replicándose en los principales dominios de recomendación, configurarían una agenda de investigación en fertilización en la región amazónica del norte del Ecuador.

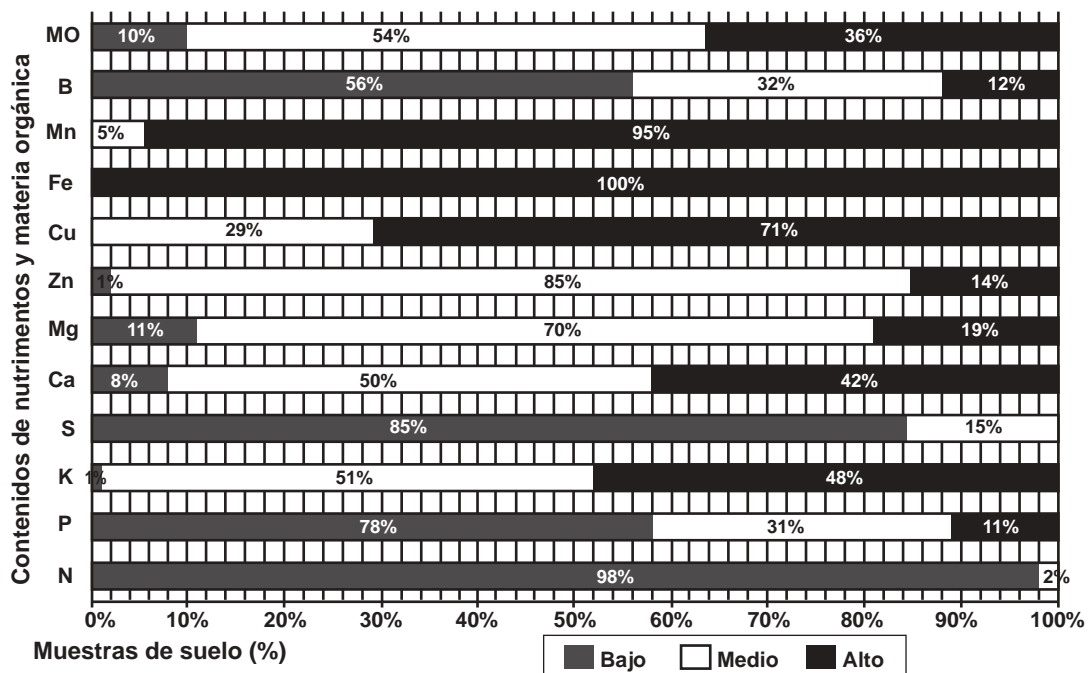


Gráfico 11. Caracterización química de los suelos de la Amazonía norte del Ecuador

**Decisión:**

Los elementos N, S, P<sub>2</sub>O<sub>5</sub>, B, Mg, Zn, Ca y K<sub>2</sub>O requieren de atención preferencial en la Amazonía norte del Ecuador. Resulta imperativo el estudio de fertilización.

**3.4 PRUEBA DE TAMIZAJE**

Las pruebas de tamizaje (en inglés: screening) son chequeos rápidos para valorar probables efectos de un factor (p.e.: agroquímicos) sobre un problema específico poco conocido o desconocido. En la agricultura hay tamizajes de quelatos para medir su efecto sobre los síntomas visuales de deficiencias de minerales en las plantas; pruebas de ingredientes activos sobre enfermedades foliares o radicales, entre otras.

Al investigador, por ejemplo, puede preocuparle la falta de uniformidad en el cultivo, constatando plantas faltantes (“fallas”). ¿Hay argumento para afirmar que se trata de una pudrición radical causada por nemátodos?. Las acciones del investigador serían: observar y explorar las plantas, tomar una porción de suelo de la zona radical afectada, colocar las muestras de suelo contaminado en macetas, sembrar granos de especies sensibles como fréjol o maíz, aplicar al substrato soluciones de plaguicidas en concentraciones conocidas y medir la respuesta en la sobrevivencia después de 15 días.

**Ejercicio 41:**

Explorar si alguna de las sustancias usadas como desinfectantes de semillas tienen efecto positivo sobre la germinación de una especie sensible, sembrada en macetas conteniendo suelo contaminado con fitopatógenos.

**Datos:**

La relación porcentual entre semillas germinadas y sembradas (germinación %). La variables es del tipo “mayor es mejor”. Los datos del tamizaje se indican en el Cuadro 22.

Cuadro 22. Prueba de tamizaje de agroquímicos midiendo la germinación de una especie sensible en substratos contaminados

Sustancias	Semillas sembradas (SS)	Semillas germinadas (SG)	Relación SG/SS	Germinación (%)	Escalar al máximo (%)
A	50	25	0,50	50	69
B	50	27	0,54	54	75
C	50	15	0,30	30	42
D	50	12	0,24	24	33
E	50	8	0,16	16	22
F	50	13	0,26	26	36
G	50	16	0,32	32	44
H	50	24	0,48	48	67
I	50	36	0,72	72	100
J	50	19	0,38	38	53
Testigo (T0)	50	3	0,06	6	8

**Decisión:**

El 72 % de germinación de la semilla se logró usando la sustancia I, que representa el valor escalar más alto (EAM = 100 %).

3.4.1 PROBABILIDADES CONDICIONALES Y MARGINALES

En los estudios exploratorios y descriptivos se pueden calcular las probabilidades condicionales y marginales. Una probabilidad es condicional cuando un evento A ocurre cuando también sucede otro evento B. Una probabilidad es marginal cuando la distribución trata un subconjunto de datos de la variable aleatoria (Martínez et al., 2020).

**Ejercicio 42:**

Calcular las probabilidades condicionales de una muestra de 300 plantas donde hubieron 240 sanas y 60 enfermas. Calcular las probabilidades marginales si del 240 sanas, 160 recibieron un tratamiento preventivo y 80 no fueron tratadas; además, de las 60 enfermas, 20 fueron tratadas y 40 no recibieron el tratamiento preventivo.

**Cálculo de las probabilidades condicionales:**

Probabilidad de tener plantas enfermas:  $p(\text{enfermas}) = \frac{60}{300} = 0,20 \sim 20\%$

Probabilidad de tener plantas sanas:  $p(\text{sanas}) = \frac{240}{300} = 0,80 \sim 80\%$

**Cálculo de las probabilidades marginales:**

$p(\text{sanas con tratamiento}) = \frac{160}{240} = 0,67 \sim 67\%$

$p(\text{sanas sin tratamiento}) = \frac{80}{240} = 0,33 \sim 33\%$

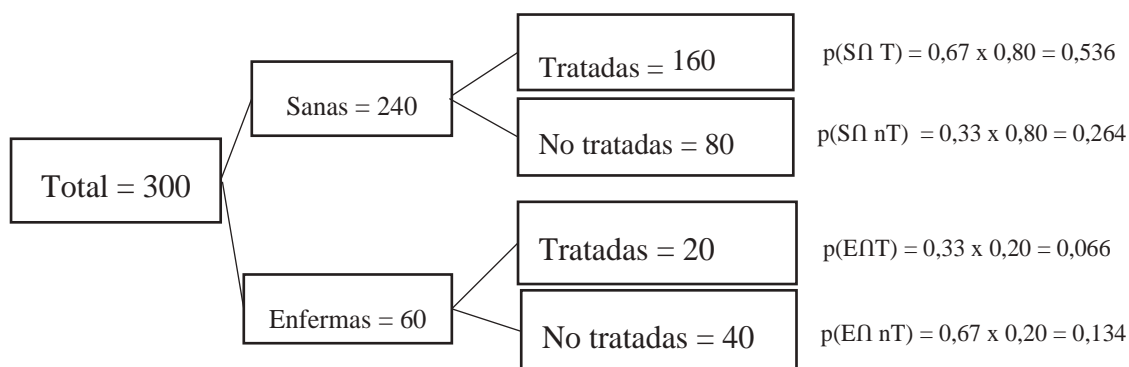
$p(\text{enfermas con tratamiento}) = \frac{20}{60} = 0,33 \sim 33\%$

$p(\text{enfermas sin tratamiento}) = \frac{40}{60} = 0,67 \sim 67\%$

**Cálculo de probabilidad de intersección de sucesos en plantas “tratadas y enfermas”:**  
 $p(E \cap T) = 0,33 \times 0,20 = 0,066 \sim 6,6 \%$

**Cálculo de probabilidad de intersección de sucesos en plantas “no tratadas y enfermas”:**  
 $p(E \cap nT) = 0,67 \times 0,20 = 0,134 \sim 13,4 \%$

La probabilidad condicional de plantas enfermas,  $p = 0,20$  se verifica como la suma de las probabilidades en las intersecciones:  $6,6 \% + 13,4 \% = 20 \%$  (Gráfico 12).



Fuente: Martínez et al. (2020)

Gráfico 12. Probabilidades condicionales y marginales de dos sucesos fitosanitarios

**Decisión:**

Se evidencia que hay diferencias entre plantas “tratadas y enfermas” ( $p = 0,066$ ) y “no tratadas y enfermas” ( $p = 0,134$ ).

**3.4.2 PRUEBA EXACTA DE FISHER**

La prueba exacta de Fisher es una técnica no paramétrica usada para determinar la asociación entre dos variables independientes, buscando respuestas a la cuestión: ¿la proporción de una variable es diferente respecto del valor que adquiere la otra variable? (Martínez et al., 2020).

Esta prueba se usa cuando las muestras son de reducido tamaño, aun cuando haya, en cualquier celda un dato con valor 0. En esto radica la diferencia con la prueba  $\chi^2$  que se usa para analizar muestras grandes. Las variables dicotómicas se organizan en tablas de contingencia  $2 \times 2$  (Cuadro 23).

Cuando una variable se mide en más de dos clases (tablas de contingencia  $3 \times 2$ ,  $4 \times 2$  u otras), se usan extensiones del método, agrupando columnas o hileras y analizando las distintas categorías por pares.

La fórmula de la prueba exacta de Fisher es la siguiente:

$$p = \frac{(A + B)! (C + D)! (A + C)! (B + D)!}{N! A! B! C! D!} \quad [42]$$

Dónde:

$p$  = Probabilidad de Fisher

A, B, C, D = Valores de los conteos para la tabla de contingencia  $2 \times 2$

N = Suma de las 4 celdas en la tabla  $2 \times 2$

Cuadro 23. Modelo de una tabla de contingencia 2 x 2

	Variable de respuesta		Total
	+	-	
Tratamientos	Grupo 1	A      B	A + B
	Grupo 2	C      D	C + D
	Total	A + C      B + D	N

Los niveles de significación, generalmente usados son:  $\alpha = 0,05$  y  $\alpha = 0,01$ .

Las hipótesis estadísticas se formulan como:

$H_0$ : La variable de respuesta no depende del factor en estudio ( $p > 0,05$ )

$H_1$ : La variable de respuesta si depende del factor en estudio ( $p < 0,05$ )

La regla de decisión, con  $\alpha = 0,05$ , es la siguiente:

Si  $p$  calculada  $> \alpha = 0,05$ : Aceptar  $H_0$

Si  $p$  calculada  $\leq \alpha = 0,05$ : Aceptar  $H_1$

**Ejercicio 43:**

Analizar 12 tesis de pregrado en agronomía, de la biblioteca de una universidad, y determinar si la calidad de los experimentos depende del diseño aplicado (Ruiz et al., 2015, p. 250).

**Datos y análisis:**

Los datos de las 12 tesis de pregrado sobre temas de agronomía, analizados, se indican en el Cuadro 24.

Cuadro 24. Calidad de los experimentos relacionados con la selección del diseño

	Calidad de los experimentos		Tesis analizadas
	Mala	Buena	
Incorrecta selección del diseño	1	2	3
Correcta selección del diseño	4	5	9
Total de tesis analizadas	5	7	12

**Nivel de significación:**

$\alpha = 0,05 \sim 5\%$

**Hipótesis estadísticas:**

$H_0$ : La calidad de los experimentos no depende de la selección del diseño

$H_1$ : La calidad de los experimentos si depende de la selección del diseño

**Regla de decisión:**

Si  $p$  calculada de Fisher  $> \alpha = 0,05$ : Aceptar  $H_0$

Si  $p$  calculada de Fisher  $\leq \alpha = 0,05$ : Aceptar  $H_1$

**Cálculos:**

	Calidad de los experimentos		
	Mala	Buena	Tesis analizadas
Incorrecta selección del diseño	A = 1	B = 2	A + B = 3
Correcta selección del diseño	C = 4	D = 5	C + D = 9
Total de tesis analizadas	A + C = 5	B + D = 7	N = 12

$$p = \frac{(A + B)! (C + D)! (A + C)! (B + D)!}{N! A! B! C! D!} \rightarrow p = \frac{(1 + 2)! (4 + 5)! (1 + 4)! (2 + 5)!}{12! 1! 2! 4! 5!}$$

$$p = \frac{(3)!(9)!(5)!(7)!}{12!1!2!4!5!} \rightarrow p = \frac{(6)(362880)(120)(5040)}{(479001600)(1)(2)(24)(120)!} = 0,477 \sim 47,7 \%$$

**Interpretación:**

$p$  calculada de Fisher = 0,477 >  $\alpha = 0,05$ : Aceptar  $H_0$

**Decisión:**

La diferencia de 2 tesis entre 12 no resultó estadísticamente significativa. La calidad de los experimentos no depende de la selección del diseño.

**Ejercicio 44:**

En 20 novillas Brahman blanco inseminadas se comparó las técnicas de “inseminación profunda” y tradicional, obteniéndose los resultados de la variable preñez. Determinar si las técnicas de inseminación influyen sobre la preñez de las vacas (Monsalve y Zumaeta, 2011).

**Datos y análisis:**

Los datos de preñez de las novillas tratadas con los métodos de inseminación profunda y tradicional se indican en el Cuadro 25.

Cuadro 25. Estado de preñez de las novillas en función de los métodos de inseminación

	Estado de preñez de las novillas		Total de novillas inseminadas
	Preñadas	No preñadas	
Inseminación profunda	9	1	10
Inseminación tradicional	4	6	10
Total de novillas inseminadas	13	7	20

**Nivel de significación:**

$\alpha = 0,05 \sim 5\%$

**Hipótesis estadísticas:**

$H_0$ : La preñez no depende del método de inseminación.

$H_1$ : La preñez si depende del método de inseminación.

**Regla de decisión:**

Si  $p$  calculada de Fisher >  $\alpha = 0,05$ : Aceptar  $H_0$

Si  $p$  calculada de Fisher  $\leq \alpha = 0,05$ : Aceptar  $H_1$

**Cálculos:**

	Estado de preñez		Total
	Preñadas	No preñadas	
Inseminación profunda	A = 9	B = 1	A + B = 10
Inseminación tradicional	C = 4	D = 6	C + D = 10
Total	A + C = 13	B + D = 7	N = 20

$$p = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{N!A!B!C!D!} \rightarrow p = \frac{(9+1)!(4+6)!(9+4)!(1+6)!}{20!9!1!4!6!}$$

$$p = \frac{(10)!(10)!(13)!(7)!}{20!9!1!4!6!} \rightarrow p = 0,0271 \sim 2,71\%$$

**Interpretación:**

$p$  calculada de Fisher = 0,0271 <  $\alpha = 0,05$ : Aceptar  $H_1$ .

**Decisión:**

La preñez de las vacas depende del método de inseminación con 97,3 % de confianza.

**Ejercicio 45:**

En 15 trabajadores agrícolas que manipulan plaguicidas se realizó un análisis de su salud integral, agrupándolos por edades: menores de 30 años y mayores de 30 años. Determinar si la salud de los trabajadores se relaciona con los grupos de edades.

**Datos y análisis:**

Los datos de los trabajadores agrícolas, sanos y enfermos, por edades, se expone en la tabla de contingencia 2 x 2 (Cuadro 26).

Cuadro 26. Estado de salud de los trabajadores agrícolas clasificados por edades

	Trabajadores agrícolas		Total
	Enfermos	Sanos	
Menores de 30 años	1	5	6
Mayores de 30 años	9	0	9
Total	10	5	15

**Nivel de significación:**

$\alpha = 0,05 \sim 5\%$

**Hipótesis estadísticas:**

$H_0$ : La salud de los trabajadores que manipulan plaguicidas no depende de la edad

$H_1$ : La salud de los trabajadores que manipulan plaguicidas si depende de la edad

**Regla de decisión:**

Si  $p$  calculada de Fisher >  $\alpha = 0,05$ : Aceptar  $H_0$

Si  $p$  calculada de Fisher  $\leq \alpha = 0,05$ : Aceptar  $H_1$

**Cálculos:**

	Trabajadores agrícolas		Total
	Enfermos	Sanos	
Menores de 30 años	A = 1	B = 5	A+B = 6
Mayores de 30 años	C = 9	D = 0	C+D = 9
Total	A+C = 10	B+D = 5	N = 15

$$p = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{N!A!B!C!D!} \rightarrow p = \frac{(6)!(9)!(10)!(5)!}{15!1!5!9!0!} = 0,002 \sim 0,2\%$$

$$p = 0,002 \sim 0,2\%$$

**Interpretación:**

$p = 0,002 < \alpha = 0,05$ : Aceptar  $H_1$ . La condición de salud de los trabajadores agrícolas que manipulan plaguicidas si depende de la edad.

**Decisión:**

La salud de los trabajadores agrícolas que manipulan plaguicidas si depende de la edad ( $p = 0,002$ ).

### 3.5 CASOS DE INVESTIGACIÓN EXPLORATORIA

**Caso 1:**

Construcción de un índice de satisfacción para clientes de supermercados mexiquenses. Una investigación exploratoria (Arroyo et al., 2008, p. 67).

El desarrollo de índices de satisfacción requiere de una cuidadosa elaboración del instrumento de medición para que resulte útil. Una encuesta aplicada entre los clientes de una cadena de supermercados de cobertura regional, por ejemplo, puede aportar información necesaria para realizar un análisis factorial y verificar la multidimensionalidad del concepto de satisfacción.

**Caso 2:**

Uso de plaguicidas y discapacidad intelectual en estudiantes de escuelas municipales, Provincia de Talca, Chile (Muñoz, 2010).

Con el objetivo de describir la relación existente entre la cercanía de escuelas municipales a predios agrícolas que utilizan plaguicidas y la presencia de escolares con discapacidad intelectual, se realizó un estudio exploratorio. Se utilizaron datos de integración escolar de niños con discapacidad intelectual de la provincia Talca (Chile). El resultado indicó que hay una mayor proporción de estudiantes con discapacidad intelectual en las zonas donde se aplican plaguicidas.

**Caso 3:**

Estudio ¿Qué significa ser médico para un estudiante del primer año de medicina? (Angulo y Corujo, 2008, p. 9).

El impulso vocacional conduce al estudiante a decidirse por la práctica de la medicina. El objetivo del estudio fue establecer el significado de lo que es vocación y el ser médico, para un estudiante que inicia la carrera de Medicina. El instrumento de recolección



de información fue una encuesta anónima y voluntaria que analizó usando frecuencia absoluta, porcentaje y tabla de contingencia. Para el 42,2 %, la vocación indicó el deseo innato de ejercer una profesión y para 40,4 % ser médico significaba ser una buena persona.

**Caso 4:**

Diagnóstico de la situación agrosocial de la comunidad agrícola indígena de Kashaama, con el fin de describir y analizar las principales variables sociales y agronómicas que caracterizan las unidades de producción agrícola (Olivares y Franco, 2015).

El grupo de participantes estuvo constituido por 40 productores agrícolas seleccionados al azar. A través de una encuesta se recolectó información demográfica, socioeconómica y agroecológica de los participantes.

**Resultados de la encuesta:**

Problemas	%
Contaminación del agua para consumo humano	23
Deterioro de los servicios de salud, asistencia médica y servicio ambulatorio	22
Ausencia de maquinaria agrícola	21
Ausencia de planes de financiamiento agrícola	15
Deterioro de los recursos naturales	14
Deterioro de la vialidad, vivienda, electricidad, educación y seguridad	5

Al explorar el modo de vida de los indígenas se determinó que tienen una cosmovisión, un sistema social, creencias, costumbres y conocimientos propios lo cual configura su patrimonio inmaterial. Esta situación merece la atención del Estado, de la academia y de los científicos sociales.

## CAPÍTULO IV: INVESTIGACIÓN DESCRIPTIVA

En los estudios descriptivos se analiza las características de poblaciones o muestras, profundizando los tipos exploratorios y relacionales, cuantificando la información e incluso haciendo comparaciones. En este nivel investigativo ya se tiene claridad acerca de las relaciones entre factores pero no se llega a su manipulación. Las comparaciones pueden ser temporales y/o espaciales, seleccionando variables concretas, aspecto que lo diferencia de los estudios exploratorios.

El análisis de datos requiere de varias etapas para su cabal comprensión, que empieza con el análisis de frecuencias y los cálculos de las medidas de tendencia central, de dispersión y de forma. En el diseño se debe considerar el tiempo y el espacio. Si un estudio en un diagnóstico socioeconómico se recopila datos sobre el ingreso de las familias de los agricultores y los analiza, está en nivel exploratorio; si clasifica los ingresos en función de salarios mínimos vitales y calcula valores máximos, mínimos y promedios, está en nivel descriptivo univariado. Si, además, el investigador tomó datos del grado de escolaridad de los jefes de hogar y busca definir la intensidad de la asociación, se trata de un estudio relacional. El análisis simultáneo de dos o más variables es estadística multivariada, que puede enmarcarse en los niveles descriptivo y hasta proyectivo.

En la Figura 25, se exponen los tipos de estudios, según el número de registros de datos a lo largo del tiempo (longitudinales) o los espacios territoriales en donde se replican (estudios transversales. Estos estudios, además, pueden ser cualitativos o cualitativos, unifactoriales o multifactoriales, descriptivos o relacionales y experimentales o predictivos.es, predictivos

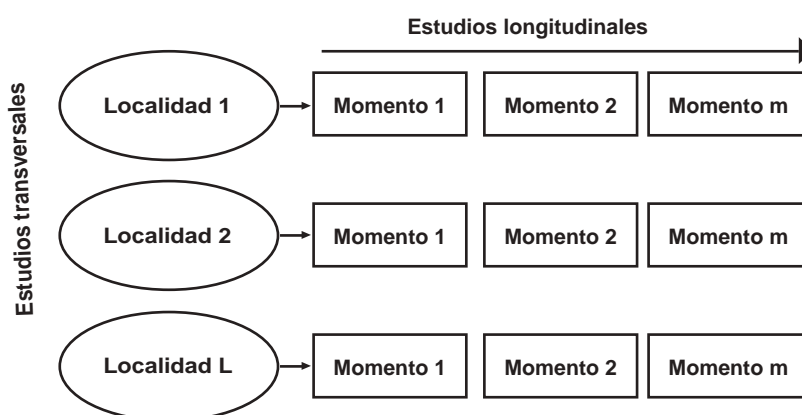


Figura 25. Estudios a lo largo del tiempo y en diversos escenarios

### 4.1 ANÁLISIS DE FRECUENCIAS

El análisis de frecuencias es una técnica utilizada para examinar la distribución de las muestras o poblaciones, implica contar la frecuencia de ocurrencia de cada categoría en un conjunto de datos y luego se procede a resumir los datos de los conteos en forma de tabla o de gráfico. El análisis de frecuencias ayuda a identificar patrones, tendencias o desequilibrios, en los datos, también permite comparar diferentes grupos o subconjuntos de datos y evaluar la asociación entre variables categóricas mediante tablas de contingencia o pruebas estadísticas como la prueba de Chi-cuadrado.

Este análisis tiene el propósito de sintetizar o reducir las series de datos en tablas de frecuencias de una variable (un sentido) o de dos variables (dos sentidos), en representaciones gráficas y en estadísticos (García, 2011, p. 26). El término frecuencia se refiere a la cantidad de veces que se repite un determinado valor en una serie de datos, en valores absolutos o relativos (Little y Hills, 1987, p. 23). Algunos conceptos usados en el análisis de frecuencias (Universidad de Granada, 2016), son los siguientes:

**Frecuencia absoluta.**- Es el número de datos que corresponden a cada clase. Se representa gráficamente en un histograma o en un polígono de frecuencias.

$$\sum n_i = n_1 + n_2 + \dots + n_k = N \quad [43]$$

Dónde:

$n_1$  = Frecuencia en la clase 1,  $n_2$  = Frecuencia en la clase 2 ...  $n_k$  = Frecuencia en la clase k  
 $N$  = Total de datos

**Frecuencia absoluta acumulada.**- Se calcula sumando las frecuencias desde el primer intervalo hasta el último valor de frecuencias acumuladas que corresponderá a la suma total.

$$N_i = n_1 + n_2 + \dots + n_i \quad [44]$$

**Frecuencia relativa.**- Se obtiene dividiendo cada frecuencia de cada clase por el total de observaciones. Se representa en proporciones o porcentajes.

$$f_i = \frac{n_i}{N} (100) \quad [45]$$

**Frecuencia relativa acumulada.**- Se calcula sumando las frecuencias relativas desde el primer intervalo hasta que el valor último que corresponde al 100 % .

$$F_i = f_1 + f_2 + \dots + f_i \quad [46]$$

**Serie de datos.**- Es el conjunto de datos de una variable perteneciente a una población o una muestra específica, tal como se recolectó o en forma ordenada,

**Serie ordenada de datos.**- Es el conjunto de datos organizados de mayor a menor o viceversa, de una variable concreta, correspondiente a una muestra o una población.

**Clase.**- Es la categoría de datos sin traslapes con la cual se agrupan los datos. Una serie de datos se divide en clases. El número de clases depende del tamaño de la serie de datos.

**Marca de clase.**- Es el punto medio de cada intervalo de clase.

**Número de clases.**- Se refiere a las categorías en que se distribuye la serie de datos.

**Intervalo de clase.**- Es la magnitud de una clase. Se han propuesto varias fórmulas para calcular el intervalo de clase. Salazar (2010, pp. 29-30) plantea calcular como:  $INC = \sqrt{n}$ ; y si  $n > 30$ , el  $INC = 1 + 3,222 \cdot \log n$ . Se destaca que es el investigador quien lo prefija de acuerdo a la naturaleza del estudio, recomendándose usar la fórmula siguiente:

$$INC = \frac{M - m}{N} \quad [47]$$

Dónde:

INC = Intervalo de clase

M = Valor máximo en la serie de datos

m = Valor mínimo en la serie de datos

N = Número de clases proyectada para el análisis de frecuencias

**Ejercicio 46:**

En una serie ordenada de datos de producción (g.planta<sup>-1</sup>), obtenida en una muestra de 30 plantas, realizar el análisis de frecuencias.

**Datos:**

131 135 138 141 145 147 151 151 154 155  
 155 160 160 160 161 162 164 165 165 167  
 168 172 174 175 179 183 185 188 197 199

**Análisis:**

El valor máximo es 199 g.planta<sup>-1</sup> y el valor mínimo es 131 g.planta<sup>-1</sup>, por lo tanto, el rango o diferencia es 68 g.planta<sup>-1</sup>. El intervalo de clase se proyecta a partir de la relación:  $68/7 = 9,71 \sim 10$  g, siendo el INC de 10 g.planta<sup>-1</sup> (Cuadro 27).

Cuadro 27. Análisis de frecuencias de la variable producción en una muestra de 30 plantas

Clases (g.planta <sup>-1</sup> )	Control	Frecuencia Absoluta (FA)	Frecuencia Absoluta Acumulada (FAA)	Frecuencia Relativa (FR)	Frecuencia Relativa Acumulada (FRA)
≤140	III	3	3	10	10
141 - 150	III	3	6	10	20
151 - 160	IIIIIIII	8	14	27	47
161 - 170	IIIIIIII	7	21	23	70
171 - 180	IIII	4	25	13	83
181 - 190	III	3	28	10	93
≥191	II	2	30	7	100
Σ		30		100	

**Decisión:**

Hay dos plantas con producciones > 190 g.planta<sup>-1</sup>, por tanto tienen potencial genético aprovechable.

**Ejercicio 47:**

En una serie de datos de altura de planta (cm), en una muestra de 51 plantas, se calculó los estadígrafos: valor máximo = 390 cm, valor mínimo = 250 cm, rango =140 cm. Realizar el análisis de frecuencias e identificar las plantas de menor altura.

**Datos y análisis:**

Los datos de altura de 51 plantas de café, organizados en siete clases y el análisis de frecuencias, se detallan en el Cuadro 28.

HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA

Cuadro 28. Análisis de frecuencias de altura (AP) en una muestra de 51 plantas de café

Clases: AP (cm)	Frecuencia Absoluta (FA)	Frecuencia Absoluta Acumulada (FAA)	Frecuencia Relativa (FR) %	Frecuencia Relativa Acumulada (FRA) %
≤ 270	8	8	15,7	15,7
271 – 290	17	25	33,3	49,0
291 – 310	11	36	21,6	70,6
311 – 330	4	40	7,8	78,4
331 – 350	9	49	17,6	96,1
351 – 370	1	50	2,0	98,0
371 – 390	1	51	2,0	100,0
Σ	51		100	

**Análisis:**

Las frecuencias absolutas o relativas se pueden representar en un histograma. El número de clases lo define el investigador, sugiriéndose un número impar. Si el objetivo es identificar plantas de menor altura, entonces hay 8 plantas que reúnen esta condición (Gráfico 13). En n = 51 datos de altura de planta, hay 8 con valores menores que 270 cm.

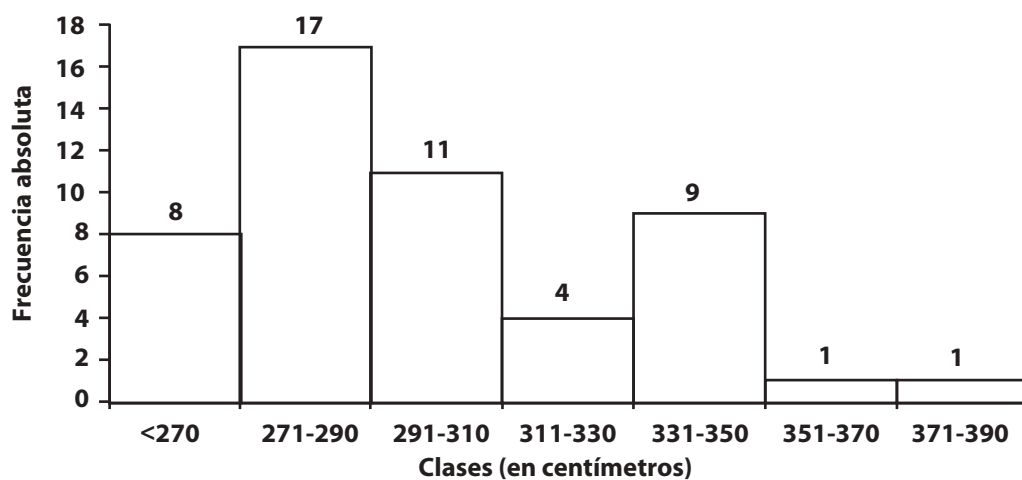


Gráfico 13. Frecuencias absolutas de altura de planta en 51 cafetos de 60 meses de edad

**Decisión:**

Hay 8 plantas con altura < 270 cm que pueden ser de interés para mejora genética.

**Ejercicio 48:**

Realizar el análisis de frecuencias usando las medias de altura de planta de 32 clones de café robusta e identificar el genotipo de porte más bajo.

**Datos y análisis:**

El análisis de frecuencias de la variable altura de planta de 32 clones de café y la identificación de los genotipos según el porte, se indica en el Cuadro 29.

*Cuadro 29. Frecuencias absolutas (FA) y relativas (FR) de los promedios de altura de planta en 32 clones de café robusta y clasificación cualitativa según el porte*

Clases: AP (cm)	FA	FR	Conteo	Clasificación según el porte	Clones de café robusta			
≤ 260	1	3	I	Porte bajo	COF-0 06			
261 - 290	7	22	IIIIII		CON-ERB-01	MA-N-01	MA-N-02	NP-2044
					NP-3056	NP-4024	NP-3072	
				Porte medio	RP-S 04	NP-3013	Brasilia-1	NP-3018
291 - 320	16	50	IIIIIIIIIIIIIIIIII		COF-0 02	COF-0 04	COF-0 05	COF-LR-01
					ETP-3753-13	COF-O 07	CON-ETP-06	ETP-3752-6
					NP-2024	CON-ETP- 01	COF-O-01	CON-ETP-04
321 - 350	4	13	IIII	Porte alto	CON-ETP-05	CON-ETP-03	ETP-3756-14	RP-S-01
351 - 380	3	9	III		ETP-3564-2	JO-P-01	CON-ETP-02	
≥ 381	1	3	I		COF-0-03			
Σ	32	100						

**Interpretación:**

El clon COF-06 tiene el porte más bajo ( $\leq 260$  cm). Si el objetivo es seleccionar clones de porte bajo, de alta producción, con resistencia a la roya y buena calidad física y sensorial, habría que ampliar los análisis hacia esas otras variables de interés.

**Decisión:**

El clon COF-06 tiene una altura de planta  $\leq 260$  cm, que se categoriza de porte bajo<sup>10</sup>.

**Ejercicio 49:**

Comparar la altura de planta (cm) de tres clones de café robusta, a los 60 meses de edad, en base a los promedios y rangos.

**Datos y análisis:**

En el Cuadro 30, se indican los datos de altura de planta, de los tres clones de café robusta a ser compararse. El objetivo es identificar los de menor altura y menor variabilidad. El COF-06 tiene un promedio de 277 cm, una altura máxima de 290 cm y un rango de 30

<sup>10</sup> El clon COF 06 fue seleccionada por el porte bajo, además posee las características de alta producción por planta, buena calidad física del grano y excelente calidad de taza.

cm. Los otros dos clones, además de tener promedios más altos, lo valores máximos y los rangos también son más altos que COF-06.

Cuadro 30. Comparación de las medias de altura de planta de tres clones de café robusta

Plantas	Altura de planta (cm) de tres clones de café robusta			
	COF-06	CON-ERB 01	NP-2024	Total
Plantas	18	16	17	51
Media (cm)	277	298	333	
Máximo (cm)	290	340	390	
Rango (cm)	30	70	110	

**Resultado:**

El clon COF 06 tiene una altura de planta de 277 cm, el CON-ERB-01 una media de 298 cm y el NP-2024 de 333 cm. El rango en COF 06 resultó menor que los otros clones.

**Decisión:**

El clon COF-06 tiene un promedio de altura de planta de 277 cm y un rango de 30 cm que se considera de porte bajo y de menor dispersión.

#### 4.2 DIAGRAMA DE “TALLOS Y HOJAS”

El diagrama de “tallos y hojas” (en inglés: Stem & leaf), separa el dato en dos partes: la unidad, decena o centena (dígito o dígitos del lado izquierdo) que constituye el tallo y el dígito de la derecha que corresponde a las hojas (Salazar, 2010, p. 44). Esta técnica permite visualizar la serie de datos como un histograma de números.

**Ejercicio 50:**

Realizar el diagrama de “tallo y hojas” usando la serie ordenada de 30 plantas que corresponde a la variable producción (g.planta<sup>-1</sup>).

**Datos:**

131	135	138	141	145	147	151	151	154	155
155	160	160	160	161	162	164	165	165	167
168	172	174	175	179	183	185	188	197	199

**Análisis:**

El diagrama de “tallo y hojas” es un histograma numérico (Gráfico 14), que contiene información idéntica al histograma de barras.

Tallo (decenas)	Hojas (unidades)									
13	1	5	8							
14	1	5	7							
15	1	1	4	5	5					
16	0	0	0	1	2	4	5	5	7	8
17	2	4	5	9						
18	3	5	8							
19	7	9								

Gráfico 14. Diagrama de “tallo y hojas” para datos de producción en una muestra de 30 plantas

**Resultados:**

La mayor frecuencia absoluta de datos se ubica en la decena 16, donde se localizan las medidas de tendencia central. Se denota una tendencia a la distribución normal.

**Decisión:**

En la muestra de 30 plantas se identificaron dos con producciones de 197 y 199 g.planta<sup>-1</sup>, considerados altos que podrían ser de interés para fitomejoramiento.

### 4.3 MEDIDAS DE TENDENCIA CENTRAL

Las medidas de tendencia central son: media, mediana y moda (Alder & Roessler, 1976, pp. 32-44). Las medias más usadas son: simple, aritmética, geométrica, armónica y ponderada.

#### 4.3.1 MEDIA SIMPLE

La media simple ( $\bar{Y}_s$ ) es el promedio de los valores extremos de la serie de datos. Es muy útil cuando se desea una media referencial de manera rápida.

La fórmula de la media simple es:

$$\bar{Y}_s = \frac{\text{Maximo} + \text{Minimo}}{2} \quad [48]$$

**Ejercicio 51:**

Calcular la media simple de la siguiente serie de datos en cm:

131	135	138	141	145	147	151	151	154	155
155	160	160	160	161	162	164	165	165	167
168	172	174	175	179	183	185	188	197	199

**Análisis:**

$$\bar{Y}_s = \frac{\text{Maximo} + \text{Minimo}}{2} \quad \rightarrow \quad \bar{Y}_s = \frac{131 + 199}{2} = \frac{330}{2} = 165$$

**Decisión:**

$$\bar{Y}_s = 165 \text{ cm}$$

#### 4.3.2 MEDIA ARITMÉTICA

La media aritmética ( $\bar{Y}$ ) es la medida central característica de una serie de datos. La fórmula usada para el cálculo de la media aritmética es la siguiente:

$$\bar{Y} = \frac{\sum Y_i}{n} \quad [49]$$

La fórmula indicada es equivalente a:  $\bar{Y} = \frac{1}{n} \sum_i^n Y_i$

En el programa Excel, la media aritmética se calcula con la función:

=promedio(valor1:valor2)



**Ejercicio 52:**

Calcular la media aritmética de la variable altura de planta, correspondiente al clon COF-06, donde la suma de los valores:  $\Sigma Y = 4990$  y  $n = 18$  plantas.

**Datos:**

$$\Sigma Y = 4990 \text{ y } n = 18 \text{ plantas}$$

**Análisis:**

$$\bar{Y} = \frac{1}{n} \sum_i^n Y_i \rightarrow \bar{Y} = \frac{1}{18} (4990) = 277 \text{ cm}$$

**Decisión:**

$$\bar{Y} = 277 \text{ cm}$$

4.3.3 MEDIA GEOMÉTRICA

La media geométrica ( $\bar{Y}_G$ ) es la raíz n-ésima del producto de todos los factores. Se usa para promediar datos donde hay valores extremos que distorsionarían la media aritmética (Salazar, 2010, p. 56).

La fórmula para calcular la media geométrica es la siguiente:

$$\bar{Y}_G = \sqrt[n]{(Y_1)(Y_2) \dots (Y_n)} \quad [50]$$

**Ejercicio 53:**

Calcular la media geométrica para altura de planta de dos clones de café robusta: COF-06 (277 cm) y NP-2024 (333 cm).

**Datos:**

$$\text{COF-06} = 277 \text{ cm}$$

$$\text{NP-2024} = 333 \text{ cm}$$

**Análisis:**

$$\bar{Y}_G = \sqrt{(277)(333)} = \sqrt{92241} = 304 \text{ cm}$$

**Decisión:**

$$\bar{Y}_G = 304 \text{ cm. Nota: la media aritmética es 305 cm}$$

**Ejercicio 54:**

Estimar el “dato faltante” de la “parcela perdida” en un ensayo en diseño en BA de 5 tratamientos en 4 repeticiones. La parcela perdida corresponde al tratamiento 2 en la repetición 4. Los promedios sin corregir de T2 es 50 g.planta<sup>-1</sup> y de R4 es 65 g.planta<sup>-1</sup>.

**Datos:**

$$\text{T2 sin corregir} = 50 \text{ g.planta}^{-1}, \text{ R4 sin corregir} = 65 \text{ g.planta}^{-1}.$$

**Análisis:**

$$\bar{Y}_G = \sqrt{(50)(65)} = \sqrt{3250} = 57,0 \text{ g.planta}^{-1}$$

**Decisión:**

$$\text{Dato estimado para la parcela perdida} = 57,0 \text{ g.planta}^{-1}.$$

#### 4.3.4 MEDIA ARMÓNICA

La media armónica ( $\bar{Y}_A$ ) de una serie de datos es igual al recíproco de la media aritmética (Salazar, 2010, p. 59). Esta sirve para calcular medias de tiempos y movimientos.

La fórmula para calcular la media armónica es la siguiente:

$$\bar{Y}_A = \frac{n}{\sum \frac{1}{Y_i}} \quad [51]$$

#### Ejercicio 55:

Se tienen dos reservorios de 5000 m<sup>3</sup> de agua y para llenarlo se dispone de las bombas A y B. Con la bomba A se bombea 1000 m<sup>3</sup>.hora<sup>-1</sup> y con la B, 500 m<sup>3</sup>.hora<sup>-1</sup>. ¿Cuál es la media armónica?

#### Datos y análisis:

Capacidad de bombeo (m<sup>3</sup>.hora<sup>-1</sup>): A = 1000 y B = 500. Reservorio = 5000 m<sup>3</sup>.

$$\bar{Y}_A = \frac{n}{\sum \frac{1}{Y_i}} \quad \rightarrow \quad \bar{Y}_A = \frac{2}{\frac{1}{10} + \frac{1}{5}} = \frac{2}{0,3} = 6,67 \text{ horas} \sim 6 \text{ horas con } 40 \text{ minutos}$$

#### Decisión:

La media armónica es  $\bar{Y}_A = 6$  horas y 40 minutos, tiempo requerido para llenar los dos reservorios con las dos bombas.

#### 4.3.5 MEDIA PONDERADA

Media ponderada es el promedio dinámico de valores cambiantes en el tiempo como selecciones de alternativas, índices de precios al consumidor, precios de venta y ganancias.

La fórmula para calcular la media ponderada es la siguiente:

$$\bar{X}_p = \frac{X_1p_1 + X_2p_2 + \dots + X_np_n}{p_1 + p_2 + \dots + p_n} \quad [52]$$

Dónde:

$X_i$  = Datos por componente de la medición

$p_i$  = Ponderaciones de la importancia de los componentes (pesos relativos)

#### Ejercicio 56:

Para ingresar a una universidad se requiere una calificación promedio ponderada  $\geq 5,0/10$  de la nota del bachillerato con peso de 60 % y del examen de admisión con peso de 40 %. Un estudiante tuvo una nota de bachillerato de 5,66/10 y de 4,01/10 en el examen de admisión. ¿Puede ingresar a la universidad?

#### Datos y análisis:

La nota de bachillerato =  $\frac{5,66}{10}$  y examen de admisión =  $\frac{4,01}{10}$ . Se requiere una puntuación ponderada  $\geq \frac{5,00}{10}$  para poder ingresar a la universidad

$$\bar{X}_p = \frac{X_1p_1 + X_2p_2 + \dots + X_np_n}{p_1 + p_2 + \dots + p_n} \quad \rightarrow \quad \bar{X}_p = \frac{5,66(60) + 4,01(40)}{60 + 40} = \frac{339,6 + 160,4}{100} = \frac{500}{100} = 5,0$$

#### Decisión:

$\bar{X}_p = 5,0$  puntos. Si puede ingresar a la universidad.

**Ejercicio 57:**

Determinar el índice de sostenibilidad de las prácticas de conservación de agua (A), suelo (S) y biodiversidad (B), en un proyecto donde se valoraron con la escala ordinal de 1 a 10 (1 = no deseable y 10 = muy deseable) y obtuvo los siguientes puntajes: A = 6, S = 4 y B = 6. La importancia o pesos por componente, fueron: A = 50 %, S = 20 % y B = 30%.

**Datos y análisis:**

A = 6 con peso de 50 %, S = 4 con peso de 20 % y B = 6 con peso de 30%.

$$\bar{X}_p = \frac{X_1p_1 + X_2p_2 + \dots + X_np_n}{p_1 + p_2 + \dots + p_n} \rightarrow \bar{X}_p = \frac{6(50) + 4(20) + 6(30)}{50 + 20 + 30} = \frac{300 + 80 + 180}{100} = \frac{560}{100} = 5,6$$

**Decisión:**

$\bar{X}_p = 5,6$  ¿Qué hacer en esta situación de sostenibilidad?

4.3.6 MEDIANA

La mediana ( $\tilde{X}$ ) es el valor central en una serie ordenada de datos, cuando n es impar y el promedio de los valores centrales cuando n es par. Un atributo importante de la  $\tilde{X}$  es la “no sensibilidad” a los valores extremos (Salazar, 2010, p.67).

En Excel, la mediana se calcula con la función:

=MEDIANA(número1, [número2], ...)

**Ejercicio 58:**

Calcular la mediana de la serie de datos siguiente: 4, 5, 7, 9, 9, 10, 12, 12, 12, 14, 15, 20, 55

**Datos y análisis:**

4, 5, 7, 9, 9, 10, 12, 12, 12, 14, 15, 20, 55

El valor central es 12

**Decisión:**

$\tilde{X} = 12$

**Ejercicio 59:**

En una serie de n = 30, los datos varían de 131 a 199 g.planta<sup>-1</sup>, siendo los valores centrales 161 y 162, ¿Cuál es la mediana?.

**Datos y análisis:**

$$\tilde{X} = \frac{\text{valor central 1} + \text{valor central 2}}{2} = \frac{161 + 162}{2} = 161,5 \text{ g.planta}^{-1}$$

**Decisión:**

$\tilde{X} = 161,5 \text{ g.planta}^{-1}$ .

4.3.7 MEDIANA EN DATOS AGRUPADOS

Cuando los datos están agrupados en clases, la mediana se calcula con la fórmula:

$$\tilde{X} = L_i + \frac{\frac{n}{2} - FAA_{(i-1)}}{FA} (INC) \quad [53]$$

Dónde:

n = Tamaño de muestra (número de observaciones)

$\frac{n}{2}$  = Indicativo de la clase en que debe ubicarse la  $\tilde{X}$

$L_i$  = Límite inferior de la clase donde está teóricamente la  $\tilde{X}$

FA = Frecuencia absoluta (FA) de la clase donde está ubicada teóricamente la  $\tilde{X}$

$FAA_{(i-1)}$  = Frecuencia absoluta acumulada (FAA), donde está la  $\tilde{X}$  restado en una posición

INC = Intervalo de clase.

**Ejercicio 60:**

Calcular la  $\tilde{X}$  para la variable producción (g.planta<sup>-1</sup>) a partir de los datos agrupados que indicados a continuación:

**Datos y análisis:**

En el Cuadro 31, se exponen los datos de producción (g.planta<sup>-1</sup>) de n = 30 plantas, agrupadas en siete clases. La clase central corresponde a la decena de 161 a 170 g.planta<sup>-1</sup>.

Cuadro 31. Análisis de frecuencias de la variable producción g.planta<sup>-1</sup> en una muestra de 30 plantas

Clases (g.planta <sup>-1</sup> )	Frecuencia Absoluta (FA)	Frecuencia Absoluta Acumulada (FAA)	Frecuencia Relativa (%)	Frecuencia Relativa Acumulada (%)
131 - 140	3	3	10	10
141 - 150	3	6	10	20
151 - 160	8	14	27	47
161 - 170	7	21	23	70
171 - 180	4	25	13	83
181 - 190	3	28	10	93
191 - 200	2	30	7	100
$\Sigma$	30		100	

$$\tilde{X} = L_i + \frac{\frac{n}{2} - FAA_{(i-1)}}{FA} (INC)$$

$$n = 30$$

$$\frac{n}{2} = \frac{30}{2} = 15. \text{ Indica que } \tilde{X} \text{ se ubica en la FAA mayor que 15. En este caso, en 161 - 170}$$

$$L_i = \text{Límite inferior de la clase donde está la } \tilde{X} = 161$$

$$FA = \text{Frecuencia absoluta (FA) de la clase donde está teóricamente la } \tilde{X} = 7$$

$$FAA_{(i-1)} = \text{Frecuencia absoluta acumulada (FAA) donde teóricamente se posiciona la } \tilde{X} \text{ menos una posición} = 14$$

$$INC = \text{Intervalo de clase o ancho de clase} = 10$$

$$\tilde{X} = 161 + \frac{15 - 14}{7} (10) = 162,4 \text{ g.planta}^{-1}$$

**Decisión:**

$$\tilde{X} = 162,4 \text{ g.planta}^{-1}.$$

#### 4.3.8 MODA

La moda ( $M_o$ ) es el valor más frecuente en la serie ordenada de datos. Es el único estadístico aplicable a datos cualitativos y cuantitativos. Las características de las modas son: Una serie de datos puede tener más de una moda (depende de las subpoblaciones), solo toma en cuenta las observaciones más frecuentes, indica el valor característico de la distribución, se puede localizar con facilidad y puede no haber moda (Salazar, 2010, p. 76).

#### Ejercicio 61:

Indicar la moda en la serie de 30 datos de la variable peso en  $g.planta^{-1}$ .

#### Datos:

131	135	138	141	145	147	151	151	154	155
155	<u>160</u>	<u>160</u>	<u>160</u>	161	162	164	165	165	167
168	172	174	175	179	183	185	188	197	199

#### Decisión:

$$M_o = 160 \text{ g.planta}^{-1}.$$

### 4.4 MEDIDAS DE DISPERSIÓN

Las medidas de dispersión son la expresión de la variabilidad de la serie de datos alrededor del valor central (Alder & Roessler, 1976, pp. 44-57) Las medidas de dispersión son los cuantiles, la varianza, la desviación típica, el error típico, el coeficiente de variación, la variación y el intervalo de confianza.

En todos los casos, hay que tener en cuenta las unidades del sistema internacional y la pertinencia de los decimales y sus magnitudes. La desviación, estándar, error típico, intervalo de confianza se expresan en las mismas unidades del promedio. La desviación estándar de la producción, por ejemplo, se puede medir en toneladas, kilos, gramos o miligramos, pero la varianza en unidades cuadrática (producción<sup>2</sup>). El Coeficiente de variación y variación relativa se expresan en porcentaje.

#### 4.4.1 ANÁLISIS DE CUANTILES

Un análisis de frecuencias y un histograma resulta insuficiente para describir un fenómeno agrosocial o ambiental, hay que complementar usando otras técnicas como los análisis de cuantiles. Los resultados conllevan a identificar las distorsiones, por lo que lo consideran una medida de posición (Salazar, 2010, pp. 80-92). Las mediciones en cuantiles, más importantes son: cuantiles, quintiles, deciles y percentiles.

Cuantiles.- La serie de datos se divide en cuatro partes.

Quintiles.- La serie de datos se divide en cinco partes

Deciles.- La serie de datos se divide en 10 partes

Percentiles.- La serie de datos se divide en 100 partes.

Un histograma que representa de los ingresos anuales de 829 hogares de agricultores<sup>11</sup>, (Gráfico 15) proporciona información en seis clases (pudo haber más o menos clases). Los resultados indican que hay cinco familias con ingresos anuales > \$5000 y 76 hogares con ingresos < \$1000.

<sup>11</sup> Información del proyecto: “Reconversión de pequeñas fincas cafetaleras en unidades autosostenibles”. Consejo Cafetalero Nacional. 2012.

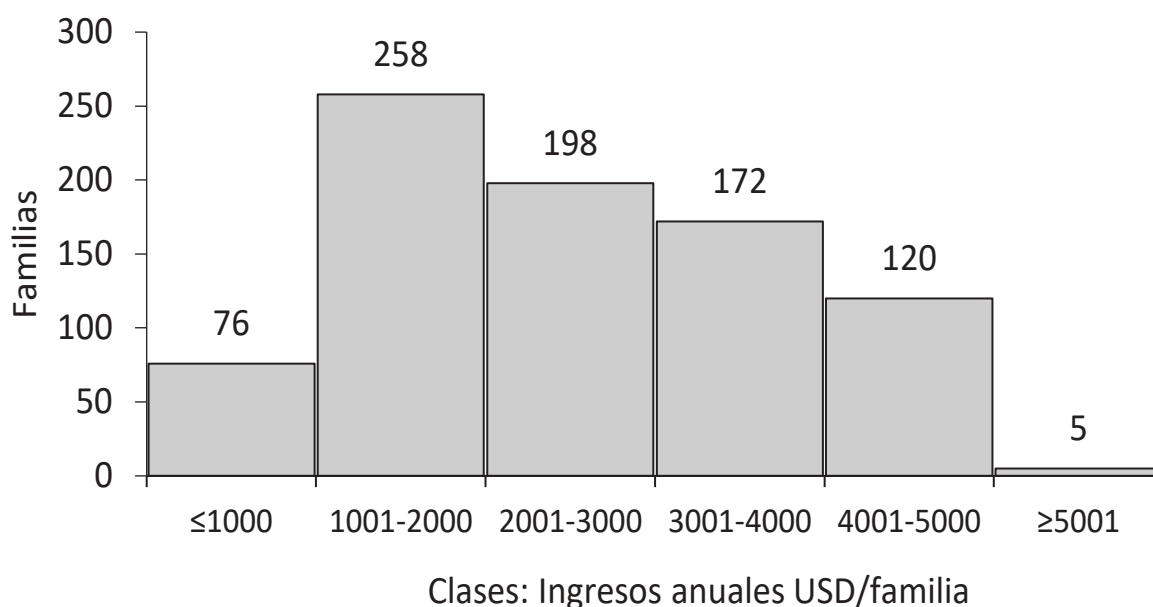


Gráfico 15. Clasificación de los ingresos anuales de 829 familias de agricultores

#### 4.4.1.1 CUARTILES

El análisis de cuartiles es la segmentación de la serie de datos en cuatro partes iguales. Cada sección contiene el 25 % de los datos. Por simple inspección se determina los valores mínimo (LI) y máximo (LS).

Se calcula los puntos que limitan los cuartiles: Q1, Q2 y Q3. El punto Q2 corresponde a la mediana general de la serie de datos. El punto Q1 es la mediana parcial del segmento LI → Q2 y el punto Q3 es la mediana parcial del segmento Q2 → LS.

Cabe indicar que el análisis de la serie de datos se inicia con el cálculo del rango (Rn) y de los valores proporcionales (teóricos) de los segmentos. El rango intercuartílico (RI) y la desviación intercuartílica (DQ), ayudan a interpretar la proporcionalidad de los cuartiles, estadísticos que se calculan con las siguientes fórmulas:

$$RI = Q_3 - Q_1 \quad [54]$$

$$DQ = \frac{RI}{2} \quad [55]$$

En Excel, el análisis de cuartiles se realiza con la función: =cuartil(valor1:valor2,q). Donde el valor q corresponde al número del cuartil

#### Ejercicio 62:

Mediante el análisis de cuartiles, valorar los ingresos anuales de 829 familias de agricultores.

#### Datos y análisis:

En el Cuadro 32, se expone el análisis de cuartiles para la variable ingresos anuales en una muestra de 829 hogares de caficultores.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 32. Análisis de cuartiles de los ingresos anuales de 829 hogares de caficultores*

Cuartiles	Proporción	Frecuencia acumulada	Ingresos anuales (\$)
Límite inferior (LI) de la serie	0	1	265
Cuartil 1 (Q1)	0,25	207	1550
Cuartil 2 (Q2 = mediana)	0,50	415	2410
Cuartil 3 (Q3)	0,75	622	3501
Límite superior (LS) de la serie	1,00	829	5019

**Resultados:**

La mediana Q2 = \$2410. El segmento LI → Q1 varía de \$265 a \$1550 (diferencia = \$1020). La distancia Q1 → Q2 varía de \$1550 a \$ 2410 (diferencia = 860). La distancia Q2 → Q3 varía de \$2410 a \$3501 (diferencia = 1091). La distancia Q3 → LS varía de \$3501 a \$5019 (diferencia = \$1518).

El rango de la serie de datos:  $R_n = 5019 - 265 = \$4754$

Segmento proporcional =  $\frac{RI}{4} = \frac{4754}{4} = \$1188,5$

El rango intercuartílico:  $RI = Q3 - Q1 = 3501 - 1550 = \$1951$ .

La desviación intercuartílica:  $DQ = \frac{RI}{2} = \frac{1951}{2} = \$975,5$

**Interpretación:**

Las diferencias entre los cuatro segmentos cuartílicos no son proporcionales. El segmento proporcional (\$1188,5) es mayor que la desviación intercuartílica (\$975,5). La diferencia LS – Q3 = \$1518 de la cola derecha, es mayor que los otros segmentos.

**Decisión:**

El ingreso familiar anual más bajo fue de \$265. En el cuartil más alto, los ingresos varían de \$3501 a \$5019.año<sup>-1</sup>. Se evidencia un ligero sesgo hacia la cola derecha.

**4.4.1.2 QUINTILES**

La serie de datos se distribuye en cinco segmentos iguales que equivale a proporciones del 20 %, comparándose las magnitudes entre segmentos llamados quintiles. Con esta información se puede analizar en detalle, como en el caso referido a los cuartiles.

**Ejercicio 63:**

Valorar los ingresos anuales en quintiles de 829 familias de caficultores.

**Datos:**

En el Cuadro 33, se expone el análisis de quintiles de los ingresos anuales, en una muestra de 829 hogares de caficultores.

*Cuadro 33. Análisis de quintiles de los ingresos anuales en una muestra de 829 hogares de caficultores*

Quintiles	Proporción	Frecuencia acumulada	Ingresos anuales (\$)	Incremento interquintil (\$)
Límite inferior (LI) de la serie	0	1	265	--
Quintil 1	0,20	166	1.364	1099
Quintil 2	0,40	332	2.000	636
Quintil 3	0,60	497	2.776	776
Quintil 4	0,80	663	3.710	934
Límite superior (LS) de la serie	1,00	829	5.019	1309

**Interpretación:**

En la muestra  $n = 829$  hogares, los quintiles son grupos ordenados de 166 hogares. El segmento de familias con ingresos anuales de \$ 265 a \$1364 corresponden al Quintil 1.

**Decisión:**

Los ingresos familiares más bajos corresponden al segmento LI  $\rightarrow$  Quintil 1 que equivale a ingresos diarios entre \$0,73 y \$3,74. Urge intervención en este segmento social.

4.4.1.3 DECILES

Un decil es cualquiera de las diez partes proporcionales ordenadas de una serie de datos. En el análisis de deciles se comparan las diferencias entre segmentos, valorando la proporcionalidad. En esta lógica, dos deciles consecutivos equivalen a un quintil.

**Ejercicio 64:**

Realizar el análisis de deciles para valorar los ingresos anuales de 829 hogares de agricultores.

**Datos:**

En el Cuadro 34, se expone el análisis de deciles de la variable ingresos anuales por familia en una muestra de 829 hogares de agricultores.

*Cuadro 34. Análisis de deciles de los ingresos anuales en 829 hogares de agricultores*

Deciles	Proporción	Frecuencia acumulada	Ingresos anuales (\$)	Incremento interdecil (\$)
Límite inferior (LI) de la serie	0	1	265	--
Decil 1 (D1)	0,10	83	1030	765
Decil 2 (D2)	0,20	166	1364	334
Decil 3 (D3)	0,30	249	1700	336
Decil 4 (D4)	0,40	332	1994	294
Decil 5 (D5)	0,50	415	2410	416
Decil 6 (D6)	0,60	497	2776	366
Decil 7 (D7)	0,70	580	3265	489
Decil 8 (D8)	0,80	663	3715	450
Decil 9 (D9)	0,90	746	4385	670
Límite superior (LS) de la serie	1,00	829	5019	634

**Resultados:**

Un decil está conformado por 83 familias que equivale al 10 % de la serie de datos. Los ingresos familiares del decil más bajo varían de \$265 a \$1030.año<sup>-1</sup>.

**Decisión:**

Los ingresos familiares, en promedio, varían de \$0,73 a \$2,82 diarios.

4.4.1.4 PERCENTILES

El percentil es una medida de posición para analizar una variable cuantitativa en una serie ordenada de datos, en muestras grandes. La serie de datos se distribuye proporcionalmente en cien partes o percentiles donde se comparan las diferencias entre las proporciones.

El límite inferior (LI) corresponde al percentil 0. El percentil 25 equivale al cuartil 1, el percentil 50 al cuartil 2, el percentil 75 al cuartil 3 y el percentil 100 al límite superior de



la serie de datos. Análogamente, el percentil 20 corresponde al quintil 1, el percentil 40 al quintil 2, el percentil 60 al quintil 3, el percentil 80 al quintil 4 y el percentil 100 es el límite superior de la serie.

En Excel, los percentiles se obtienen con la función:

=percentil(valor1:valor2;p). Dónde la p es el percentil buscado en la serie de datos.

Por ejemplo: el percentil 50, que tendrá el mismo valor de la mediana (Q2). En el caso del ejercicio de los ingresos de 829 familias, el percentil 50 = \$2410.

#### 4.4.2 DIAGRAMA DE “CAJA”

El diagrama de “caja y bigote” o simplemente “diagrama de caja” (box-plot) es la representación gráfica del análisis de cuartiles, que permite visualizar la segmentación de la serie de datos original en proporciones del 25 %, así como, la valoración de los potenciales sesgos o inclusive la detección de datos “fuera de tipo”.

#### Ejercicio 65:

Elaborar el diagrama de caja e interpretar los resultados con los datos de producción (gramos.planta<sup>-1</sup>) en una muestra aleatoria de 30 plantas.

La serie de datos de producción (g.planta<sup>-1</sup>) en 30 plantas, se analizó en cuartiles, siendo el límite inferior 131 y el límite superior 199, la mediana (Q2) es 161,5 El valor Q1 es 151,75 y el Q3 de 173,5 (Gráfico 16).

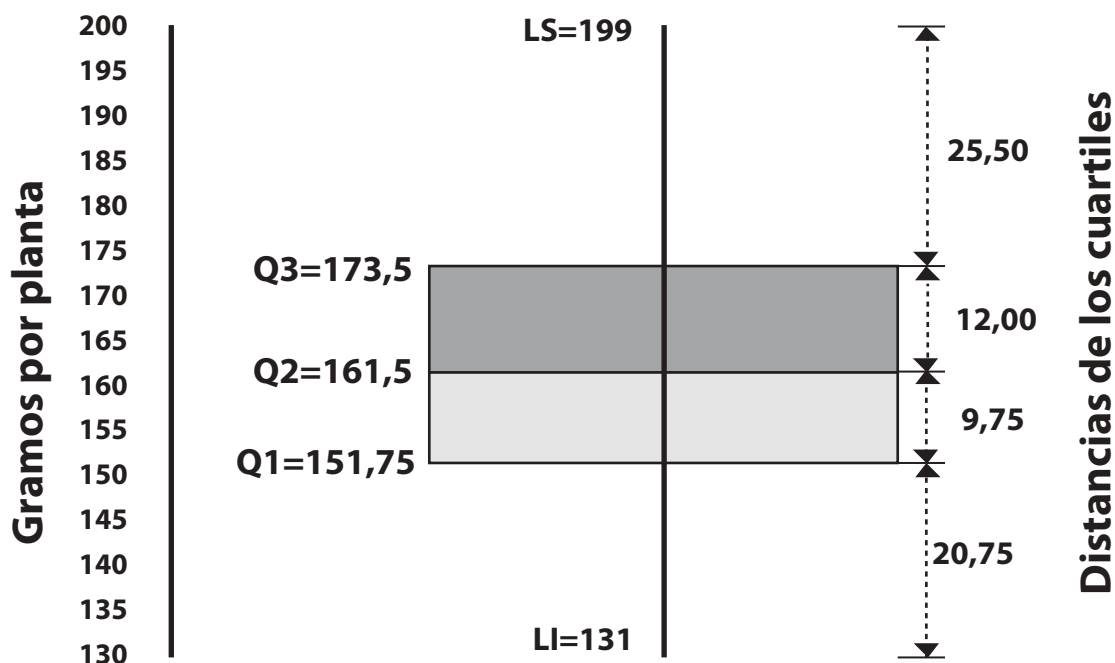


Gráfico 16. Diagrama de caja para la serie de datos de producción g.planta<sup>-1</sup>

#### Interpretación:

Las distancias entre cuartiles son:

$$Q1 - LI = 151,75 - 13,0 = 20,75 \text{ g.planta}^{-1}$$

$$Q2 - Q1 = 161,50 - 151,75 = 9,75 \text{ g.planta}^{-1}$$

$$Q3 - Q2 = 173,50 - 161,5 = 12,00 \text{ g.planta}^{-1}$$

$$LS - Q3 = 199,00 - 173,5 = 25,50 \text{ g.planta}^{-1}$$

$$RI = Q3 - Q1 = 173,5 - 151,75 = 21,75 \text{ g.planta}^{-1}$$

**Decisión:**

Se observa un ligero sesgo hacia la derecha. El cuartil más alto tiene una distancia de  $25,50 \text{ g.planta}^{-1}$ , mayor que el rango intercuartil  $= 21,75 \text{ g.planta}^{-1}$ .

4.4.3 RANGO

El rango ( $R_n$ ), también conocido como amplitud, es la diferencia entre los valores máximo y mínimo, de una serie ordenada de datos. Es una medida efectiva para determinar la dispersión de los datos.

$$R_n = \text{Valor máximo} - \text{Valor mínimo} \quad [56]$$

Según las propiedades de la distribución normal, el 99,74 % de los datos, en series cuantitativas, se localiza bajo la curva normal. Este hecho significa que la desviación estándar ( $S$ ) se aproxima a la relación siguiente:

$$\frac{R_n}{6} \sim S \quad [57]$$

Por lo tanto:  $S \sim \frac{1}{6} R_n$ .

El rango es un estadígrafo de amplio uso en los estudios biológicos y no biológicos, que facilita interpretar la dispersión de los datos y aproximar a una estimación de la desviación estándar ( $S$ ).

**Ejercicio 66:**

La estatura de los estudiantes de la carrera de ingeniería agrícola es 170 cm, con un rango de 60 cm. Cuáles son las estatura más alta y más baja en la población estudiantil?

**Datos y análisis:**

La estatura más baja se calcula como:  $\bar{X} - \frac{1}{2} R_n = 170 - .30 = 140 \text{ cm}$

La estatura más alta se calcula como:  $\bar{X} + \frac{1}{2} R_n = 170 + .30 = 200 \text{ cm}$

Si  $\bar{X} = 170 \text{ cm}$  y se acepta que  $R_n = 60 \text{ cm}$ , hay que verificar los puntos extremos.

**Decisión:**

Verificar los datos extremos de estatura.

**Ejercicio 67:**

Calcular el rango y la desviación típica en una serie de datos distribuidos normalmente donde el valor mínimo es 131 y el máximo  $199 \text{ g.planta}^{-1}$ ,

**Datos y análisis:**

El valor mínimo = 131 y el valor máximo =  $199 \text{ g.planta}^{-1}$

$$R_n = 199 - 131 = 68 \text{ g. planta}^{-1}$$

**Estimación de la desviación típica:**

Los valores extremos abarcan una amplitud de:  $\bar{X} \pm 3S$

Por lo tanto:  $R_n \sim \bar{X} \pm 3S \sim 6S \rightarrow S \sim \frac{1}{6}R_n = \frac{68}{6} = 17 \text{ cm}$

El valor calculado de la desviación estándar, en la muestra fue:  $S = 17,26 \text{ cm}$ .

**Decisión:**

La desviación estándar estimada a partir de la relación:  $S \sim \frac{1}{6}R_n = 17 \text{ cm}$

**4.4.4 VARIANZAS DE MUESTRAS Y DE LAS POBLACIONES**

La varianza es una medida de dispersión calculada como la esperanza del cuadrado de la desviación de dicha variable respecto de su media, en muestras (n) o en poblaciones (N). Se expresa en unidades cuadráticas. Las varianzas para las muestras (n) se calcula para datos no agrupados y agrupados.

El cálculo de la varianza ( $S^2$ ) a partir de datos no agrupados, se realiza con las siguientes fórmulas:

$$S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n - 1} \quad [58]$$

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1} \quad [59]$$

En el cálculo de la varianza poblacional ( $\sigma^2$ ) o parámetros se realiza mediante la relación entre la suma de cuadrados y el tamaño de población (N).

Las fórmulas usadas son las siguientes:

$$\sigma^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N} \quad [60]$$

$$\sigma^2 = \frac{\sum (X_i - \bar{X})^2}{N} \quad [61]$$

Dónde:

$S^2$  = Varianza de la muestra (estadístico)

$\sigma^2$  = Varianza de la población (parámetro)

x = Observaciones individuales

$\bar{X}$  = Promedio de la serie de datos de la variable X

n = Número de observaciones total (muestra)

N = Tamaño de la población

**Ejercicio 68:**

Calcular la varianza del peso de nueve aguacates, en decagramos, usando la fórmula 58

**Datos y análisis:**

Los datos del ensayo se exponen en el Cuadro 35.

Cuadro 35. Datos unitarios de peso de aguacates y cálculo de los cuadrados

n	Peso Dg (X)	X <sup>2</sup>
1	20	400
2	25	625
3	26	676
4	28	784
5	30	900
6	30	900
7	31	961
8	33	1089
9	34	1156
Σ	257	7491
$\bar{X}$	28,6	

$$n = 9 \quad \sum X = 257 \quad (\sum X)^2 = (257)^2 = 66049 \quad \sum X^2 = 7491$$

$$S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n - 1} \quad \rightarrow S^2 = \frac{7491 - \frac{(257)^2}{9}}{8} = 19,03$$

**Decisión:**

$$S^2 = 19,03 \text{ Dg}^2$$

**Ejercicio 69:**

Calcular la varianza de los datos de peso en decagramos, en una muestra de 9 aguacates, usando la fórmula 59.

**Datos y análisis:**

Los datos de peso (Dg) y el cálculo de la varianza, se indica en el Cuadro 36.

Cuadro 36. Cálculo de la varianza del peso de los aguacates

n	Peso en Dg (X)	X <sup>2</sup>	$\bar{X}$	$D = (X_i - \bar{X})$	D <sup>2</sup>
1	20	400	28,6	-8,60	73,96
2	25	625	28,6	-3,60	12,96
3	26	676	28,6	-2,60	6,76
4	28	784	28,6	-0,60	0,36
5	30	900	28,6	1,40	1,96
6	30	900	28,6	1,40	1,96
7	31	961	28,6	2,40	5,76
8	33	1.089	28,6	4,40	19,36
9	34	1.156	28,6	5,40	29,16
Σ	257	7.491		0	152,24
$\bar{X}$	28,6				

**Resultado:**

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1} \quad \rightarrow S^2 = \frac{152,24}{8} = 19,03 \text{ Dg}^2$$

**Decisión:**

$$S^2 = 19,03 \text{ Dg}^2$$

4.4.5 VARIANZA EN DATOS AGRUPADOS

La varianza de datos agrupados es una estimación menos precisa que se basa en datos organizados en clases, en las distribuciones de frecuencias.

La fórmula más usada para calcular la varianza en datos agrupados es la siguiente:

$$S^2 = \frac{\sum(M - \bar{X})^2(FA)}{\sum FA} \quad [62]$$

Dónde:

$S^2$  = Varianza de los datos agrupados en clases

FA = Frecuencias absolutas

$\sum FA$  = Sumatoria de las frecuencias absolutas

M = Marca de clase o punto medio del intervalo de clase

X = Promedio

### Ejercicios 70:

Calcular la media ponderada y la varianza en base a la tabla de frecuencias de la variable producción (gramos.planta<sup>-1</sup>) en una muestra de 30 plantas.

#### Datos:

Los datos agrupados en clases, marcas de clase, productos M.(FA) y  $(M - \bar{X})^2(FA)$ , usados para calcular la media ponderada y la varianza, se exponen en el Cuadro 37.

Cuadro 37. Clases, marca de clase (M), frecuencia absoluta (FA) y cálculos de la media ponderada y varianza

Clases (g.planta <sup>-1</sup> )	M	FA	$\bar{X}$	$(M - \bar{X})^2$	$(M - \bar{X})^2 \cdot (FA)$
[130-140]	135	3	162,7	765	2296
[140-150]	145	3	162,7	312	936
[150-160]	155	8	162,7	59	470
[160-170]	165	7	162,7	5	38
[170-180]	175	4	162,7	152	608
[180-190]	185	3	162,7	499	1496
[190-200]	195	2	162,7	1045	2091
$\sum$		30			7937

#### Cálculo de la media ponderada:

$$\bar{X} = \frac{\sum M(FA)}{\sum FA} \quad \rightarrow \quad \bar{X} = \frac{4880}{30} = 162,67 \text{ g.planta}^{-1}$$

#### Cálculo de la varianza de la muestra:

$$S^2 = \frac{\sum (M - \bar{X})^2(FA)}{\sum FA} \quad \rightarrow \quad S^2 = \frac{7937}{30} = 264,5 \text{ (g. planta}^{-1}\text{)}^2$$

#### Decisión:

$$\bar{X} = 162,7 \text{ g.planta}^{-1} \text{ y } S^2 = 264,5 \text{ (g. planta}^{-1}\text{)}^2$$

#### 4.4.6 DESVIACIÓN ESTÁNDAR

La desviación estándar (S) es la raíz cuadrada de la varianza y se expresa en las mismas unidades de la media. La S siempre tendrá valores positivos. La S igual a cero indica que los datos son iguales. Si a la variable se suma o se resta un número, la S no varía. Si se

multiplica por un número, la S queda multiplicada por dicho número. Si se divide por un número, la S queda dividida por dicho número.

Las fórmulas para calcular la desviación estándar son:

$$S = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}} \quad [63]$$

$$S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}} \quad [64]$$

$$S = \sqrt{\frac{\sum (M - \bar{X})^2 f}{\sum f}} \quad [65]$$

$$S = \sqrt{S^2} \quad [66]$$

$$S = (S^2)^{0,5} \quad [67]$$

Las propiedades de la desviación estándar lo convierten en una valiosa herramienta de la estadística descriptiva e inferencial (Gráfico 17).

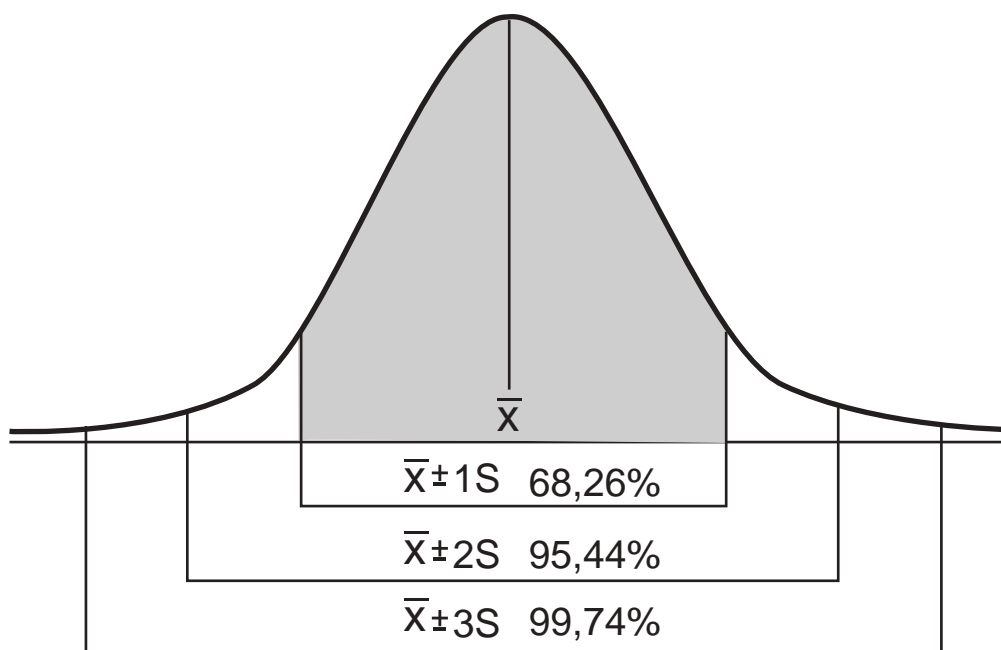


Gráfico 17. Curva normal o campana gaussiana

### Ejercicio 71:

Dado el promedio de peso,  $\bar{X} = 28,6 \text{ g por fruto}$  y  $S^2 = 19,03 \text{ g}^2$  ¿Cuál es la desviación estándar?

#### Datos y análisis:

$$S = \sqrt{S^2} = \sqrt{19,03} = 4,4 \text{ gramos.fruto}^{-1}$$

#### Decisión:

La desviación estándar es  $4,4 \text{ g.fruto}^{-1}$ . El peso varía de  $15,4$  a  $41,8 \text{ g.fruto}^{-1}$ .

**Ejercicio 72:**

Calcular la probabilidad de tener frutos con promedios de peso  $> 28,6 \text{ g.fruto}^{-1}$ , cuando la media  $\bar{X} = 28,6 \text{ g}$  y la desviación estándar  $S = 4,4 \text{ g}$ .

**Análisis:**

Dada la media de  $28,6 \text{ g.fruto}^{-1}$ , se asume que el 50 % de datos están arriba de ese valor. La probabilidad de tener valores mayores que  $28,6 \text{ g.fruto}^{-1}$  es  $p = 0,50 \sim 50 \%$ .

**Decisión:**

El 50 % de frutos tendrían pesos mayores que la media.

**Ejercicio 73:**

Dados los valores:  $S^2 = 264,6$  y  $\bar{X} = 162,7 \text{ cm}$ . ¿Cuáles son los valores:  $\bar{X} \pm 2S$  ?

**Datos y análisis:**

$$S = \sqrt{S^2} = \sqrt{264,6} = 16,3 \text{ cm}$$

$$\bar{X} + 2S = 162,7 + 2(16,3) = 195,3 \text{ cm}$$

$$\bar{X} - 2S = 162,7 - 2(16,3) = 130,1 \text{ cm}$$

$$\bar{X} \pm 2S = 0,9544 \sim 95,44 \% \text{ del área bajo la curva}$$

**Decisión:**

El área bajo la curva del segmento  $\bar{X} \pm 2S$  equivale al 95.44 %

4.4.7 ERROR ESTÁNDAR

El error estándar ( $S_{\bar{x}} = EE$ ) es una estimación de los cambios de las medidas, de muestra a muestra, en los estadísticos como la media, la asimetría o curtosis. El error estándar depende del tamaño de muestra y de la dispersión. El EE se calcula, no es asignado como en el caso del tamaño óptimo de muestra “d” que es el error máximo aceptado en el muestreo.

El error experimental se integra de tres errores parciales: de medición, sistemático y estadístico (Gráfico 18). El error de medición es la diferencia entre valor real y valor medido, el error sistemático es inherente a las especificaciones técnicas de equipos y el error estadístico corresponde a los desvíos en la toma de datos por causas inherentes al ambiente, redondeo de cifras y falta de experiencia de los operadores (Gil y Rodríguez, 2000, p. 3).

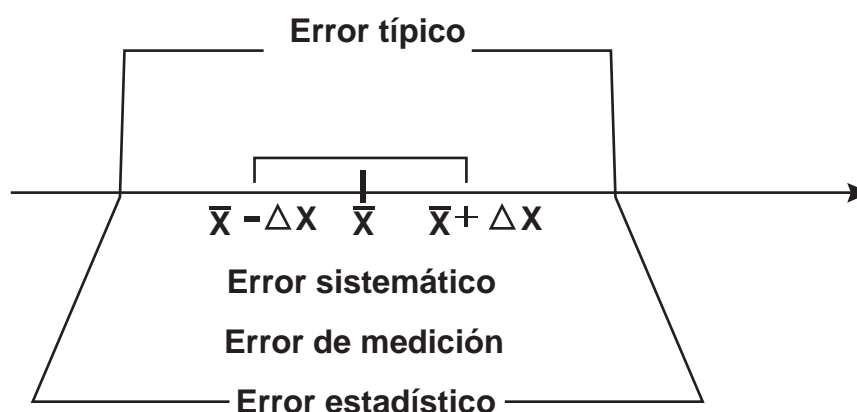


Gráfico 18. Relación entre la media muestral y su error típico

**Ejercicio 74:**

Calcular el error estándar  $S = 4,4 \text{ g.fruto}^{-1}$ , en una muestra  $n=9$ , con peso promedio  $\bar{X} = 28,6$ .

**Datos:**

$$S = 4,4 \text{ g.fruto}^{-1}, n = 9 \text{ y } \bar{X} = 28,6$$

**Análisis:**

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{4,4}{\sqrt{9}} = \frac{4,4}{3} = 1,47 \text{ g.fruto}^{-1}$$

$$\bar{X} \pm S_{\bar{x}} = 28,6 \pm 1,47 \text{ g.fruto}^{-1}$$

**Decisión:**

El error típico EE es 1,47 g. La media  $\bar{X}$  varía de 27,13 a 30,07 g.fruto<sup>-1</sup>

**Ejercicio 75:**

Interpretar el gráfico indicado:  $\bar{X} \pm S_{\bar{x}}$

**Datos y análisis:**

Las medias y errores típicos de tres clones de café se indican en el Gráfico 19. El clon COF 06:  $277 \pm 3 \text{ cm}$ , el CON-ERB-0:  $298 \pm 5 \text{ cm}$  y el NP-2024:  $333 \pm 6 \text{ cm}$ .

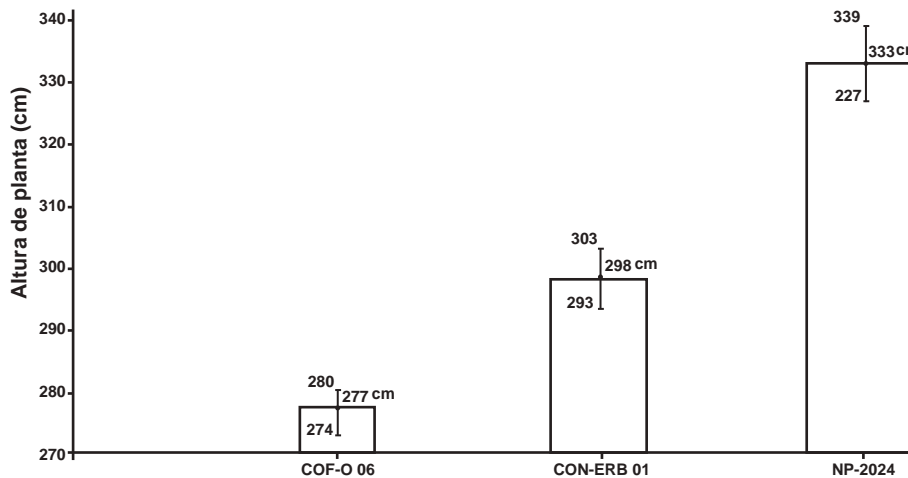


Gráfico 19. Medias ± Errores típicos de tres clones de café robusta

**Decisión:**

El clon COF-06 tiene el porte más bajo y menor error típico.

4.4.8 INTERVALO DE CONFIANZA DEL PARÁMETRO  $\mu$

En las poblaciones reales se analizan las colecciones de datos censales para calcular los parámetros. En el caso de un censo no hay error de estimación, se calculan directamente los parámetros. En las muestras, además de calcular los estadígrafos, se proyecta una aproximación a los parámetros  $\mu$  y  $\sigma$ , que se ubicarán dentro de un intervalo de confianza con una probabilidad específica (Badii et al., 2011, p. 90).

La media de la población  $\mu$ , se ubica siempre entre un límite inferior (LI) y un límite superior (LS):  $\mu = \bar{X} \pm \text{IC}$  con un nivel de confianza  $(1 - \alpha)$  concreto. Si la probabilidad de



equivocarse se ha fijado en 5 % ( $\alpha = 0,05$ ), la confianza es 95 % en la prueba de una cola y para las pruebas de dos colas, el nivel de confianza se representa como:  $1 - \frac{\alpha}{2}$

En el Cuadro 38, se exponen las fórmulas para calcular el intervalo de confianza de  $\mu$ , en series de datos cuantitativos y categóricos.

$$IC_{(1-\frac{\alpha}{2})} = Z_{\frac{\alpha}{2}} \left( \frac{S}{\sqrt{n}} \right) \quad [68]$$

$$IC_{(1-\frac{\alpha}{2})} = \pm Z_{\alpha} \left[ \sqrt{\frac{p \cdot q}{n}} \right] \quad [69]$$

Dónde:

IC = Intervalo de confianza para un nivel de significación dado

n = Número de observaciones o tamaño de la muestra

$Z_{\alpha}$  = Estadístico que se asocia a la probabilidad de equivocarse en la decisión

Si  $\alpha = 0,05 \rightarrow Z = 1,96$

Si  $\alpha = 0,01 \rightarrow Z = 2,58$

p = Proporción observada de *ocurrencia* de un evento

q = Proporción complementaria de *no ocurrencia* de un evento:  $q = 1 - p$

Cuadro 38. Fórmulas para calcular los límites de confianza de  $\mu$  en series de datos numéricos y categóricos

	Varianza	Desviación estándar	Error estándar	Intervalo de confianza	Límites de confianza de $\mu$
Datos numéricos	$S^2$	$S$	$EE = \frac{S}{\sqrt{n}}$	$Z_{\alpha}(EE)$	$\mu = \bar{x} \pm Z_{\alpha}(EE)$
Datos categóricos	$p \cdot q$	$\sqrt{p \cdot q}$	$EE = \frac{\sqrt{p \cdot q}}{\sqrt{n}}$	$z_{\alpha} \sqrt{\frac{p \cdot q}{n}}$	$\mu = p \pm z_{\alpha} \sqrt{\frac{p \cdot q}{n}}$

### Ejercicio 76:

Estimar  $\mu$  con sus límites inferior (LI) y superior (LS) de confianza (95 %) siendo la media  $\bar{X} = 162,7$  cm y el error típico  $EE = 2,96$  cm, en una prueba de dos colas.

#### Datos y análisis:

$$\bar{X} = 162,7 \text{ cm}, EE = 2,96 \text{ cm y } \alpha = 0,05$$

$$IC_{(1-\frac{\alpha}{2})} = Z_{\frac{\alpha}{2}} \left( \frac{S}{\sqrt{n}} \right) \quad \rightarrow \quad IC_{95\%} = 1,96 (2,96) = 5,8 \text{ cm}$$

$$LI \text{ de } \mu = \bar{X} - IC = 162,7 - 5,8 = 156,9 \text{ cm}$$

$$\bar{X} = 162,7 \text{ cm}$$

$$LS \text{ de } \mu = \bar{X} + IC = 162,7 + 5,8 = 168,5 \text{ cm}$$

#### Decisión:

La media poblacional  $\mu$  se ubica entre 156,9 y 168,5 cm, con el 95 % de confianza.

**Ejercicio 77:**

Determinar el efecto de las dietas sobre la ganancia en cerdos con el 95 % de confianza. La ganancia de peso promedio (GP) en un grupo de 25 cerdos, en un determinado período, fue 15 kg con una desviación típica de 3 kg. En otro grupo de 25 cerdos, en un período similar, se obtuvo una ganancia promedio de 17 kg con una desviación típica de 3 kg.

**Datos y análisis:**

Grupo I:

$n_I = 25$  cerdos

GP (I) = 15 kg y  $S = 3$  kg

$$\text{Error típico} = \frac{S}{\sqrt{n}} \rightarrow EE = \frac{3}{\sqrt{25}} = 0,60 \text{ kg}$$

$$\text{Intervalo de confianza (95 \%)} = Z_{\alpha} \left[ \frac{S}{\sqrt{n}} \right] = 1,96 (0,60) \rightarrow IC_{95 \%} = 1,2 \text{ kg}$$

Grupo II:

$n_{II} = 25$  cerdos

GP (II) = 17 kg y  $S = 3$  kg

$$\text{Error típico} = \frac{S}{\sqrt{n}} \rightarrow EE = \frac{3}{\sqrt{25}} = 0,60 \text{ kg}$$

$$\text{Intervalo de confianza (95 \%)} = Z_{\alpha} \left[ \frac{S}{\sqrt{n}} \right] = 1,96 (0,60) \rightarrow IC(95 \%) = 1,2 \text{ kg}$$

**Análisis gráfico:**

La diferencia entre los promedios de la ganancia de peso de los dos grupos alcanzó los 2 kg, sin embargo, el traslape indica que esa diferencia como efecto de las dietas aplicadas a los dos grupos de cerdos, no es estadísticamente diferente (Gráfico 20).

Grupo I:  $\bar{Y} = 15$  kg con  $IC_{95 \%} = 1,2$  kg, indica que  $\mu$  varía de 13,8 a 16,2 kg

Grupo II:  $\bar{Y} = 17$  kg con  $IC_{95 \%} = 1,2$  kg, indica que  $\mu$  varía de 15,8 a 18,2 kg

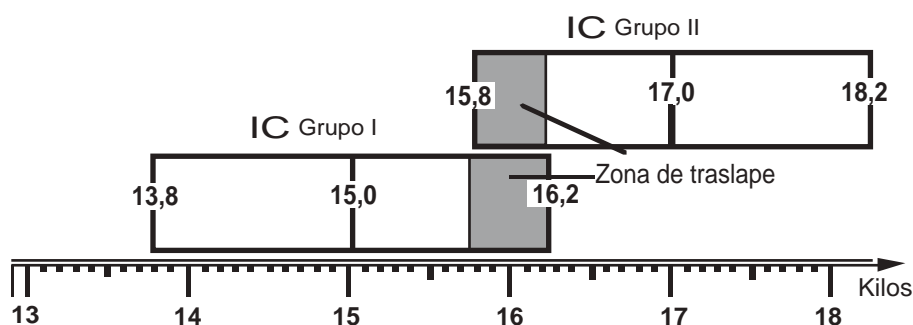


Gráfico 20. Intervalos de confianza de  $\mu$  (95 %) para ganancia de peso en dos grupos cerdos

**Decisión:**

No hubo efecto significativo de las dietas sobre la ganancia de peso en los cerdos.

**Ejercicio 78:**

Estimar el intervalo de confianza (95 %) para el porcentaje de plantas enfermas en un cultivo de frutales, donde la muestra  $n = 100$  plantas con una media de plantas enfermas es 7 % que representa la media histórica del estado sanitario.

**Datos y análisis:**

$$\alpha = 0,05; n = 100; \bar{Y} = 7 \% \sim p = 0,07 \rightarrow q = 1 - p = 0,93$$

$p = 0,07$  plantas enfermas

$q = 1 - p \rightarrow q = 1 - 0,07 = 0,93$  plantas sanas

$$IC_{\alpha} = z_{\alpha} \sqrt{\frac{p \cdot q}{n}} = 1,96 \sqrt{\frac{0,07(0,93)}{100}} = 1,96(0,0255) = 0,05$$

$$\mu = p \pm z_{\alpha} \sqrt{\frac{p \cdot q}{n}} = 0,07 \pm 1,96 \sqrt{\frac{0,07(0,93)}{100}} = 0,07 \pm 0,05$$

LI de  $p = 0,07 - 0,05 = 0,02 \sim 2 \%$

$\bar{Y} = 7 \%$

LS de  $p = 0,07 + 0,05 = 0,12 \sim 12 \%$

**Decisión:**

La proporción de plantas enfermas en la población varía desde 2 hasta el 12 % con una confianza del 95 %.

4.4.9 COEFICIENTE DE VARIACIÓN DE LA MUESTRA

El Coeficiente de variación (CV) es una medida estandarizada de la dispersión. Es la relación entre la desviación estándar y la media de la serie de datos, expresada en por ciento (Badii et al., 2004, p.43). En las curvas normales, las desviaciones estándar se relacionan con la media, alejándose o acercándose a ella en función de la dispersión de los datos. El CV sirve para comparar la dispersión de dos o más muestras, independientemente de las magnitudes de las series de datos (Salazar, 2010, p. 105).

La fórmula para calcular el Coeficiente de variación (CV) es la siguiente:

$$CV(\%) = \frac{s}{\bar{x}} (100) \quad [70]$$

Cuando el CV (%) es elevado significa que existe alta heterogeneidad en la serie de datos y si fuese reducido habría poca variabilidad. Si en una población se constata que hay subpoblaciones, es necesario estratificar y analizar por partes. Un CV (%) > 50 %, indica que la dispersión de los datos es alta y si CV > 100 % es probable que haya sesgos.

**Ejercicio 79:**

Calcular el Coeficiente de variación de una muestra para la variable peso de frutos, donde:  $\bar{X} = 28,6$  g.fruto<sup>-1</sup> y  $S = 4,4$  g.fruto<sup>-1</sup>.

**Datos y análisis:**

$$\bar{X} = 28,6 \text{ g.fruto}^{-1} \text{ y } S = 4,4 \text{ g.fruto}^{-1}$$

$$CV(\%) = \frac{s}{\bar{x}} (100) \rightarrow CV(\%) = \frac{4,4}{28,6} (100) = 15,4 \%$$

**Decisión:**

El Coeficiente de variación es 15,4 %, lo cual indica que la desviación estándar y la media se relacionan con una baja intensidad.

4.4.10 VARIACIÓN RELATIVA

La variación relativa (VR) en una muestra, es la proporción del error típico de la media con respecto de su media, que puede expresar en porcentaje. Se calcula en base a la relación entre el error estándar y el promedio de la muestra (Badii et al., 2004, p. 43).

La fórmula para calcular VR es la siguiente:

$$VR(\%) = \frac{S}{\bar{X}} (100) \quad [71]$$

**Ejercicio 80:**

Calcular la variación relativa en una muestra  $n = 9$ , siendo los estadísticos desviación estándar  $S = 4,4 \text{ g.fruto}^{-1}$  y el peso promedio del frutos  $\bar{X} = 28,6 \text{ g.fruto}^{-1}$ .

**Datos:**

$$n = 9, \bar{X} = 28,6 \text{ g.fruto}^{-1} \text{ y } S = 4,4 \text{ g.fruto}^{-1}$$

**Cálculo del error estándar:**

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{4,4}{\sqrt{9}} = \frac{4,4}{3} = 1,47 \text{ g.fruto}^{-1}$$

**Cálculo de la variación relativa:**

$$VR(\%) = \frac{EE}{\bar{X}} (100) = \frac{1,47}{28,6} (100) = 5,14 \%$$

**Decisión:**

La VR = 5,14 % indica que la proporción del error típico de la media es reducida.

4.4.11 PRECISIÓN

En la estadística, se denomina precisión a la capacidad de obtener el mismo resultado en distintas mediciones, en las muestras tomadas de una población. La variación relativa (VR) es la proporción de error típico con relación al promedio y la diferencia de VR respecto del total estima a la precisión (Badii et al., 2011, p. 90).

La precisión (P) se estima usando la fórmula:

$$P (\%) = 100 - VR \quad [72]$$

**Ejercicio 81:**

En un estudio de  $n = 9$  frutos,  $\bar{X} = 28,6 \text{ g.fruto}^{-1}$  y  $S = 4,4 \text{ g.fruto}^{-1}$  ¿Cuál es la precisión?

**Datos:**

$$n = 9, \bar{X} = 28,6 \text{ g.fruto}^{-1} \text{ y } S = 4,4 \text{ g.fruto}^{-1}$$

**Cálculo del error estándar:**

$$EE = \frac{S}{\sqrt{n}} = \frac{4,4}{\sqrt{9}} = \frac{4,4}{3} = 1,47 \text{ g.fruto}^{-1}$$

**Cálculo de la variación relativa:**

$$VR(\%) = \frac{EE}{\bar{X}} (100) = \frac{1,47}{28,6} (100) = 5,14 \%$$

**Cálculo de la precisión estadística:**

$$P(\%) = 100 - VR = 100 - 5,14 = 94,86 \%$$

**Decisión:**

La precisión es 94,86 %.

## 4.5 MEDIDAS DE CONCENTRACIÓN

Las medidas de concentración se refieren al "grado de igualdad en el reparto de la totalidad de los valores de una variable". Si pocos individuos acaparan una gran parte de los salarios, de la tierra agrícola, del acceso al agua o del capital, ¿se podría medir ese grado de concentración?. Si, se mide usando el índice de Gini (IG) y representando gráficamente en la curva de Lorenz.

### 4.5.1 ÍNDICE DE GINI

El índice de Gini (IG) es un estadístico que mide el grado de desigualdad de los datos, organizados en frecuencias absolutas, hacia un lado u otro de la curva gaussiana. El índice de Gini tiene algunas limitaciones, ya que no proporciona información sobre la naturaleza específica de la desigualdad ni sobre las causas subyacentes.

El IG toma valores entre 0 y 1. Cuando el IG se aproxima a cero, los datos están uniformemente repartidos y cuando el IG se aproxima a 1 significa que la distribución está concentrada hacia una de las colas, por tanto, habría mucha desigualdad.

El IG se usa para analizar la desigualdad de los salarios, la inequitativa distribución de la tierra o de las propiedades rurales o urbanas.

La fórmula para calcular el IG es la siguiente:

$$IG = \frac{\sum(p_i - q_j)}{\sum p_i} \quad [73]$$

Dónde:

IG = Índice de Gini

X = Variable de análisis (por ejemplo: salarios mensuales)

FA = Frecuencia absoluta por clase

FAA = Frecuencia absoluta acumulada

$p_i$  y  $q_j$  son variables relacionadas, no probabilidades.

$$p_i = \frac{FAA}{\sum FA} (100)$$

$$q_j = \frac{[FA(X)]Acum.}{\sum(FA.X)} (100)$$

Nota: Al calcular IG no se deben incluir los valores:  $p_i = 100$ ,  $q_j = 100$  y  $p_i - q_j = 0$

### Ejercicio 82

Calcular el Índice de Gini con los datos de remuneraciones mensuales que paga una empresa agrícola a sus 10 trabajadores e interpretar los resultados.

#### Datos y análisis:

Las remuneraciones mensuales que paga una empresa agrícola a sus 10 trabajadores, se exponen en el Cuadro 39. En el Cuadro 40, se detalla el cálculo del IG.

Cuadro 39. Salarios mensuales de los trabajadores de una empresa agrícola

Salarios: X	Salarios mensuales (\$)			
		600	900	1200
Trabajadores	6	2	1	1

Cuadro 40. Frecuencias de las remuneraciones de 10 trabajadores en una empresa agrícola y procedimiento de cálculo del índice de Gini

X (\$)	FA	(FA).X	FAA	$p_i = \frac{FAA}{\sum FA} (100)$	[(FA).X]Acumulada	$q_j = \frac{[(FA).X]Acum.}{\sum (FA.X)} (100)$	$p_i - q_j$
600	6	3600	6	60	3600	43	17
900	2	1800	8	80	5400	64	16
1200	1	1200	9	90	6600	79	11
1800	1	1800	10	-	8400	-	-
$\Sigma$	10	8400		230	24000		44

#### Interpretación:

$$IG = \frac{\sum(p_i - q_j)}{\sum p_i} \rightarrow IG = \frac{44}{230} = 0,1913$$

#### Decisión:

El IG = 0,1913 está próximo a 0 y lejos de 1, por tanto, no hay evidencia de la desigualdad en las remuneraciones de los 10 trabajadores.

### 4.5.2 CURVA DE LORENZ

La “Curva de Lorenz” es la graficación de los valores del IG. Es una herramienta que ayuda a interpretar el grado de desigualdad de los datos organizados en frecuencias.

### Ejercicio 83:

Elaborar la curva de Lorenz con la información del IG referido a los salarios de 10 trabajadores, en una empresa agrícola.

#### Datos y análisis:

La curva de Lorenz, elaborada con los datos  $p_i$  y  $q_j$  de los resultados del índice de Gini (Ejercicio 82), IG = 0,196, se presenta en el Gráfico 21.

$p_j$	$q_j$
0	0
60	43
80	64
90	79
100	100

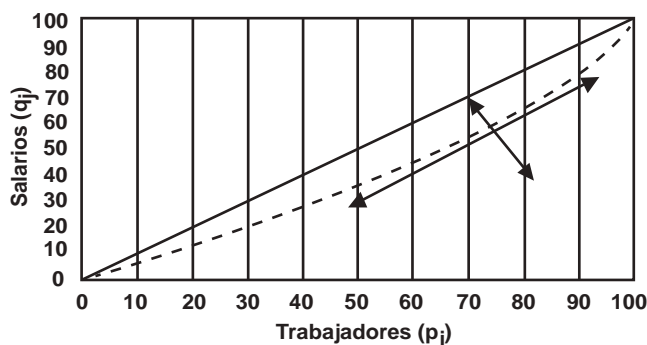


Gráfico 21. Curva de Lorenz de la distribución salarial en una empresa agrícola

**Decisión:**

El  $IG = 0,1913$ . indica que no es notoria la tendencia a la desigualdad en la distribución salarial, lo cual se confirma al interpretar la curva de Lorenz.

### 4.6 MEDIDAS DE FORMA DE LA DISTRIBUCIÓN

Las medidas de forma son aquellos estadísticos que indican si la distribución de frecuencias se aleja o se aproxima a distribución gaussiana.

#### 4.6.1 COEFICIENTE DE ASIMETRÍA

El Coeficiente de asimetría (A) es una medida que indica el alejamiento de la curva normal basado en las magnitudes y el número de elementos a cada lado del promedio (Gráfico 22). Al elaborar un histograma y el polígono de frecuencias, se proyectan las colas de la curva que pueden distorsionar la campana gaussiana. Si el 50 % de los datos se encuentran a cada lado de la media, la curva gaussiana es simétrica y si esto no ocurre es asimétrica (Molina y Rodrigo, 2010, p. 2).

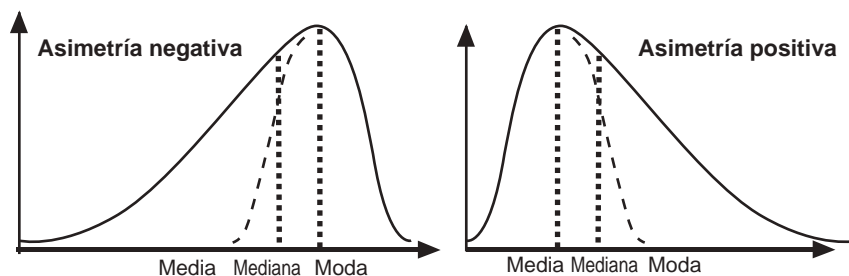


Gráfico 22. . Curvas asimétricas que distorsionan la campana gaussiana

Las fórmulas para calcular el Coeficiente de asimetría (A) son las siguientes:

$$A = \frac{n}{(n-1)(n-2)} \sum \left[ \frac{X_i - \bar{X}}{S} \right]^3 \quad [74]$$

$$A = \frac{\frac{1}{n} \sum (X_i - \bar{X})^3}{\left[ \frac{1}{n} \sum (X_i - \bar{X})^2 \right]^{3/2}} \quad [75]$$

Dónde:

A = Coeficiente de asimetría de Fisher (Algunos autores usan el símbolo g1).

n = Tamaño de la muestra

$X_i$  = Datos individuales de la variable aleatoria

$\bar{X}$  = Media

S = Desviación estándar de la muestra

Las hipótesis estadísticas se formulan como:

$H_0$ : Hay simetría en la distribución

$H_1$ : No hay simetría en la distribución, es asimétrica

La regla de decisión se plantea de la siguiente manera:

Si  $A \leq |0,50|$ : Hay simetría

Si  $A > |0,50|$ : Hay asimetría

En Excel, el coeficiente de asimetría se obtiene con la función: =COEFICIENTE.ASIMETRIA(valor inicial: valor final).

Si  $A > |0,50|$  se interpreta que la cola de la curva se prolonga hacia la derecha y se llama asimetría positiva; de otro modo, si  $A > |-0,50|$  significa que la cola de la curva se prolonga hacia la izquierda y tiende a mostrar asimetría negativa. En estos casos hay distorsión de la campana gaussiana que puede originar una curva “no normal”.

#### Ejercicio 84:

Calcular el Coeficiente de asimetría de Fisher para la variable peso del fruto, medida en gramos, en una muestra aleatoria  $n = 9$ , siendo los estadígrafos:  $\bar{X} = 28,6$  g y  $S = 4,4$  g.

#### Datos y análisis:

El procedimiento del cálculo de la asimetría, se detalla en el Cuadro 41.

Cuadro 41. Cálculo del Coeficiente de asimetría (A) de la variable peso del fruto

$n_i$	$X_i$	$\bar{X}$	S	$X_i - \bar{X}$	$\frac{X_i - \bar{X}}{S}$	$\left[\frac{X_i - \bar{X}}{S}\right]^3$
1	20	28,6	4,4	-8,6	-1,961	-7,55
2	25	28,6	4,4	-3,6	-0,815	-0,54
3	26	28,6	4,4	-2,6	-0,586	-0,20
4	28	28,6	4,4	-0,6	-0,127	0,00
5	30	28,6	4,4	1,4	0,331	0,04
6	30	28,6	4,4	1,4	0,331	0,04
7	31	28,6	4,4	2,4	0,560	0,18
8	33	28,6	4,4	4,4	1,019	1,06
9	34	28,6	4,4	5,4	1,248	1,94
$\Sigma$	257					-5,039

#### Cálculo de la asimetría:

$$A = \frac{n}{(n-1)(n-2)} \sum \left[ \frac{X_i - \bar{X}}{S} \right]^3 \quad \rightarrow \quad A = \frac{9}{(9-1)(9-2)} (-5,039)$$

$$A = (0,1607)(-5,039) = -0,81$$

#### Hipótesis estadísticas:

$H_0$ : Hay simetría. La distribución es normal

$H_1$ : No hay simetría. La distribución es asimétrica

#### Regla de decisión:

Si  $A \leq |0,50|$ : Hay simetría

Si  $A > |0,50|$ : Hay asimetría



**Interpretación:**

$A = -0,81 > |0,50|$ : Se evidencia un sesgo hacia la izquierda. Hay asimetría negativa.

**Decisión:**

La distribución tiende a mostrar asimetría negativa. Se aleja de la curva normal.

4.6.2 COEFICIENTE DE CURTOSIS

El Coeficiente de Curtosis (K) es una medida estadística del grado de concentración de los datos alrededor del promedio (Gráfico 23). Por la Curtosis se distinguen tres tipos de distribución: mesocúrtica, leptocúrtica y platicúrtica (Molina y Rodrigo, 2010, p. 7). La mesocúrtica es la típica curvatura alrededor del valor central (normal).

La distribución leptocúrtica es una curva puntiaguda porque hay una elevada concentración de datos alrededor de la media, alejándola de la normal. La distribución platicúrtica es la curva que tiende a aplanarse por la alta dispersión de los datos de extremo a extremo con un nítido alejamiento de la curva gaussiana.

Las fórmulas usadas para calcular la curtosis (K) son las siguientes:

$$K = \left[ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left( \frac{X_i - \bar{X}}{s} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)} \quad [76]$$

$$K = \frac{\frac{1}{n} \sum (X_i - \bar{X})^4}{\left[ \frac{1}{n} \sum (X_i - \bar{X})^2 \right]^2} \quad [77]$$

Las hipótesis estadísticas se formulan como:

$H_0$ : la distribución es mesocúrtica (curva normal)

$H_1$ : la distribución se aleja de la normalidad

La regla de decisión se plantea del modo siguiente:

Si  $K \leq |0,50|$ : La distribución es mesocúrtica. La curva es normal

Si  $K > |0,50|$ : La distribución se aleja de la mesocúrtica. Si tiene signo positivo es leptocúrtica y si tiene signo negativo es platicúrtica.

En Excel, la curtosis se obtiene con la función:

=curtosis(valor1: valor2).

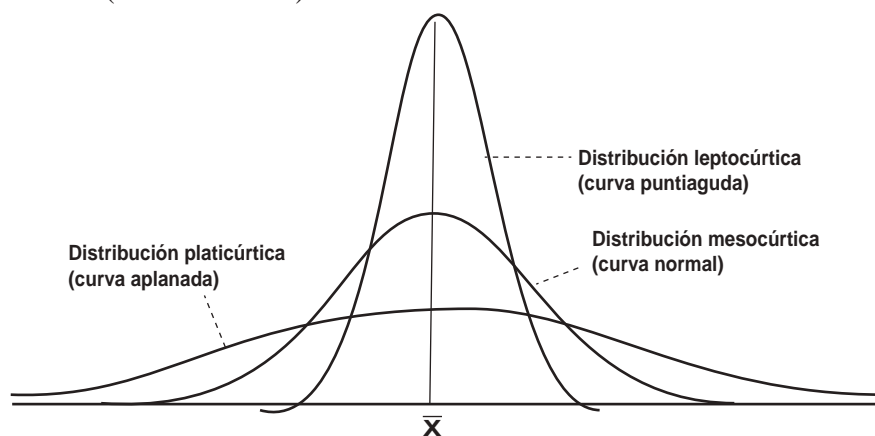


Gráfico 23. Curvas de las series de datos según los tipos de curtosis

**Ejercicio 85:**

Calcular el Coeficiente de Curtosis para la variable peso del fruto, medida en gramos, en una muestra aleatoria  $n = 9$  con estadígrafos:  $\bar{X} = 28,6$  g y  $S = 4,4$  g.

**Datos y análisis:**

El procedimiento de cálculo de la curtosis (K), se indica en el Cuadro 42.

Cuadro 42. Cálculo del Coeficiente de curtosis para peso en una muestra de 9 frutos

$n_i$	$X_i$	$\bar{X}$	S	$X_i - \bar{X}$	$\frac{X_i - \bar{X}}{S}$	$\left[\frac{X_i - \bar{X}}{S}\right]^4$
1	20	28,6	4,4	-8,6	-1,961	14,80
2	25	28,6	4,4	-3,6	-0,815	0,44
3	26	28,6	4,4	-2,6	-0,586	0,12
4	28	28,6	4,4	-0,6	-0,127	0,00
5	30	28,6	4,4	1,4	0,331	0,01
6	30	28,6	4,4	1,4	0,331	0,01
7	31	28,6	4,4	2,4	0,560	0,10
8	33	28,6	4,4	4,4	1,019	1,08
9	34	28,6	4,4	5,4	1,248	2,43
$\Sigma$	257					18,98

**Cálculo de la Curtosis:**

$$K = \left[ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left( \frac{X_i - \bar{X}}{S} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)}$$

$$K = \left[ \frac{9(10)}{(8)(7)(6)} (18,985) \right] - \frac{3(8)^2}{(7)(6)} = -0,514$$

**Hipótesis estadísticas:**

$H_0$ : la distribución es mesocúrtica

$H_1$ : la distribución se aleja de la normal

**Regla de decisión:**

Si  $K \leq |0,50|$ : La distribución es mesocúrtica

Si  $K > |0,50|$ : La distribución se aleja de la mesocúrtica

**Interpretación:**

$K = -0,514 > |0,50|$ : La distribución se aleja de la curva mesocúrtica.

**Decisión:**

$K = -0,514$ . La distribución muestral tiende a ser platicúrtica.

## 4.7 CONDICIONES PARA USAR LAS TÉCNICAS PARAMÉTRICAS

Para que los resultados tengan validez, en los análisis con la estadística paramétrica, las series de datos cuantitativos tienen que cumplir algunas condiciones, como: tamaño óptimo de muestra, distribución normal de los datos, homogeneidad de las varianzas de los tratamientos, independencia de datos y aditividad en las respuestas experimentales. En estas condiciones se puede aplicar con certeza las pruebas paramétricas, entre ellas: F, t, correlación r de Pearson y el análisis de varianza.

En el caso de muestras pequeñas (ejemplo:  $n = 6$ ), reducido número de grados de libertad del error en los experimentos ( $GLE < 12$ ), datos provenientes de mediciones con escalas nominales y ordinales y distribuciones alejadas de curva normal (distribución libre), solo es pertinente el uso de técnicas no paramétricas, entre ellas: pruebas  $\chi^2$ , binomial,

Poisson y de Fisher, correlación de Spearman y los análisis de varianza por rangos de Friedman y de Kruskal- Wallis.

#### 4.7.1 PRUEBAS DE LA NORMALIDAD

La normalidad de una serie de datos, se valora a través del histograma, observándose una tendencia de la forma de campana gaussiana. La distribución normal varía de  $-\infty$  a  $+\infty$ , con forma simétrica. El sesgo de la curva se mide con el Coeficiente de asimetría (A), el apuntamiento lo define la Curtosis y la posición de la curva sobre la abscisa lo determina la media ( $\mu$ ).

Se han desarrollado varios métodos para probar la normalidad, entre ellos, los de Jarque-Bera (JB) y de Kolmogorov-Smirnov (KS).

##### 4.7.1.1 PRUEBA DE JARQUE Y BERA DE LA NORMALIDAD

La prueba de Jarque - Bera (1987) es una prueba de bondad de ajuste usada para comprobar si una serie de datos tiene la asimetría y la curtosis característicos de una curva normal (p. 167). El estadístico JB se calcula con la siguiente fórmula:

$$JB = \frac{n}{6} \left[ A^2 + \frac{1}{4}(K)^2 \right] \quad [78]$$

Dónde:

JB = Estadístico de normalidad

n = Número de observaciones

A = Coeficiente de Asimetría

K = Coeficiente de Curtosis

Las fórmulas para calcular los Coeficientes Asimetría (A) y Curtosis (K) son:

$$A = \frac{\frac{1}{n} \sum (X_i - \bar{X})^3}{\left[ \frac{1}{n} (\sum (X_i - \bar{X})^2) \right]^{3/2}} \quad K = \frac{\frac{1}{n} \sum (X_i - \bar{X})^4}{\left[ \frac{1}{n} (\sum (X_i - \bar{X})^2) \right]^2}$$

La significación se determina contrastando el estadístico *JB* con Chi-cuadrado, considerando el nivel de significación  $\alpha$  y los grados de libertad (GL), predefinidos.

En Excel, el valor *p* de  $\chi^2$  se obtiene con la función:

=DISTR.CHI(*JB*;1).

La hipótesis estadística se formula como:

H<sub>0</sub>: La distribución es normal

H<sub>1</sub>: La distribución difiere de la normal

La regla de decisión es la siguiente:

Si  $JB < \chi^2_{0,05}$ : Aceptar H<sub>0</sub>

Si  $JB \geq \chi^2_{crítico}$ : Aceptar H<sub>1</sub>

**Ejercicio 86:**

Probar la normalidad de cuatro clones de café usando el método de Jarque-Bera (1987).

**Datos y análisis:**

Los datos Asimetría y Curtosis, así como la prueba JB, se exponen en el Cuadro 43.

**Hipótesis:**

H<sub>0</sub>: La distribución es normal

H<sub>1</sub>: La distribución difiere de la normal

**Regla de decisión:**

Si  $JB < \chi^2_{crítico}$ : Aceptar H<sub>0</sub>

Si  $JB \geq \chi^2_{crítico}$ : Aceptar H<sub>1</sub>

**Valor  $\chi^2_{crítico}$ :**

Para  $\alpha = 0,05$  y  $GL = 1$ , el valor crítico  $\chi^2_{0,05} = 3,841$

**Fórmula JB:**

$$JB = \frac{n}{6} \left[ A^2 + \frac{1}{4} (K)^2 \right]$$

JB para clon COF-03-A7

$$JB = \frac{10}{6} \left[ (1,44)^2 + \frac{1}{4} (1,43)^2 \right] = 4,30 > \chi^2_{0,05} = 3,841: \text{ Aceptar } H_1$$

JB para clon COF-04-A7

$$JB = \frac{10}{6} \left[ (0,12)^2 + \frac{1}{4} (-0,95)^2 \right] = 0,40 < \chi^2_{0,05} = 3,841: \text{ Aceptar } H_0$$

JB para clon COF-LE-A1

$$JB = \frac{10}{6} \left[ (-1,42)^2 + \frac{1}{4} (-0,32)^2 \right] = 3,403 < \chi^2_{0,05} = 3,841: \text{ Aceptar } H_0$$

JB para clon NP-3018-A19

$$JB = \frac{10}{6} \left[ (0,75)^2 + \frac{1}{4} (-0,59)^2 \right] = 1,083 < \chi^2_{0,05} = 3,841: \text{ Aceptar } H_0$$

Cuadro 43. Prueba de la normalidad de cuatro clones de café robusta con el método de Jarque-Bera

Estadísticos	COF-03-A7	COF-04-A7	LE-A1	NP-3018-A19
N	10	10	10	10
Media (kg.ha <sup>-1</sup> )	1257	1248	1686	956
Coficiente de asimetría (A)	1,44	0,12	-1,42	0,75
Coficiente de Curtosis (K)	1,43	-0,95	-0,32	-0,59
n/6	1,67	1,67	1,67	1,67
n/4	0,25	0,25	0,25	0,25
A <sup>2</sup>	2,07	0,01	2,02	0,56
K <sup>2</sup>	2,0449	0,9025	0,1024	0,3481
JB	4,308	0,400	3,403	1,083
$\chi^2$ con $\alpha = 0,05$ y $GL = 1$	3,841	3,841	3,841	3,841
p de Chi-cuadrado	0,038	0,527	0,065	0,298
Distribución	No normal	Normal	Normal	Normal

**Decisión:**

Los clones COF 04-A7, LE-A1 y NP 3018-A19 tienen distribución normal. La serie de datos del clon COF 03-A7 no tiene distribución normal.

#### 4.7.1.2 PRUEBA DE KOLMOGOROV-SMIRNOV DE LA NORMALIDAD

La prueba de Kolmogorov-Smirnov (KS) es una de las pruebas no paramétricas de “bondad de ajuste” usada para docimar la hipótesis de que una variable tiene distribución normal (Quintero y Duran, 2004, p. 46). El concepto “docimar” es la suposición de que la hipótesis nula es verdadera hasta encontrar la evidencia estadística que permita decidir lo contrario (Cavada, 2009, 256). Esta prueba posibilita determinar si los puntajes de una serie de datos provienen de una población con distribución normal (Martínez y Martínez, 2010).

La fórmula para calcular el estadístico KS es la siguiente:

$$D = \text{Máx}|FRA - SRAE| \quad [79]$$

Dónde:

$D_{\text{calculada}}$  = Estadístico KS

FRA = Frecuencia relativa acumulada observada

SRAE = Frecuencia relativa acumulada esperada

Máx = Valor absoluto más alto de la diferencia: FRA - SRAE.

Para probar la significación se contrasta  $D_{\text{calculada}}$  con  $D_{\text{crítico}}$  con un nivel de significación  $\alpha = 0,05$  o  $0,01$ .

El valor  $D_{\text{crítico}}$  se obtiene en la tabla KS o usando la siguiente fórmula:

$$D_{\text{crítico}} = \sqrt{\frac{\ln\left(\frac{\alpha}{2}\right)}{\frac{2}{n}}} \quad [80]$$

Dónde:

$D_{\text{crítico}}$ : Estadístico de Kolmogorov y Smirnov para un nivel de significación  $\alpha$  dado.

$\ln$  = Logaritmo natural o neperiano cuya base es  $e = 2,71828$

$\alpha$  = Nivel de significación fijado por el investigador (0,05 o 0,01)

$n$  = Número de observaciones

La hipótesis estadística se formula como:

$H_0$ : La distribución es normal

$H_1$ : La distribución difiere de la normal

La regla de decisión es la siguiente:

Si  $D_{\text{calculada}} < D_{\text{crítica}}$ : Aceptar  $H_0$

Si  $D_{\text{calculada}} \geq D_{\text{crítica}}$ : Aceptar  $H_1$

#### Ejercicio 87:

Determinar la normalidad de la distribución de frecuencias del rendimiento ( $\text{kg}\cdot\text{ha}^{-1}$ ) del clon de café robusta LE-A1, con un nivel de confianza del 95 %, usando la prueba de Kolmogorov-Smirnov.

#### Datos y análisis:

La prueba de Kolmogorov y Smirnov (KS) se realiza por partes, individualmente de clon en clon, en base a su análisis de frecuencias (Cuadro 44). Se detalla el procedimiento para la prueba KS de la normalidad solo para el clon LE-A1. Para las pruebas de los otros clones, se repite el proceso.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 44. Análisis de frecuencias absolutas de los clones de café robusta en estudio*

Clases (kg.ha <sup>-1</sup> )	COF-03-A7	COF-04-A7	NP-3018-A19	LE-A1	Σ
<600	0	2	1	0	3
601-1200	5	2	7	3	17
1201-1800	4	4	2	2	12
1801-2400	0	2	0	4	6
> 2400	1	0	0	1	2
Observaciones (n)	10	10	10	10	40

**Hipótesis estadísticas:**

H<sub>0</sub>: La distribución de frecuencias del clon LE-A1 es normal

H<sub>1</sub>: La distribución de frecuencias del clon LE-A1 difiere de la normal

**Regla de decisión:**

Si D calculado < D crítico: Aceptar H<sub>0</sub>

Si D calculado ≥ D crítico: Aceptar H<sub>1</sub>

El procedimiento de la prueba Kruskal-Wallis (KS) para la variable rendimiento del clon LE-A1, se indica en el Cuadro 45.

*Cuadro 45. Prueba KS de la normalidad de los datos de rendimiento del clon LE-A1 de café robusta*

Estadísticos	Clases (kg.ha <sup>-1</sup> )					n
	<600	601-1200	1201-1800	1801-2400	>2400	
Frecuencia absoluta (FA)	0	3	2	4	1	10
Frecuencia absoluta acumulada (FAA)	0	3	5	9	10	--
Frecuencia relativa (FR)	0,0	0,30	0,20	0,40	0,10	1
Frecuencia relativa acumulada (FRA)	0,0	0,30	0,50	0,90	1,00	--
Frecuencia absoluta según H <sub>0</sub> (SAE)	2	2	2	2	2	10
Frecuencia relativa según H <sub>0</sub> (SRE)	0,2	0,2	0,2	0,2	0,2	1
Frecuencia relativa acumulada según H <sub>0</sub> (SRAE)	0,20	0,40	0,60	0,80	1,00	--
FRA – SRAE	-0,20	-0,10	-0,10	0,10	0,00	

$D_{calculada} = Máx|FRA - SRAE| = 0,20$ . (Valor absoluto)

**Cálculo del valor crítico D<sup>12</sup>:**

$$D_{crítica} = \sqrt{\frac{\ln(\frac{\alpha}{2})}{\frac{2}{n}}} \rightarrow D_{0,05} = \sqrt{\frac{\ln(0,025)}{\frac{2}{10}}} \rightarrow D_{0,05} = \sqrt{\frac{3,69}{2}} \rightarrow D_{0,05} = \sqrt{\frac{1,845}{10}} = 0,43 \rightarrow D_{0,05} = 0,43$$

**Hipótesis estadísticas:**

H<sub>0</sub>: La distribución es normal

H<sub>1</sub>: La distribución difiere de la normal

**Regla de decisión:**

Si  $D_{calculada} < D_{0,05}$ : Aceptar  $H_0$  con el 95 % de confianza

Si  $D_{calculada} \geq D_{0,05}$ : Aceptar  $H_1$  con el 95 % de confianza

**Interpretación:**

$D_{calculada} = 0,20 < D_{0,05} = 0,43$ : Aceptar  $H_0$

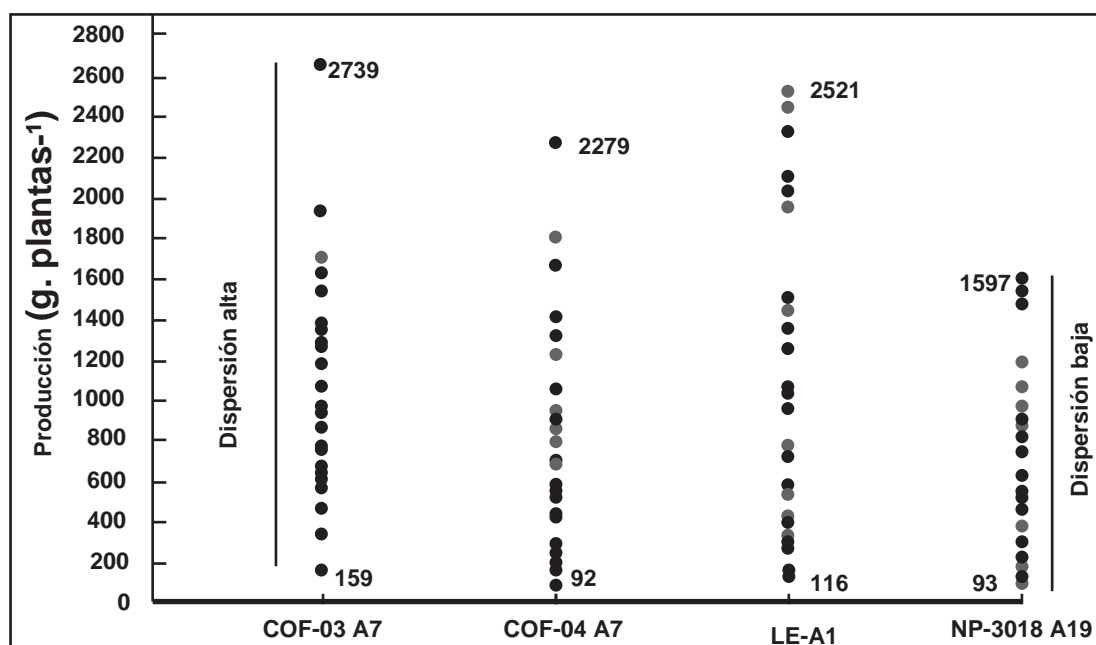
**Decisión:**

El clon LE-A1 tiene distribución normal con el 95 % de confianza.

4.7.2 PRUEBAS DE LA HOMOCEDASTICIDAD

Las transformaciones de datos se realizan para corregir la asimetría en la distribución y homogenizar las varianzas, condiciones que deben verificarse previo a las pruebas estadísticas paramétricas (Di Rienzo et al., 2009, p. 349). La heterogeneidad de las varianzas se asocia a la ausencia de normalidad. Cuando se comprara las varianzas de dos tratamientos, se usa la prueba F, aunque se puede extender el método, probando de dos en dos. En cambio, cuando se desea probar la homocedasticidad de todo un conjunto de tratamientos se usa la prueba de Levene.

La elaboración de un dispersograma<sup>13</sup> por tratamientos, en un experimento, posibilita valorar el comportamiento de las series de datos detectar potenciales datos atípicos que pudiesen tener efectos en el incremento de la varianza. En el Gráfico 24, se indica un dispersograma de los cuatro clones de café, observándose que NP 3018-A19 tiene una dispersión de 93 a 1597 g.planta<sup>-1</sup> y con seguridad menor varianza, en comparación de los otros clones.



Fuente: Duicela (2021)

Gráfico 24. Dispersograma de los datos de producción de cuatro clones de café robusta

<sup>12</sup> El valor D crítico se calcula con la fórmula específica, verificándose la coincidencia con los valores críticos de la tabla Kolmogorov-Smirnov.

<sup>13</sup> El dispersograma es una herramienta eficaz para validar la calidad de los datos

4.7.2.1 PRUEBA F DE HOMOGENEIDAD DE LAS VARIANZAS

La homogeneidad de las varianzas se prueba con el estadístico F (Di Rienzo et al., 2009, p. 268, p. 170). La fórmula para calcular F es la siguiente:

$$F_{\text{calculada}} = \frac{\text{Varianza mayor}}{\text{Varianza menor}} \quad [81]$$

La prueba F requiere, del cálculo previo, de las varianzas de las muestras ( $S^2$ ).

Los valores críticos F se obtienen en la tabla específica<sup>14</sup>, para un nivel de significación  $\alpha$  y grados de libertad (GL) del numerador ( $n_1 - 1$ ) y del denominador ( $n_2 - 1$ ). Al comparar las varianzas de un grupo de tratamientos, se identifica aquel tratamiento de menor varianza con el que se contrastan los demás, aunque también puede usarse el de mayor varianza, debiendo explicarse en la interpretación.

En el programa Excel, los valores críticos de F, se obtienen con la función: =DISTR.F.INV( $\alpha$ ;grados de libertad del numerador; grados de libertad del denominador).

La hipótesis estadística se define como:

H<sub>0</sub>: Las varianzas son homocedásticas

H<sub>1</sub>: Las varianzas son heterocedásticas

La regla de decisión es la siguiente:

Si  $F_{\text{calculada}} < F_{\text{crítico}}$ : Aceptar H<sub>0</sub>

Si  $F_{\text{calculada}} \geq F_{\text{crítico}}$ : Aceptar H<sub>1</sub>

Tener presente que el valor  $F_{\text{crítico}}$  depende del nivel de significación  $\alpha$  y de los grados de libertad ( $n - 1$ ), tanto del numerador como del denominador.

**Ejercicio 88:**

Probar la homogeneidad de las varianzas usando la prueba F, en los datos de producción de cuatro clones de café robusta en Ecuador. Se toma como referencia el de menor varianza.

**Datos y análisis:**

La prueba de la homogeneidad de las varianzas, se indica en el Cuadro 46.

Cuadro 46. Prueba F de la homocedasticidad de tres clones de café robusta respecto del NP-3018.

Estadísticos	COF-03-A7	COF-04-A7	LE-A1	NP-3018-A19
<i>n</i>	33	33	33	33
Varianza	286.006	267.685	504.882	187.548
F calculada	1,525 ns	1,427 ns	2,692 **	1,000
F <sub>0,05</sub>	1,804	1,804	1,804	1,804
F <sub>0,01</sub>	2,318	2,318	2,318	2,318
Tipo de varianza	Homocedástica	Homocedástica	Heterocedástica	
p de F	0,119	0,160	0,003	

Nota: ns = no hay diferencia entre varianzas, \*\* Hay diferencia altamente significativa entre varianzas

**Decisión:**

El clon NP-3018-A19 mostró la menor varianza. La varianza del clon LE-A1 resultó heterocedástica con respecto a NP-3018-A19 (\*\*). Los clones COF-03 y COF-04 resultaron homocedásticas con relación a NP-3018.

<sup>14</sup> En el Anexo 6 se indica el procedimiento de lectura en la tabla F para obtener los valores críticos



#### 4.7.2.2 PRUEBA DE LEVENE DE HOMOGENEIDAD DE LAS VARIANZAS

La prueba de Levene se usa para determinar la homogeneidad de las varianzas de dos o más tratamientos, en variables cuantitativas. Los datos originales ( $X_i$ ) se transforman a nuevas series de datos (Levene data), con los cuales se realiza un análisis de varianza.

La fórmula para transformar los datos originales a la Levene data es:

$$Y = |X_i - \tilde{X}| \quad [82]$$

Dónde:

Y = Data Levene para la variable n estudio

$X_i$  = Observaciones individuales por tratamiento

$\tilde{X}$  = Valor de la mediana de cada uno de los tratamientos

Esta transformación se fundamenta en que la mediana y la media tienden a coincidir en el mismo punto cuando la distribución es normal, situación que se explica por la forma simétrica de la curva y por la intensidad del apuntamiento. Los datos de la variable experimental, de cada tratamiento, se organizan en una tabla de medias, se calcula la mediana por tratamiento y posteriormente se calculan las diferencias de las medias con la respectiva mediana de cada tratamiento.

Los datos originales transformados a los valores de las diferencias  $X_i - \tilde{X}$  de cada tratamiento se identifica como “Levene-Data” con cuyos valores se realiza el análisis de varianza, interpretando el valor  $p$  o la significación estadística del contraste  $F_{calculada}$  con los  $F_{críticos}$ , si hay o no diferencias estadísticas significativas.

Las hipótesis estadísticas se conjeturan como:

$H_0$ : Las varianzas son homogéneas

$H_1$ : Las varianzas no son homogéneas

La regla de decisión se planea como:

Si  $p > \alpha$ : Aceptar  $H_0$

Si  $p \leq \alpha$ : Aceptar  $H_1$

#### **Ejercicio 89:**

Probar la homogeneidad de las varianzas en tres clones de café robusta, de la “Levene-data” en la variable producción ( $\text{kg.planta}^{-1}$ ).

#### **Datos y análisis:**

La “Levene-data” (Cuadro 47) corresponden a las diferencias entre las medias de las observaciones individuales y las medianas de los tratamientos de la variable producción.

#### **Hipótesis estadísticas:**

$H_0$ : Las varianzas de los tres clones de café robusta son iguales

$H_1$ : Las varianzas de los tres clones de café robusta son desiguales

#### **Regla de decisión:**

Si  $p \text{ valor} > \alpha = 0,05$ : Aceptar  $H_0$

Si  $p \text{ valor} \leq \alpha = 0,05$ : Aceptar  $H_1$

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

Cuadro 47. “Levene-data” (Y) de la producción de tres clones experimentales de café robusta

Datos originales de producción X: kg.planta <sup>-1</sup>				Data Levene de producción Y = $ X_i - \bar{X} $			
Plantas	COF-O 06	CON- ERB 01	NP-2024	Plantas	COF-O 06	CON- ERB 01	NP- 2024
1	1,19		0,67	1	0,16		0,12
2	1,01		0,96	2	0,02		0,17
3	1,24	1,00	0,94	3	0,21	0,27	0,15
4	0,94	1,24	0,69	4	0,09	0,03	0,09
5	0,92	1,19	0,82	5	0,11	0,08	0,04
6	0,93	0,94	1,05	6	0,10	0,33	0,26
7	0,83	1,27		7	0,20	0,00	
8	0,71	1,45	0,78	8	0,32	0,18	0,01
9	1,11	1,28		9	0,08	0,02	
10	1,22	1,63	0,97	10	0,19	0,36	0,18
11	1,07	0,92	0,83	11	0,05	0,34	0,04
12	1,03	1,58	0,45	12	0,00	0,32	0,34
13	1,01	1,51	0,79	13	0,02	0,24	0,00
14	1,31	1,19	0,45	14	0,28	0,08	0,34
15	1,20	1,03	0,62	15	0,17	0,24	0,17
16	1,33	1,70	0,71	16	0,30	0,43	0,08
17	1,03	1,41	0,86	17	0,00	0,15	0,07
$\bar{X}$	1,06	1,29	0,77				
$\bar{X}$	1,03	1,27	0,79				

**Análisis de la varianza:**

El ADEVA de la “Levene-data”, se expone en el Cuadro 48.

Cuadro 48. Análisis de varianza de las diferencias entre medias y medianas o data de Levene

Fuente de variación	GL	SC	CM	F calculada	F <sub>0,05</sub>	Valor p
Clones	2	0,047	0,024	1,71 ns	3,21	0,199
Error experimental	44	0,614	0,014			
Total	46	0,661				

Nota: ns = no hay diferencia estadística significativa entre las varianzas de los tres clones.

**Interpretación:**

$F_{\text{calculada}} = 1,71 < F_{0,05} = 3,21$ : Aceptar  $H_0$

Valor  $p = 0,199 > \alpha = 0,05$ : Aceptar  $H_0$

**Decisión:**

Las varianzas de los tres clones de café robusta son homocedásticas.

#### 4.7.3 ADITIVIDAD DE LOS EFECTOS

En un modelo como:  $Y_i = \mu \pm \epsilon_i$ , el valor de una observación corresponde al promedio  $\pm$  error de muestreo. Al estudiar un factor (diseño DCA), como la dosis de fertilizante nitrogenado (F), el modelo adiciona el efecto del fertilizante:  $Y_i = \mu + F \pm \epsilon_i$ . Si además del fertilizante (F), se asume que hay efecto de los bloques (diseño BA), el modelo incorpora el efecto de bloques (B), por lo tanto:  $Y_i = \mu + F + B \pm \epsilon_i$

En los referidos modelos, se asume que hay aditividad de los efectos, evidenciándose que el total es igual a la suma de los efectos parciales. Pero hay respuestas con efectos antagónicos o sinérgicos. Cuando los efectos de un factor potencian a otro factor ocurre la sinergia pero cuando se atenúan o merman ocurre el antagonismo (Melo et al., 2020, p. 255). Al constatar sinergias o antagonismos en las series de datos, el análisis de varianza paramétrico pierde validez, debiendo usar como la alternativa el análisis de regresión.

La heterosis, en fitomejoramiento, es un caso de sinergia, que expresa la superioridad observada en la primera generación filial (F1), respecto del promedio de sus progenitores. En fertilizantes se ha evidenciado sinergias entre nitrógeno y manganeso, fósforo y magnesio, azufre y nitrógeno.

En el antagonismo se evidencia que una sustancia química frena la acción de otra, como es el caso del potasio y boro, del fósforo y potasio, del calcio y magnesio, del calcio y zinc, y del azufre y boro.

#### 4.7.4 PRUEBA DE RACHAS DE ALEATORIEDAD

Esta prueba es una técnica no paramétrica usada para verificar la aleatoriedad en una muestra, en base a puntajes, sin hacer supuestos acerca de la forma de la distribución. Se considera como una extensión de la prueba Z.

La fórmula usada para la prueba de rachas es la siguiente:

$$Z = \frac{r - \left(\frac{2n_1 n_2}{n_1 + n_2} + 1\right)}{\sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}} \quad [83]$$

Dónde:

Z = Estadístico de rachas que se contrasta con  $Z_{crítica}$

r = Número de rachas o grupos sujetos a la prueba

$n_1$  = Número de elementos de una categoría

$n_2$  = Número de elementos de la categoría alterna

Las hipótesis que se prueban son:

$H_0$ : La distribución si es aleatoria

$H_1$ : La distribución no es aleatoria

La regla de decisión es la siguiente:

Si  $p$  calculada de  $Z_{(1-\frac{\alpha}{2})} \leq 0,025$ : Aceptar  $H_0$

Si  $p$  calculada de  $Z_{(1-\frac{\alpha}{2})} > 0,025$ : Aceptar  $H_1$

**Ejercicio 90:**

Determinar si hay aleatoriedad en la distribución de plantas de dos especies de frutales en un huerto (N = 50) con el 95 % de confianza. Se constatan 30 plantas de la especie A y 20 de la especie B.

**Datos y análisis:**

En el Cuadro 49, se presenta la distribución de las dos especies frutales en el huerto.

Cuadro 49. Croquis del huerto con las dos especies frutales

<u>A</u> <sup>1</sup>	<u>B</u> <sup>2</sup>	<u>A</u> <sup>3</sup>	<u>B</u> <sup>4</sup>	<u>A</u> <sup>5</sup>	<u>A</u> <sup>5</sup>	<u>A</u> <sup>5</sup>	<u>B</u> <sup>6</sup>	<u>B</u> <sup>6</sup>	<u>A</u> <sup>7</sup>	<u>B</u> <sup>8</sup>	<u>A</u> <sup>9</sup>	<u>B</u> <sup>10</sup>
<u>A</u> <sup>11</sup>	<u>B</u> <sup>12</sup>	<u>A</u> <sup>13</sup>	<u>A</u> <sup>13</sup>	<u>A</u> <sup>13</sup>	<u>A</u> <sup>13</sup>	<u>B</u> <sup>14</sup>	<u>B</u> <sup>15</sup>	<u>B</u> <sup>16</sup>	<u>A</u> <sup>17</sup>	<u>B</u> <sup>18</sup>	<u>A</u> <sup>19</sup>	<u>B</u> <sup>19</sup>
<u>B</u> <sup>20</sup>	<u>B</u> <sup>20</sup>	<u>B</u> <sup>20</sup>	<u>A</u> <sup>24</sup>	<u>B</u> <sup>22</sup>	<u>A</u> <sup>23</sup>	<u>B</u> <sup>24</sup>	<u>B</u> <sup>25</sup>	<u>B</u> <sup>26</sup>	<u>A</u> <sup>27</sup>	<u>A</u> <sup>27</sup>	<u>B</u> <sup>28</sup>	
<u>A</u> <sup>29</sup>	<u>A</u> <sup>29</sup>	<u>B</u> <sup>30</sup>	<u>A</u> <sup>31</sup>	<u>A</u> <sup>31</sup>	<u>A</u> <sup>31</sup>	<u>A</u> <sup>31</sup>	<u>B</u> <sup>32</sup>	<u>A</u> <sup>33</sup>	<u>B</u> <sup>34</sup>	<u>A</u> <sup>35</sup>	<u>A</u> <sup>35</sup>	

Fuente: Siegel (1990, p. 80)

Especie A : n1 = 30  
 Especie B : n2 = 20  
 n : n1 + n2 = 50  
 Rachas : r = 35

**Cálculo de Z:**

$$Z = \frac{r - \left(\frac{2n_1 n_2}{n_1 + n_2} + 1\right)}{\sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}} \rightarrow Z = \frac{35 - \left(\frac{2(30)(20)}{30 + 20} + 1\right)}{\sqrt{\frac{2(30)(20)[(2)(30)(20) - 30 - 20]}{(30 + 20)^2 (30 + 20 - 1)}}} = 2,98$$

En Excel, se obtienen los valores p de Z con la función:

=DISTR.NORM.ESTAND (Z)

Valor p para una cola = 0,00144 y valor p dos colas = 0,00288

El valor acumulado p = 0,9986 de la tabla de valores críticos, hay que restarlo de la unidad.

Para la prueba de una cola: p = 1 - 0,9986 = 0,00144.

Para la prueba de dos colas (se multiplica por 2): p = (0,00144 )(2) = p = 0,00288.

**Nivel de significación:**

El nivel de significación se fija en α = 0,05 para la prueba de dos colas.

**Hipótesis estadística:**

H<sub>0</sub>: La distribución de las dos especies frutales en el huerto es aleatoria

H<sub>1</sub>: La distribución de las dos especies frutales en el huerto no es aleatoria

**Regla de decisión:**

Si p (Z) ≥ α/2 = 0,025: Aceptar H<sub>0</sub>

Si p (Z) < α/2 = 0,025: Aceptar H<sub>1</sub>

**Interpretación:**

Valor  $p(Z) = 0,00289 < \alpha = 0,025$ : Aceptar  $H_1$

**Decisión:**

La distribución de las dos especies frutales en el huerto no es aleatoria, con el 95 % de confianza.

**4.7.5 PRUEBA DE LOS SIGNOS DE ALEATORIEDAD**

La prueba de los signos es una técnica no paramétrica usada para comparar dos muestras relacionadas, basada en los signos (+ y -) adquiridos por las diferencias en las series de datos. Esta técnica es una adaptación de la prueba de rachas, donde se aplica la misma fórmula y las rachas (grupos) corresponden a los signos.

$$Z = \frac{r - \left(\frac{2n_1 n_2}{n_1 + n_2} + 1\right)}{\sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}$$

Dónde:

$Z$  = Estadístico de rachas que se contrasta con  $Z_{\text{crítico}}$

$r$  = Número de rachas o grupos sujetos a la prueba

$n_1$  = Número de elementos de una categoría

$n_2$  = Número de elementos de la categoría alterna

El nivel de significación  $\alpha$  se fija en 1 % o 5 %. Esta prueba corresponde a dos colas.

Las hipótesis estadísticas se formulan como:

$H_0$ : Las diferencias entre muestras se deben al azar

$H_1$ : Las diferencias entre muestras no se deben al azar

La regla de decisión es la siguiente:

Si  $p$  calculada de  $Z_{(1-\frac{\alpha}{2})} \leq 0,025$ : Aceptar  $H_0$

Si  $p$  calculada de  $Z_{(1-\frac{\alpha}{2})} > 0,025$ : Aceptar  $H_1$

**Ejercicio 91:**

En dos tiendas de productos agro artesanales se vende café tostado y molido en empaques de 400 gramos. En un registro de 10 semanas consecutivas se evidencia diferencias en las ventas. Determinar si esas diferencias se deben a la casualidad.

**Datos y análisis:**

En el Cuadro 50, se indica la secuencia de los datos de paquetes de café molido vendidos en 10 semanas consecutivas, las diferencias  $B - A$  y el conteo de rachas.

Cuadro 50. Ventas en las tiendas A y B, diferencias, signos y rachas de los signos

	Número de semanas										Media de ventas
	1	2	3	4	5	6	7	8	9	10	
Ventas en la tienda A	345	450	500	600	450	356	650	500	780	420	505
Ventas en la tienda B	500	400	600	500	600	390	700	700	540	470	540
Diferencia	155	- 50	100	- 100	150	34	50	200	- 240	50	35
Signos	+	-	+	-	+	+	+	+	-	+	
Rachas de signos	(1)	(2)	(3)	(4)	(5)				(6)	(7)	

Nota: Las rachas de signos (grupos) se determinan de izquierda a derecha, en forma secuencial

Total de semanas	: n = 10
Semanas donde tienda B tiene mayores ventas (+)	: n <sub>1</sub> = 7
Semanas donde tienda B tiene menores ventas (-)	: n <sub>2</sub> = 3
Rachas de signos	: r = 7

**Cálculo de Z:**

$$Z = \frac{r - \left(\frac{2n_1 n_2}{n_1 + n_2} + 1\right)}{\sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}} \rightarrow Z = \frac{7 - \left(\frac{2(7)(3)}{7+3} + 1\right)}{\sqrt{\frac{2(7)(3)[(2)(7)(3) - 7 - 3]}{(7+3)^2 (7+3-1)}}} \rightarrow Z = \frac{7 - \left(\frac{42}{10} + 1\right)}{\sqrt{\frac{42[32]}{(10)^2 (9)}}} = 1,473$$

El valor *p* de Z acumulada = 0,9296 8 (tabla de distribución normal estandarizada)  
 Este valor restado de la unidad corresponde a *p* de una cola: 1 – 0,9296 = 0,0704.

El valor *p* en la prueba de dos colas es: 0,0704 (2) = 0,141 ~ 14,1 %.

**Hipótesis estadística:**

H<sub>0</sub>: Las diferencias en ventas es aleatoria, ocurren por casualidad.

H<sub>1</sub>: Las diferencias en ventas no es aleatoria, habría otros factores que lo afectan

**Regla de decisión:**

Si  $p(Z) \geq \frac{\alpha}{2} = 0,025$ : Aceptar H<sub>0α</sub>

Si  $p(Z) < \frac{\alpha}{2} = 0,025$ : Aceptar H<sub>1</sub>

**Interpretación:**

Valor  $p(Z) = 0,141 > \alpha = 0,025$ : Aceptar H<sub>0</sub>

**Decisión:**

No hay diferencia estadística significativa en las ventas, entre las dos tiendas. La diferencia es resultado de la casualidad.

**4.8 PRUEBAS DE LA INDEPENDENCIA**

Las pruebas de independencia implican el uso de una tabla de valores observados para establecer si hay dependencia o independencia entre los grupos a compararse.

**4.8.1 PRUEBA DE LA INDEPENDENCIA CON LA t de WELCH**

La suposición de independencia implica que las perturbaciones de series separadas no se encuentran sistemáticamente relacionadas. La aleatorización, en la asignación de tratamientos y en el muestreo, y el evitar las autocorrelaciones entre variables, asegura la independencia de las series de datos (Di Rienzo et al., 2009, p. 42). Un ejemplo de autocorrelación en granos es el peso seco de 100 granos (al 10 % de humedad) y el índice de semilla (peso promedio de un grano). La independencia de los datos se puede probar usando la *t* de Student cuando hay homocedasticidad, normalidad y sin datos faltantes en las series o con la prueba *t* de Welch cuando hay desigual número de observaciones.

Para calcular la *t* de Welch se usa la fórmula:

$$T_{W \text{ calculada}} = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \quad [84]$$

Dónde:

$T_W$  = t de Welch calculada

$\bar{X}_A$  = Media de la serie de datos A

$\bar{X}_B$  = Media de la serie de datos B

$S^2_A$  = Varianza de la serie de datos A

$S^2_B$  = Varianza de la serie de datos B

$n_A$  = Número de observaciones en la serie A

$n_B$  = Número de observaciones en la serie B

La t de Welch se basa en un ajuste o corrección de los grados de libertad. Para obtener el valor crítico de t, en este caso particular, luego de especificar el nivel de significación ( $\alpha$ ), se modifica los grados de libertad, usando la fórmula siguiente:

$$GL = \frac{\left(\frac{S^2_A}{n_A} + \frac{S^2_B}{n_B}\right)^2}{\frac{1}{n_A - 1} \left(\frac{S^2_A}{n_A}\right)^2 + \frac{1}{n_B - 1} \left(\frac{S^2_B}{n_B}\right)^2} \quad [85]$$

Dónde:

GL = Grados de libertad

$S^2_A$  = Varianza del tratamiento A,  $S^2_B$  = Varianza del tratamiento B

$n_A$  = Observaciones en el tratamiento A,  $n_B$  = Observaciones en el tratamiento B

Las hipótesis estadísticas se formulan como:

$H_0$ : Hay independencia

$H_1$ : No hay independencia

La regla de decisión es la siguiente:

Si  $t_{calculada} < t_{crítica}$ : Aceptar  $H_0$

Si  $t_{calculada} \geq t_{crítica}$ : Aceptar  $H_1$

### Ejercicio 92:

Determinar si hay independencia entre las series de datos de dos clones de café robusta, usando la prueba t de Welch.

#### Datos y análisis:

En el Cuadro 51, se indican los datos y cálculos de la t de Welch.

Cuadro 51. Prueba t de Welch de la independencia entre dos clones de café robusta

Estadísticos	Clones de café robusta	
	LE-A1	NP-3018-A19
n	33	33
Media de café oro (g.planta <sup>-1</sup> )	1006	703
Desviación estándar	711	433
t de Welch (calculado)	2,092	--
$t_{0,05}$	2,308	--

#### Interpretación:

$t_{calculada} = 2,092 < t_{crítica} = 2,308$ : Aceptar  $H_0$ .

#### Decisión:

Las series de datos de los clones LE-A1 y NP-3018-A19 son independientes con el 95 % de confianza.

#### 4.8.2 PRUEBA U DE MANN-WHITNEY

Esta es una técnica no paramétrica alternativa a la prueba t usada para comparar dos tratamientos, con igual o desigual número de observaciones, que no requiere del cumplimiento de las condiciones exigidas por la estadística paramétrica, referidas al tamaño de muestra, distribución normal y homocedasticidad.

Los datos pueden ser cualitativos o cuantitativos. La prueba se basa en la asignación de rangos a los datos de los dos tratamientos y posterior reagrupamiento en sumas de los rangos/tratamiento (R1 y R2). Con esta información se calcula los estadísticos U. La significación estadística se determina usando la tabla de valores críticos U o la prueba Z. En este caso se requiere calcular los estadísticos:  $\mu_U$  y  $\sigma_U$  que corresponden a la media y desviación estándar de Mann-Whitney (MW)

Las fórmulas para calcular U se adaptan a las sumas de los rangos por tratamiento:

Cuando:  $R_2 > R_1$ , el estadístico U de Mann-Whitney se calcula con la fórmula:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \quad [86]$$

Cuando:  $R_1 > R_2$ , el estadístico U de Mann-Whitney se calcula con la fórmula:

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 \quad [87]$$

Dónde:

U = Estadístico de Mann-Whitney

$n_1$  = Número de observaciones para el tratamiento 1

$n_2$  = Número de observaciones para el tratamiento 2

R1 y R2 = Suma de los rangos de los dos tratamientos comparados

El valor Z calculado se obtiene con la fórmula:

$$Z = \frac{U - \mu_U}{\sigma_U} \quad [88]$$

Dónde:

Z = Estadístico del contraste

U = Estadístico de Mann-Whitney

$\mu_U$  = Media

$\sigma_U$  = Desviación estándar

La media MW se calcula con la fórmula:

$$\mu_U = \frac{n_1 n_2}{2} \quad [89]$$

La desviación estándar MW se calcula con la fórmula:

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \quad [90]$$



El valor  $Z_{\text{calculado}}$  se contrasta con  $Z_{\text{crítico}}$  obtenido de la tabla de distribución normal.

El Excel, el valor  $p$  de  $Z$  se obtiene directamente con la función:

$$=DISTR.NORM.ESTAND(Z_{\text{calculado}})$$

**Ejercicio 93:**

Usando la prueba U con el 95 % de confianza, comparar los promedios de frutos.planta<sup>-1</sup> en función de dos frecuencias de aplicación en prefloración de bioestimulantes al follaje.

**Datos y análisis:**

En el Cuadro 52, se indican los datos de frutos.planta<sup>-1</sup> como respuesta a dos tratamientos de bioestimulantes aplicados en prefloración.

Cuadro 52. Frutos por planta y transformación a rangos de dos tratamientos con bioestimulantes

Tratamiento 1		Tratamiento 2	
Frutos.planta <sup>-1</sup>	Rango T1	Frutos.planta <sup>-1</sup>	Rango T2
45	1	51	2
64	4	53	3
75	6	70	5
78	7	110	9
82	8	--	--
n	5	4	
ΣR	26		19

**Hipótesis estadísticas:**

H<sub>0</sub>: Las medias de los tratamientos son estadísticamente iguales

H<sub>1</sub>: Las medias de los tratamientos son estadísticamente diferentes

**Regla de decisión:**

Si  $p$  de  $Z \geq \alpha$ : Aceptar H<sub>0</sub>

Si  $p$  de  $Z < \alpha$ : Aceptar H<sub>1</sub>

**Cálculos de U:**

Selección de la fórmula U: Como R1 > R2: Usar la fórmula U2

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 \quad \rightarrow U_2 = 5(4) + \frac{4(4 + 1)}{2} - 19 = 11$$

**Cálculo de la media MW:**

$$\mu_U = \frac{n_1 n_2}{2} = \frac{(5)(4)}{2} = 10$$

:

**Cálculo de la desviación estándar MW**

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \rightarrow \sigma_U = \sqrt{\frac{5(4)(5 + 4 + 1)}{12}} \rightarrow \sigma_U = \sqrt{\frac{200}{12}} = 4,083$$

**Cálculo de Z:**

$$Z = \frac{U - \mu_U}{\sigma_U} \rightarrow Z = \frac{11 - 10}{4,083} = 0,245$$

p de Z acumulada = 0,597 (valor obtenido en Excel)

**Interpretación:**

p de  $Z \geq \alpha = 0,05$ : Aceptar  $H_0$ . Las medias de los tratamientos comprados son estadísticamente iguales ( $p > 0,05$ ).

**Decisión:**

No hay diferencia significativa entre tratamientos.

**4.8.3 PRUEBA DE LA INDEPENDENCIA CON CHI CUADRADO**

La prueba  $\chi^2$  compara si las frecuencias observadas (O) son independientes de las frecuencias esperadas (E) en datos organizados en tablas de contingencia 2 x 2 (Cuadro 53).

*Cuadro 53. Modelo de tabla de contingencia 2 x 2 de muestras relacionadas*

Factor en estudio	Variable categórica		Total
	+	-	
Con tratamiento	A	B	A + B
Sin tratamiento	C	D	C + D
Total	A + C	B + D	N

Fuente: Badii et al. (2021, p. 11)

**Ejercicio 94:**

Determinar si la posición de las hojas en la planta, en los tercios superior, medio e inferior, tiene relación con el nivel de incidencia de una plaga insectil. Se tomó una muestra de 180 plantas y se clasificó las hojas infestadas, según la prevalencia de las posiciones.

**Datos y análisis:**

Los datos de hojas infestadas, clasificadas por la prevalencia de las posiciones en la planta, se exponen en el Cuadro 54. En el Cuadro 55, se indica el procedimiento para calcular el estadístico Chi-cuadrado.

*Cuadro 54. Posición de las hojas infestadas en las plantas y valores observados y esperados*

Frecuencias	Prevalencia de la posición de las hojas infestadas en las plantas			Total de plantas
	Tercio alto	Tercio medio	Tercio bajo	
Frecuencia observada	51	44	85	180
Frecuencia esperada	60	60	60	180

Cuadro 55. Cálculo de los valores Chi-cuadrado para posición de las hojas infestadas

Frecuencias y cálculos	Prevalencia de la posición de las hojas infestadas en las plantas			Total de plantas
	Tercio alto	Tercio medio	Tercio bajo	
Frecuencia observada	51	44	85	180
Frecuencia esperada	60	60	60	180
O - E	-9,0	-16,0	25,0	---
(O - E) <sup>2</sup>	81,0	256,0	625,0	---
$\chi^2$	1,4	4,3	10,4	16,03

$$\chi^2 = \frac{(O - E)^2}{E} \rightarrow \chi^2_{\text{calculada}} = 16,03$$

**Hipótesis estadística:**

H<sub>0</sub>: La posición de las hojas en la planta no influye sobre la incidencia (independencia)

H<sub>1</sub>: La posición de las hojas en la planta si influye sobre la incidencia (no independencia)

**Regla de decisión:**

Si  $\chi^2_{\text{calculada}} < \chi^2_{\text{valor crítico}}$ : Aceptar H<sub>0</sub>

Si  $\chi^2_{\text{calculada}} \geq \chi^2_{\text{valor crítico}}$ : Aceptar H<sub>1</sub>

**Grados de libertad:**

$$GL = (C - 1)(H - 1) = (3 - 1)(2 - 1) = 2$$

**Valores críticos:**

$$\chi^2_{0,05} = 5,991 \text{ y } \chi^2_{0,01} = 9,210$$

**Interpretación:**

$\chi^2_{\text{calculado}} = 16,03 > \chi^2_{0,05} = 5,991$ : Aceptar H<sub>1</sub> con el 95% de confianza.

$\chi^2_{\text{calculado}} = 16,03 > \chi^2_{0,01} = 9,210$ : Aceptar H<sub>1</sub> con el 99% de confianza.

La posición de las hojas en la planta incide en el comportamiento de la plaga (\*\*). El valor *p* de  $\chi^2 = 0,00033$  indica el 99,97 % de confianza en la decisión estadística.

En Excel, el valor *p* de  $\chi^2$  se obtiene con la función:

=PRUEBA.CHI(rango\_real;rango\_esperado).

**Decisión:**

La posición de las hojas en la planta si influye en la incidencia de la plaga (*p* < 0,01).

**Ejercicio 95:**

En una pista circular de carreras de caballos hay ocho carriles (C). Se dispone de información de 144 eventos con la especificación de los triunfos. Un grupo de aficionados desea conocer si la posición en las postas de carreras de caballos tiene influencia en los triunfos.

**Datos y análisis:**

La información histórica de 144 carreras y el número de triunfos por cada carril se expone en el Cuadro 56.

Cuadro 56. Triunfos observados en las carreras de caballos según carriles en la pista circular

	Posición de los caballos ganadores en la pista circular de carreras de caballos								Total de carreras
	C1	C2	C3	C4	C5	C6	C7	C8	
Nº triunfos observados	29	19	18	25	17	10	15	11	144

**Análisis de los valores esperados por carril:**

Los triunfos esperados (TE) en cada uno de los 8 carriles (C), se determina mediante la siguiente relación:

$$T_E = \frac{1}{C}(n) \qquad T_E = \frac{1}{C}(n) = \frac{1}{8}(144) = 18 \text{ triunfos por carril}$$

**Hipótesis estadística:**

H<sub>0</sub> = La posición en la pista no influye en los triunfos. Observados = Esperados

H<sub>1</sub> = La posición en la pista si influye en los triunfos. Observados ≠ Esperados

**Regla de decisión:**

Si  $\chi^2_{\text{calculada}} < \chi^2_{\text{valor crítico}}$ : Aceptar H<sub>0</sub>

Si  $\chi^2_{\text{calculada}} \geq \chi^2_{\text{valor crítico}}$ : Aceptar H<sub>1</sub>

En el Cuadro 57, se indican los resultados de 144 carreras y los triunfos por carril.

Cuadro 57. Triunfos observados y esperados de 144 carreras de caballos y cálculo de Chi cuadrada

Estadísticos	Posición de los caballos ganadores en la pista circular de carreras de caballos								Total
	C1	C2	C3	C4	C5	C6	C7	C8	
Triunfos observados por carril (O)	29	19	18	25	17	10	15	11	144
Triunfos esperados por carril (E)	18	18	18	18	18	18	18	18	144
Diferencia: O - E	11	1	0	7	-1	-8	-3	-7	0
$\chi^2$	6,72	0,06	0,00	2,72	0,06	3,56	0,50	2,72	<b>16,33</b>

Fuente: Siegel (1990, p. 67)

$$\chi^2 = \frac{\Sigma(O - E)^2}{E} \quad \rightarrow \quad \chi^2_{\text{calculado}} = 16,33$$

**Valores críticos:**

$$\chi^2_{0,05} = 14,07 \text{ y } \chi^2_{0,01} = 18,48$$

$$GL = (C - 1)(H - 1) = 7$$

**Interpretación:**

$\chi^2_{\text{calculada}} = 16,33 > \chi^2_{0,05} = 14,07$ : Aceptar H<sub>1</sub> con el 95 % de confianza. La posición en la pista si influye en los triunfos.

$\chi^2_{\text{calculada}} = 16,33 < \chi^2_{0,01} = 18,48$ : Aceptar H<sub>0</sub> con el 99 % de confianza. La posición en la pista no influye en los triunfos.

En la actualidad, los estadísticos recomiendan tomar decisiones en base al valor *p*, obtenido en los distintos programas computacionales. En este caso, *p* de  $\chi^2 = 0,022 \sim 2,2 \%$ .

En Excel, el valor  $p$  de  $\chi^2$  se obtiene con la función:

=PRUEBA.CHI(rango\_real;rango\_esperado)

**Decisión:**

Con el 97,8 % de confianza se afirma que los triunfos en las carreras de caballos si dependen de la posición en la pista.

#### 4.9 INTRODUCCIÓN AL ANÁLISIS MULTIVARIADO

Los métodos del análisis multivariado (AM) permiten explorar, describir y relacionar series de datos de varias variables sobre un mismo objeto de estudio (Balzarini et al., 2015, p. 8), posibilitando hacer aproximaciones predictivas. Esto significa que las técnicas multivariadas pueden adaptarse a los objetivos de estudio de otros niveles investigativos, aunque ciertamente tiene énfasis en la descripción.

El carácter multivariante se fundamenta en el número de variables analizadas y en las combinaciones que pueden existir entre ellas. Un análisis multivariado se inicia con la organización de la información de las variables, lo más homogéneo que sea posible y la formación de grupos, los más distintos que sean posibles. El análisis multivariante puede clasificarse de distintas maneras en función de la relación entre variables cuantitativas y cualitativas (Figura 26).

Para el tratamiento de variables cuantitativas se usa el análisis de componentes principales (ACP) y el análisis factorial (AF). En el tratamiento de variables cualitativas se usa el análisis de correspondencias (AC). Para relacionar casos se usa el análisis de conglomerados jerárquicos o clúster (ACJ) y en los análisis de objetos se emplea el escalamiento multidimensional (EM).

El estudio multivariado debe planearse con antelación definiendo objetivos con exactitud. Por ejemplo, en un diagnóstico agrosocial, se requiere información de los agricultores de una comunidad o de varias comunidades, sobre el tamaño de las fincas, la productividad, los ingresos, los costos de producción, el grado de escolaridad, los ingresos mensuales, el nivel de tecnología que usa y el grado de satisfacción con la vida.

En una zona ganadera puede estudiarse el tamaño de las fincas, el área dedicada a la ganadería, la carga animal y la productividad diaria de leche/vaca, entre otras variables. Los animales de un hato ganadero pueden clasificarse por sus características productivas y reproductivas, edad, rusticidad, prolificidad, fertilidad, raza y condiciones de salud.

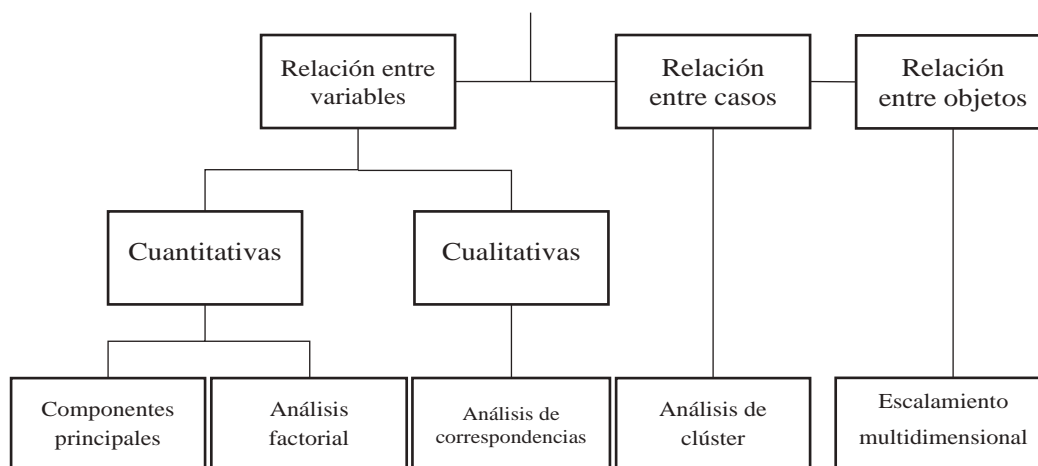


Figura 26. Técnicas de análisis multivariados de mayor uso en la investigación agropecuaria

#### 4.9.1 ANÁLISIS DE COMPONENTES PRINCIPALES

El Análisis de Componentes Principales (ACP) es un método estadístico que permite transformar el conjunto inicial de datos en un conjunto reducido de nuevas variables llamadas componentes principales que resultan de las combinaciones lineales de las variables que expliquen la mayor parte de la variación total. El primer factor o componente es aquel que explica una mayor parte de la varianza total, el segundo factor explica la otra parte de la varianza restante, y así sucesivamente, hasta explicar el 100 % de la variación.

El ACP es un análisis de ordenamiento que utiliza las estructuras de los autovectores (raíces latentes o eigens) de la matriz de correlación o de una matriz de varianza-covarianza calculadas a partir de las variables originales (Yengle, 2012, p. 146). Los vectores propios, son los vectores “no nulos” que se transforman por un operador para dar origen a un múltiplo escalar de sí mismo, que recibe el nombre de valor propio, autovalor, valor característico o eigenvalor. A menudo, una transformación queda determinada por sus vectores propios y valores propios. Un espacio propio, eigenespacio se asocia al valor propio lambda ( $\lambda$ ) como conjunto de vectores. Un vector propio (AX) se refiere al dato original transformado en valor propio ( $\lambda X$ ) que representa una magnitud, una asociación y una dirección.

El ACP tiene aplicaciones en la determinación de la similitud morfológica entre especies y variedades vegetales, entre grupos raciales de animales, entre sistemas forestales o comunidades rurales. En las poblaciones vegetales se analizan variables como: altura de planta, diámetro del tallo, número de hojas, número de ramas, longitud de rama, distancia entrenudos, número de frutos, peso de frutos, longitud del pistilo, ancho de la corola u otras. En la producción ganadera se analizan variables como: carga animal, edad de los animales, producción diaria de leche, rendimiento quesero, entre otras.

#### Ejercicio 96:

Realizar el ACP de los parámetros productivos de 12 fincas ganaderas

#### Datos y análisis:

Los datos de las variables sobre la producción lechera se exponen en el Cuadro 58

Cuadro 58. Variables productivas en 12 fincas ganaderas

Fincas	Área de la finca (ha)	Superficie ganadera (%)	Número de animales	Tasa de reposición	Tasa de mortalidad	Carga animal	Índice de terneros comerciales	Índice de litros de leche comerciales
1	20	60	15	0,5	0,2	1,25	0,120	12,5
2	30	70	25	0,4	0,1	1,19	0,090	13,1
3	35	65	21	0,6	0,3	0,92	0,300	11,1
4	33	70	30	0,6	0,2	1,3	0,200	19,5
5	25	80	35	0,3	0,1	1,75	0,900	24,5
6	28	90	28	0,5	0,1	1,11	0,100	14,4
7	30	60	29	0,5	0,2	1,61	0,200	20,9
8	34	70	30	0,9	0,3	1,26	0,270	12,6
9	35	80	40	0,2	0,1	1,43	0,088	25,7
10	36	90	50	0,3	0,15	1,54	0,150	15,4
11	40	90	60	0,4	0,1	1,67	0,200	16,7
12	50	70	65	0,2	0,05	1,86	0,500	18,6

Los componentes principales, usando el programa INFOSTAT, representados en una gráfica biplot (dos dimensiones), explican el 67,3 % de la variación total, al CP1 correspondió el 50,3 % y al CP2 el 17,0 %. Esta información se deriva de la interpretación de los autovalores. En los autovectores se aprecia la importancia de los parámetros individuales dentro de los componentes principales.

Las tasas de mortalidad y de reposición son parámetros productivos asociados. El número de animales, la superficie dedicada a la ganadería y el área total de las fincas conforman otro conjunto distinto de parámetros relacionados. La carga animal, el índice de litros de leche comercial y el índice de terneros conforman otro grupo de parámetros. Se observa también que el grado de similitud entre las fincas sometidas al ACP, destacándose que la finca 12 es muy diferente del resto, principalmente de las fincas 3, 8 y 1.

El ACP calculado<sup>15</sup> con el uso del programa INFOSTAT, en primer lugar arroja los valores propios y proporciones de la contribución a la variación total de las variables productivas asociadas a la ganadería y magnitud de los autovectores en los componentes principales 1 y 2, expuesto en el Cuadro 59. La magnitud de los autovectores en los componentes principales 1 y 2, se indican en el Cuadro 60.

En el CP1 tienen más peso las variables: número de animales, tasa de reposición, tasa de mortalidad y carga animal. En el Gráfico 25, se observa el diagrama ACP de los parámetros productivos de las 12 fincas ganaderas. La tasa de mortalidad y tasa de reposición tienen fuerte relación. En otro grupo se encuentra el índice de terneros comerciales, índice de litros de leche comercial y carga animal. En un tercer grupo, el número de animales y las superficies total de las fincas y las dedicadas a la ganadería.

*Cuadro 59. Valores propios y proporciones de la contribución a la variación total de las variables productivas asociadas a la ganadería y magnitud de los autovectores en los componentes principales 1 y 2*

Valores propios y porcentaje explicado de la variación total				
Lambda	Valor propio	Proporción	Proporción acumulada	Porcentaje explicado
<b>1</b>	4,03	<b>0,50</b>	0,50	<b>50</b>
<b>2</b>	1,36	<b>0,17</b>	0,67	<b>67</b>
3	1,12	0,14	0,81	81
4	0,66	0,08	0,90	90
5	0,43	0,05	0,95	95
6	0,26	0,03	0,98	98
7	0,13	0,02	1,0	100
8	0,0039	0,00048	1,0	100

<sup>15</sup> Cálculo usando el programa INFOSTAT versión 2020

Cuadro 60. Magnitud de los autovectores en los componentes principales 1 y 2

Variables de producción lechera	e1	e2
Área total de la finca	0,28	<b>-0,46</b>
Superficie dedicada a la ganadería	0,26	<b>-0,37</b>
Número de animales	<b>0,43</b>	-0,32
Tasa de reposición	<b>-0,41</b>	-0,07
Tasa de mortalidad	<b>-0,41</b>	0,03
Carga animal	<b>0,43</b>	0,21
Índice de terneros comerciales	0,20	<b>0,54</b>
Índice de litros de leche comerciales	0,33	<b>0,45</b>

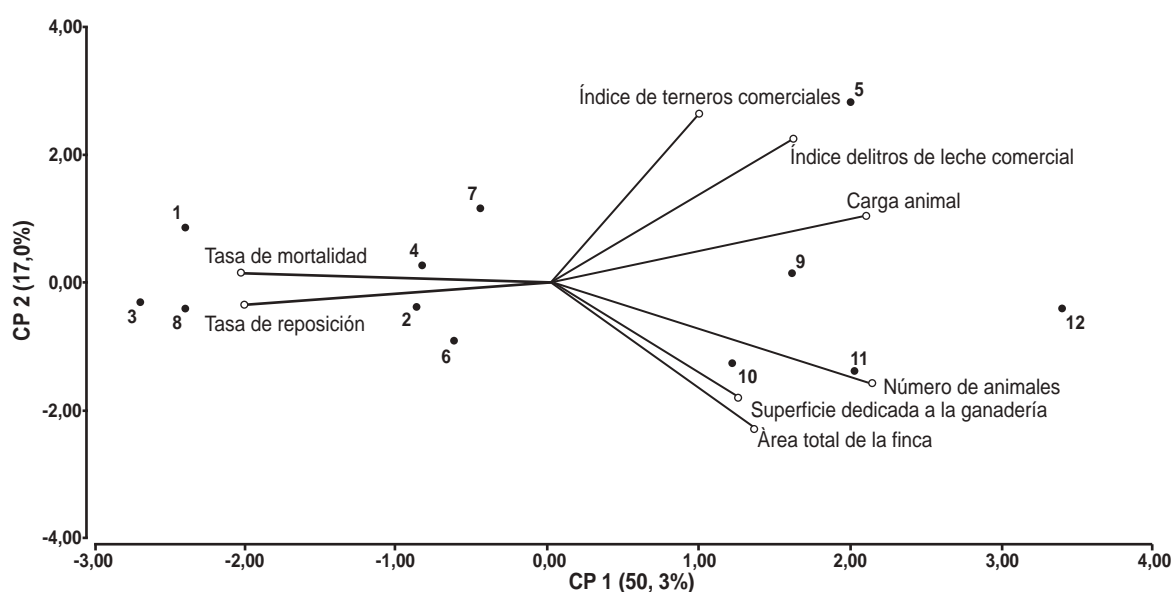


Gráfico 25. Análisis de componentes principales de los parámetros productivos en 12 fincas ganaderas

**Decisión:**

Las variables productivas asociadas a la ganadería bovina forman tres grupos: El primero lo integran las tasas de mortalidad y de reposición; el segundo, el número de animales, la superficie ganadera y el área total de la finca y el tercer grupo está conformado por carga animal, índice de litros de leche comercial y el índice de terneros.

4.9.2 ANÁLISIS FACTORIAL

El objetivo del análisis factorial (AF) es analizar si hay factores que puedan explicar el comportamiento de las variables. Los factores latentes explican los comportamientos visibles en las variables fenotípicas. El AF y el análisis de componentes principales (ACP) están relacionados, pero existen diferencias. El ACP busca hallar las combinaciones lineales de las variables originales para explicar la mayor parte posible de la varianza total, mientras que el AF reduce el número de variables respecto del total original, que expliquen la mayor proporción de la varianza total (Yengle, 2012, p. 147).

El AF supone que hay “factores comunes” subyacentes a todas las variables, lo cual no lo presupone el ACP. El AF se realiza en dos etapas: a) La primera etapa corresponde a



**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

un análisis exploratorio donde se incluyen todas las variables registradas; y, b) posteriormente, se hace un AF confirmatorio, reprocesando los datos solo de las variables que más contribuyen a la varianza total, identificadas en la etapa exploratoria.

En el análisis factorial<sup>16</sup> (AF), se obtiene, en primer lugar, la matriz de correlaciones lineales (Cuadro 61). El área total de la finca se correlaciona positivamente con número de animales ( $r = 0,819^{**}$ ). El número de animales se correlaciona positivamente con la carga animal ( $r = 0,763^{**}$ ) y negativamente con la tasa de mortalidad ( $r = -0,584^*$ ). La tasa de reposición y la tasa de mortalidad se correlacionan positivamente ( $r = 0,840^{**}$ ), así como, la carga animal y el índice de litros de leche comerciales ( $r = 0,639^*$ ).

El análisis de las comunalidades (Cuadro 62), a partir de la extracción, valora como las variables de mayor importancia: número de animales/finca, área total del predio y tasa de mortalidad. En el análisis confirmatorio (Gráfico 26), se determinó que las variables de mayor importancia fueron: número de animales, área total de la finca, tasa de mortalidad, carga animal y tasa de reposición, excluyéndose las variables SG, ITer e ILec. El ACP con las variables selectas amplía la varianza explicada a 83,3 % (60,8 en CP1 y 22,5 en CP2).

*Cuadro 61. Coeficientes de correlaciones entre variables productivas en fincas ganaderas*

Variables	ATF	SG	NA	TR	TM	CA	ITer	ILec
Área total de la finca (ATF)	1							
Superficie dedicada a la ganadería (SG)	0,204	1						
Número de animales (NA)	0,819**	0,531	1					
Tasa de reposición (TR)	-0,273	-0,360	-0,553	1				
Tasa de mortalidad (TM)	-0,248	-0,516	-0,584*	0,840**	1			
Carga animal (CA)	0,411	0,218	0,763**	-0,586*	-0,584*	1		
Índice de terneros comerciales (ITer)	0,065	-0,004	0,225	-0,201	-0,159	0,516	1	
Índice de litros de leche comerciales (ILec)	0,082	0,174	0,338	-0,564	-0,480	0,639*	0,400	1

Nota: Valores críticos de  $r_{0,05} = 0,576$  y de  $r_{0,01} = 0,708$  con  $GL = 10$

*Cuadro 62. Análisis de las comunalidades entre variables productivas en las fincas ganaderas*

Comunalidad	Inicial	Extracción
Número de animales (NA)	1,00	0,984
Área total de la finca (ATF)	1,00	0,919
Tasa de mortalidad (TM)	1,00	0,862
Carga animal (CA)	1,00	0,860
Tasa de reposición (TR)	1,00	0,796
Índice de leche comercial (ILec)	1,00	0,729
Índice de terneros comercial (ITer)	1,00	0,721
Superficie dedicada a la ganadería (SG)	1,00	0,636

Nota: SG, ITer e ILec se excluyen del AF confirmatorio por su baja contribución a la varianza

<sup>16</sup> Análisis AF calculado en INFOSTAT versión 2020.

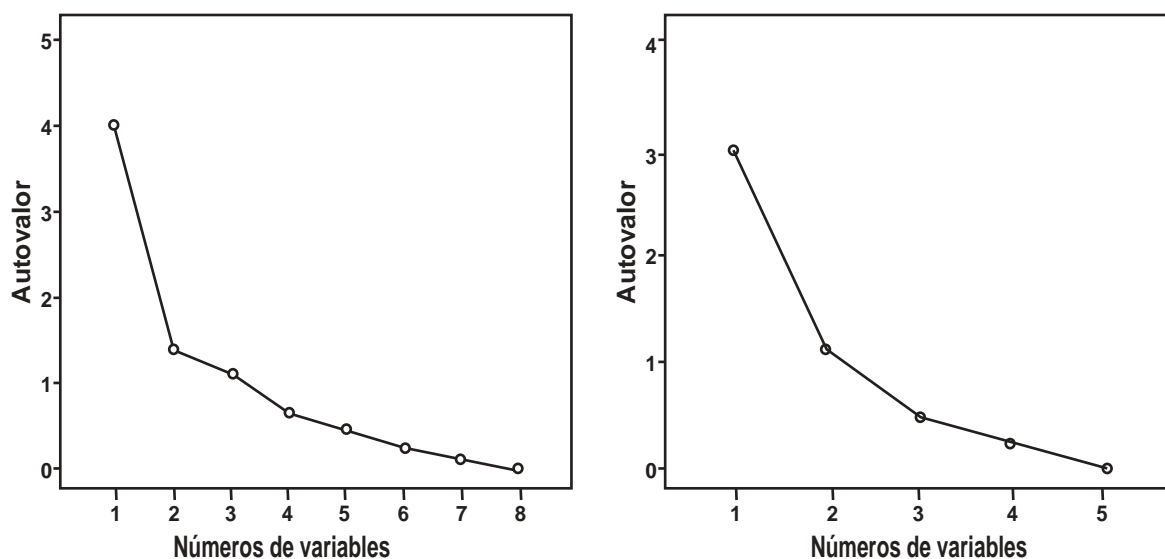


Gráfico 26. Reducción de variables en un estudio ganadero mediante análisis factorial

En el AF exploratorio se valoran las ocho variables, luego se descartan tres de ellas por su baja contribución a la variación total, reprocesando el AF con cinco variables (análisis confirmatorio). Estas mismas cinco variables identificadas como de mayor importancia se usan para el análisis de componentes principales (ACP). En este nuevo escenario, los componentes 1 y 2 explican el 83,3 % de la variabilidad total (Cuadro 63).

Cuadro 63. Análisis de componentes principales en el AF confirmatorio

Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado		
	Total	Varianza (%)	Varianza acumulada (%)	Total	Varianza (%)	Varianza acumulada (%)
1	3,040	60,798	60,798	3,040	60,798	60,798
2	1,126	22,522	<b>83,320</b>	1,126	22,522	83,320
3	0,512	10,236	93,557			
4	0,276	5,523	99,080			
5	0,046	0,920	100,000			

Para tener confiabilidad en una matriz de datos, ésta debe someterse a la prueba KMO (Kaiser-Meyer-Olkin) y a la prueba de esfericidad de Bartlett. La KMO es una medida estadística usada para probar qué tan adecuados son los datos para el análisis factorial. En los análisis, exploratorio y confirmatorio, del ejemplo, los Coeficientes KMO variaron de 0,37 a 0,525 que se considera apropiado, pues mientras más cercano a 1 es mejor. La prueba de esfericidad de Bartlett contrasta si la matriz de correlaciones es una matriz la identidad, por lo tanto, es un indicativo de que el modelo factorial es adecuado. El estadístico de Bartlett se interpreta como un valor  $p$ , pues, si  $p < 0,05$  hay significación estadística. En el ejemplo, en las pruebas exploratoria y confirmatoria, el valor  $p < 0,01$ .

#### Decisión:

El número de animales, el área total del predio, la tasa de mortalidad, la carga animal y la tasa de reposición explican 83,3 % de la variación total. La significación de Bartlett  $p < 0,001$  y el Coeficiente KMO = 0,525 indican que hay confiabilidad en las matrices de datos.

#### 4.9.3 ANÁLISIS DE CORRESPONDENCIAS

El análisis de correspondencias (AC) es una técnica descriptiva para datos cualitativos (Parra, 1996, p.183). En un AC hay dos enfoques: AC simples y AC múltiples. En el simple se utiliza las tablas de contingencia de dos variables nominales u ordinales y en el múltiple se tratan varias variables en tablas de contingencia multidimensionales.

El AC tiene como objetivos:

- Medir la proximidad entre individuos en términos de parecido: Dos individuos se parecen si un conjunto de atributos (modalidades) son semejantes.
- Medir la proximidad entre modalidades de una variable en términos de parecido: resultan excluyentes por construcción o son semejantes porque presentan el mismo comportamiento.

La suma de las distancias  $\chi^2$ , desde cada punto hasta el centro de gravedad, es la inercia. Los autovalores corresponden a la raíz cuadrada de la inercia. La obtención de los vectores propios (factores) y autovalores (valores propios) corresponden a la varianza explicada. Las medidas de distancia entre las filas y las columnas, se realiza mediante: distancia Euclídea y la distancia Chi-cuadrado. La distancia euclídea es la raíz cuadrada de la suma cuadrática de las diferencias entre pares de filas y de columnas. Algunos autores no aconsejan el uso de esta diferencia porque consideran a todos los elementos con el mismo peso, pudiendo desvirtuar el análisis de correspondencias (De la Fuente, 2011, p. 26).

La distancia  $\chi^2$  cumple el principio de la equivalencia distribucional, postulando que si dos categorías tienen perfiles idénticos pueden ser sustituidas por una que sea la suma de sus pesos, sin modificar la distancia entre filas o columnas. Pues, mientras las distancias se encuentran lo más alejadas del origen, las variables estarán mejor representadas.

#### Ejercicio 97:

Realizar el análisis de correspondencias (AC) a partir de los datos promedios de las variables productivas registradas en las 12 fincas ganaderas, de un territorio específico.

#### Datos y análisis:

Esta es una técnica de análisis cualitativo. Las variables cuantitativas pueden transformarse a cualitativas en base a criterios propios del productor o del investigador. En este caso, la escala ordinal tiene una jerarquía de tres puntos: baja, media y alta (Cuadro 64).

Cuadro 64. Variables productivas de 12 fincas ganaderas

Fincas	Área total de la finca	Superficie dedicada a la ganadería	Número de animales	Carga animal
1	Baja	Baja	Bajo	Baja
2	Media	Baja	Bajo	Baja
3	Alta	Baja	Bajo	Muy baja
4	Alta	Baja	Bajo	Baja
5	Media	Media	Medio	Muy alta
6	Media	Alta	Bajo	Baja
7	Media	Muy baja	Bajo	Alta
8	Alta	Baja	Bajo	Baja
9	Alta	Media	Medio	Media
10	Alta	Alta	Alta	Alta
11	Alta	Alta	Muy alta	Alta
12	Muy alta	Baja	Muy alta	Muy alta

Los datos cuantitativos se transforman a datos escalares, como en el presente caso, a una escala ordinal de 1 a 5, donde: 1 = muy bajo y 5 = muy alto. Los análisis cuantitativos deben complementarse con los cualitativos, pues, un enfoque mixto posibilita una mejor comprensión de las relaciones entre las variables.

El análisis de correspondencias se realizó usando el programa INFOSTAT. La contribución a Chi cuadrado, usando las variables: área total de la finca, carga animal y número de animales, explica 42,88 % de la variación total, desagregado en 23,75 % para el eje 1 y 19,13 % para el eje 2.

Autovalor	Inercias	Chi-cuadrado	(%)	Acumulado (%)
1	0,88	0,78	43,855	23,3
2	0,78	0,62	34,71	18,5

Las correspondencias entre carga animal, número de animales y área total de la finca son altas. El área total del predio, en su nivel medio, tiene correspondencia con los niveles bajos y muy bajos de carga animal, y con el bajo número de animales. El área del predio tiene correspondencia con la alta carga animal y el número de animales (Gráfico 27).

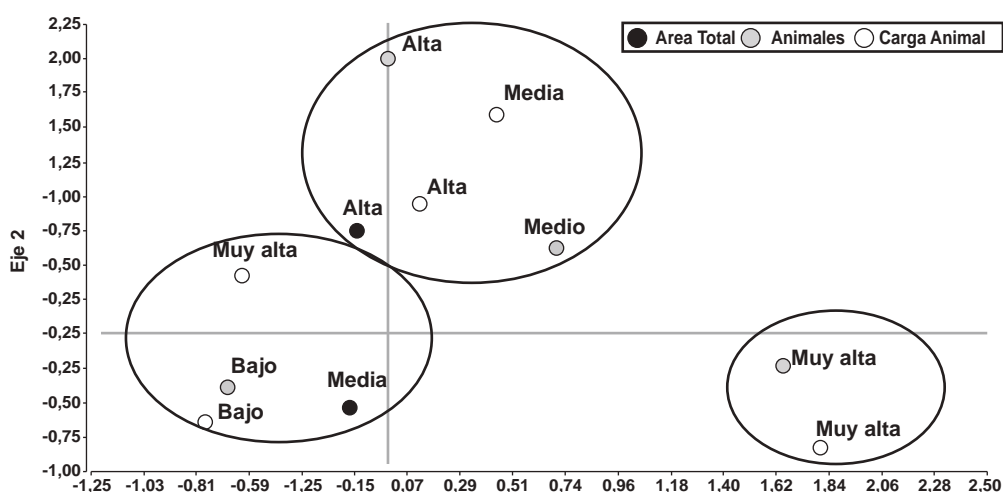


Gráfico 27. Análisis de correspondencias de las variables asociadas a la producción ganadera

**Decisión:**

Las variables productivas en fincas dedicadas a la producción lechera, tienen niveles de correspondencia muy altos, en las variables carga animal y número de animales. Esto denota la existencia de fincas ganaderas de alta eficiencia.

**Ejercicio 98:**

Determinar las correspondencias entre las variables de calidad en el faenamiento de peces de la especie “vieja azul” sometidos a cuatro dietas.

**Datos:**

Los análisis de frecuencias de las variables referidas al faenamiento de peces en función de las dietas<sup>17</sup>, se indican en el Cuadro 65.

<sup>17</sup> Data compartida por el Dr. Carlos Mazón Paredes (+), docente de la UTEQ.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

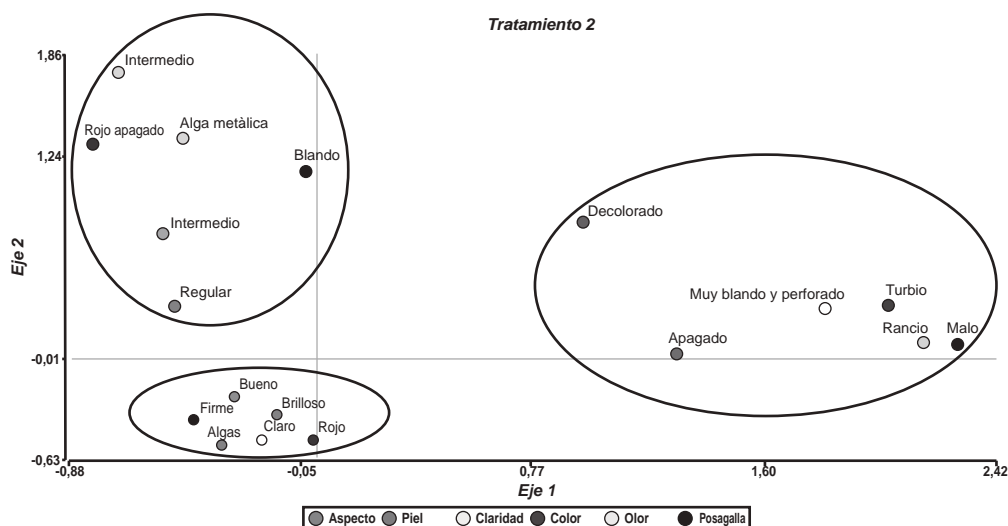
*Cuadro 65. Frecuencias absolutas (FA) de los atributos de calidad en faenamiento de peces de la especie “vieja azul” en función a cuatro dietas*

Variables / frecuencia	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	Σ
Aspecto FA	29	Bueno 27	28	28	12	Regular 13	12	12	7	Malo 8	8	8	192
Piel FA	30	Brillante 28	30	28	11	Intermedia 12	10	12	7	Apagada 8	8	8	192
Claridad FA	30	Claro 32	32	32	10	Intermedio 8	8	8	8	Turbio 8	8	8	192
Color FA	32	Rojo 32	36	33	9	Rojo apagado 8	8	7	7	Decolorado 8	4	8	192
Olor FA	28	Olor a algas 29	28	28	12	Olor a alga metálica 11	12	12	8	Olor rancio 8	8	8	192
Posagalla FA	31	Firme 31	32	32	10	Blando 10	8	8	7	Muy blando y perforado 7	8	8	192

**Contribución a Chi cuadrado:**

Tratamiento 1					
Ejes	Autovalor	Inercias	Chi cuadrado	%	% acumulado
1	0,70	0,49	231,47	24,71	24,71
2	0,58	0,34	157,91	16,86	41,47
Tratamiento 2					
Ejes	Autovalor	Inercias	Chi cuadrado	%	% acumulado
1	0,77	0,59	682,99	29,50	29,50
2	0,65	0,42	491,50	21,23	<b>50,73</b>
Tratamiento 3					
Ejes	Autovalor	Inercias	Chi cuadrado	%	% acumulado
1	0,66	0,43	636,79	21,53	21,53
2	0,65	0,42	623,20	21,07	42,60
Tratamiento 4					
Ejes	Autovalor	Inercias	Chi cuadrado	%	% acumulado
1	0,76	0,58	1432,54	29,13	29,13
2	0,68	0,46	1126,41	22,91	<b>52,04</b>

Los tratamientos 2 y 4 explican la variación total en mayor proporción, calculada con Chi cuadrado. Las variables asociadas a la calidad tuvieron un similar comportamiento con todas las dietas. En el Gráfico 28, se expone el mejor tratamiento (T2), evidenciándose las correspondencias entre las variables: buen aspecto, piel brillante, color rojo, claridad de la pulpa, olor a algas y posagalla, que tienen una contribución acumulada de 50,73 %.



Nota: El tratamiento T2 se desagrega con fines didácticos.

Gráfico 28. Análisis de correspondencias de los atributos de la calidad de peces de la especie “vieja azul”

**Decisión:**

En el tratamiento T2 se expresa de forma nítida la correspondencia entre las variables de calidad, en peces de la “especie vieja azul”.

**Ejercicio 99:**

Realizar el análisis de correspondencias entre los contenidos de boro (B), azufre (S), zinc (Zn), potasio (K<sub>2</sub>O) y fósforo (P<sub>2</sub>O<sub>5</sub>), niveles de pH y materia orgánica (MO) en 150 fincas cafetaleras de la amazonia norte del Ecuador.

**Datos y análisis:**

Los resultados de los análisis químicos del suelo se exponen en el Cuadro 66.

Cuadro 66. Promedios de los contenidos de nutrientes, materia orgánica y pH del suelo de 150 fincas cafetaleras de la amazonia norte

Contenidos de Nutrientes	N	P <sub>2</sub> O <sub>5</sub>	K <sub>2</sub> O	S	Ca	Mg	Zn	Cu	Fe	Mn	B	MO	pH
Bajo	148	101	7	127	18	40	22	3	0	1	94	18	76
Medio	2	36	90	22	84	89	112	62	0	22	43	72	73
Alto	0	13	53	1	48	21	16	85	150	127	13	60	1

La contribución de la Chi cuadrado para explicar la variación es 24,88 % del total (Cuadro 67). Con la información de las correspondencias entre los principales nutrientes del suelo, contenidos de materia orgánica y niveles de pH, se consideró pertinente analizar todas las variables. El análisis con las variables de mayor peso mejoró la contribución, aunque el objetivo solo fue detectar las correspondencias. Los pesos de las variables del suelo, ponderadas para las coordenadas, en los ejes 1 y 2, se indica en el Cuadro 68.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 67. Contribución a la variación explicada mediante Chi cuadrado*

Ejes	Autovalor	Inercias (autovalor <sup>2</sup> )	Chi-cuadrado $\frac{(O - E)^2}{E}$	%	Acumulado (%)
1	0,54	0,29	817,64	15,39	15,39
2	0,42	0,18	504,57	9,49	24,88

*Cuadro 68. Valores de peso de las variables en las coordenadas*

VARIABLES	Niveles	Eje 1	Eje 2
pH	Adecuado	<b>0,81</b>	0,20
	Muy ácido	<b>-0,59</b>	-0,13
	Alcalino	<b>-1,65</b>	<b>-2,34</b>
Boro	Alto	0,20	<b>-0,98</b>
	Bajo	0,13	0,14
	Medio	-0,31	0,05
Zinc	Medio	-0,18	-0,33
	Alto	<b>1,78</b>	0,29
	Bajo	-0,60	<b>2,18</b>
Azufre	Bajo	0,12	0,17
	Medio	<b>-0,65</b>	<b>-0,90</b>
	Alto	<b>-1,43</b>	<b>-0,82</b>
Potasio	Alto	<b>1,09</b>	-0,15
	Medio	-0,60	-0,10
	Bajo	-0,97	<b>2,84</b>
Fósforo	Medio	<b>1,23</b>	0,05
	Bajo	-0,52	0,12
	Alto	<b>0,64</b>	<b>-0,91</b>
Nitrógeno	Bajo	0,0037	0,03
	Medio	-0,21	<b>-1,80</b>
Materia orgánica	MO-Alto	-0,12	-0,08
	MO-Medio	0,09	0,13
	MO-Bajo	0,04	-0,34

***Interpretación:***

Los suelos dedicados al café en la amazonia norte del Ecuador son altamente diferenciados en pH. El suelo alcalino encontrado en una sola finca es un caso excepcional. El pH adecuado se encuentra a una distancia de 0,89 a la derecha y a 0,20 arriba del cuadrante. Los contenidos de materia orgánica están próximos al cero, vertical y horizontal, compartiendo el espacio con el nitrógeno bajo, fósforo bajo, azufre bajo y boro bajo, todos estos próximos al pH muy ácido. El adecuado pH del suelo tiene correspondencia con los niveles altos de zinc y de potasio, y con los medios-altos de fósforo. Los contenidos de materia orgánica medios, bajos y altos no inciden directamente en los niveles de fertilidad.

Los niveles bajos de azufre, boro y nitrógeno se asocian al nivel medio de zinc. El pH muy ácido tiene fuerte correspondencia con los niveles bajo de fósforo y medio de potasio. Los contenidos de materia orgánica, alta, media y baja se encuentran próximos al centro (Gráfico 29).

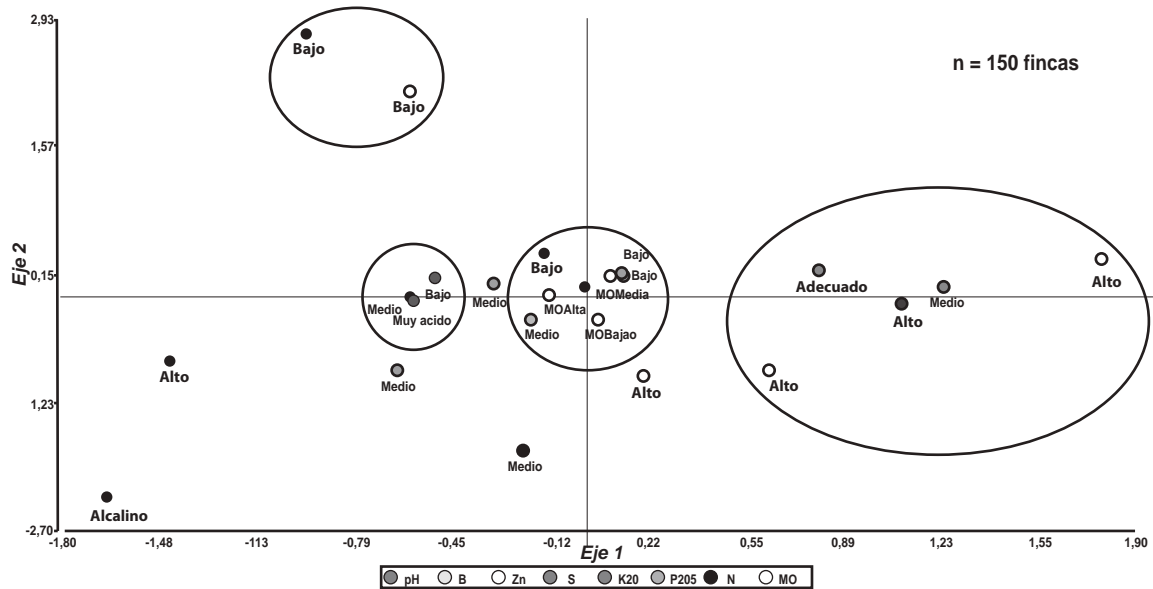


Gráfico 29. Análisis de correspondencias entre pH, materia orgánica y nutrientes en suelos amazónicos

**Decisión:**

Los contenidos medios, bajos y altos de materia orgánica no son determinantes de la fertilidad de los suelos. El pH muy ácido tiene una correspondencia directa con los niveles bajos de fósforo y medios de potasio.

4.9.4 ESCALAMIENTO MULTIDIMENSIONAL

El escalamiento multidimensional (EM) es un procedimiento no paramétrico para determinar preferencias o percepciones de los encuestados y representarlos en un diagrama visual, llamado mapa perceptual. El EM es una técnica de análisis cualitativo que tiene el propósito de describir los datos mediante la reducción de las dimensiones y su representación gráfica procurando encontrar relaciones no lineales. Este análisis solo es factible usando rangos ordenados o mediciones en escala ordinal.

A diferencia de las otras técnicas multivariadas, como análisis factorial (AF) y análisis de componentes principales (ACP), donde se dimensionan los atributos, en el Escalamiento multidimensional (EM) se identifican las distancias, interpretándose que mientras más reducidas sean habría mayor similitud. La distancia puede interpretarse geoméricamente mediante varias técnicas como la distancia euclidiana considerada una medida de similitud o disimilitud.

La distancia euclidiana es una medida de la distancia entre dos puntos en un espacio euclidiano. La distancia euclidiana se calcula con la fórmula siguiente:

$$D = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \quad [91]$$

Dónde:

D = Distancia euclidiana

X<sub>2</sub> - X<sub>1</sub> = Diferencia entre dos puntos sucesivos de una variable

Y<sub>2</sub> - Y<sub>1</sub> = Diferencia entre dos puntos sucesivos de la variable comparada



**Ejercicio 100:**

Determinar las similitudes de los efectos de 5 plaguicidas en base a la información de la mortalidad (%) de insectos plaga, medidos 15 días después de la aplicación de los productos.

**Datos y análisis:**

En el Cuadro 69, se indica la mortalidad, la eficacia Abbott y los rangos de eficacia en función de la aplicación de agroquímicos.

**Cálculo de la mortalidad y eficacia Abbott:**

La mortalidad se calcula mediante la fórmula:

$$M(\%) = \frac{IM}{IV + IM} (100) \quad [92]$$

Dónde:

- M (%) = Porcentaje de mortalidad
- IM = Número de individuos muertos
- IV = Número de individuos vivos
- IM + IV = Total de individuos en la muestra

La eficacia Abbott modificada, referida por Andujar et al. (1997, p. 32) y el ICA (2016, p.18), se calcula con la fórmula:

$$E_{\text{Abbot}} = \frac{M_T - M_t}{100 - M_t} (100) \quad [93]$$

Dónde:

- E<sub>Abbot</sub> = Eficacia Abbot modificada, en porcentaje
- M<sub>T</sub> = Porcentaje de mortalidad en el tratamiento
- M<sub>t</sub> = Porcentaje de mortalidad en el testigo

La asignación de rangos de eficacia, basado en la mortalidad ocasionada por los productos en estudio, consiste en identificar con letras del alfabeto según los valores de eficacia, en orden, correspondiendo al producto de mayor eficacia la letra A y así sucesivamente hasta la letra F.

*Cuadro 69. Mortalidad, eficacia y rangos de eficacia en función del uso de agroquímicos*

Agroquímico	Mortalidad (%)	Eficacia Abbot	Rango de eficacia
P1	90	88	A
P2	80	76	B
P3	68	62	C
P5	55	47	D
P4	45	35	E
Testigo	15	0	F

Las diferencias de la eficacia, organizados por rangos, de menos a más, de modo escalonado, proporciona información sobre la similitud entre productos (Cuadro 70). Los rangos de eficacia ABCDE se asocian a los productos P1, P2, P3 y P4. El testigo se codifica con la letra F.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 70. Distancias entre los niveles de eficacia de los productos*

	Rangos de eficacia					
	F	E	D	C	B	A
<i>Eficacia</i>	0	35	47	62	76	88
F (0)	0					
E (35)	35	0				
D (47)	47	12	0			
C (62)	<b>62</b>	27	15	0		
B (76)	<b>76</b>	41	29	14	0	
A (88)	<b>88</b>	<b>53</b>	41	26	12	0

Nota: En negrilla se indican las mayores distancias en eficacia entre agroquímicos

Las “matrices de distancias y similaridad” (INFOSTAT, 2020), considera dos unidades euclidianas como valor crítico. Se determina que la eficacia del producto A tiende a ser similar a la del B y disímil del producto E. La eficiencia del producto B tiende a ser similar al C y disímil o alejado del E (Cuadro 71).

*Cuadro 71. Escalamiento multidimensional de la eficacia de agroquímicos usando la distancia euclídea*

Productos	A	B	C	D	E	F
A	0					
B	0,53	0				
C	1,16	0,63	0			
D	1,84	1,31	0,68	0		
E	<b>2,37</b>	1,84	1,21	0,53	0	
F	<b>3,94</b>	<b>3,41</b>	<b>2,78</b>	2,10	1,57	0

**Decisión:**

El producto P1 (A) tuvo una eficacia del 88 % que resultó próxima al producto P2 (B) que tuvo el 76 % de eficacia.

**4.9.5 ANÁLISIS DE CLÚSTER**

El análisis de conglomerados o de clúster es una técnica multivariada que clasifica los grupos de objetos o individuos por su similitud, en base a mediciones distancias y uso de algoritmos escogidos para formar agrupamientos. En el análisis de conglomerados no se conoce *a priori* el grupo de pertenencia de los elementos a agrupar. Para cualquier algoritmo de clúster será necesario seleccionar una medida de distancia entre ellas y un método de agrupamiento (Balzarini et al., 2015, p. 90). El análisis de clúster se usa para clasificar: especies vegetales, sistemas de producción, suelos, socioeconomía de los productores, condiciones de mercados, entre otros.

Las poblaciones vegetales pueden considerar en la clasificación las características agronómicas, productivas, sanitarias, organolépticas y químicas. Los sistemas de producción pueden incluir variables como: tamaño de la finca, área dedicada a la ganadería, cultivos anuales, de ciclo corto y perennes, intensidad de uso de la tecnología, productividad y rentabilidad.

En la clasificación de los suelos agrícolas se puede considerar variables como: características físicas (color, textura, estructura, profundidad), químicas (pH, contenido de nutrientes, conductividad eléctrica, materia orgánica) y biológicas (especies y poblaciones de microorganismos).

Los pasos a seguir para el análisis de clúster son los siguientes:

- Organizar la base de datos cuantitativos promedios de  $n$  casos y  $k$  variables
- Establecer indicadores de distancia que indique la medida en que cada par se parece entre sí
- Conformar grupos de acuerdo a las medidas de similaridad o de distancia
- Comparar los grupos, unos con otros
- Validar los resultados a nivel de campo

Las medidas de similaridad se basan en el cálculo de las distancias como: euclídea, euclídea al cuadrado y Manhattan. Los Coeficientes de correlación de Pearson ( $r$ ) y de Spearman ( $r_s$ ), también se usan como medidas de distancia para datos continuos (Balzarini et al., 2015, p. 50).

Además de la distancia euclídea, referida para la medición de las distancias multivariadas, se recomiendan usar la distancia euclídea al cuadrado y la distancia Manhattan.

La distancia euclídea al cuadrado se calcula con la fórmula:

$$D^2 = (X_2 - X_1)^2 + (Y_2 - Y_1)^2 \quad [94]$$

La distancia Manhattan se calcula con la fórmula

$$M = \sum[(X_2 - X_1) + (Y_2 - Y_1)] \quad [95]$$

En el análisis de conglomerados jerárquico para agrupar  $n$  muestras (cada una  $k$  dimensional), se calcula primero una matriz de distancias ( $n \times n$ ) que contiene las interdistancias entre todos los pares de muestras (Balzarini et al., 2015, pp. 90-93).

Uno de los métodos es el de *k-medias*, donde se forman  $k$  grupos y el algoritmo evalúa múltiples particiones de los objetos en  $k$  grupos, de tal modo que la distancia entre grupos sea máxima y dentro de grupos sea mínima. Se conoce *a priori* el número de  $k$  grupos. El número de  $k$  grupos debe ser especificado "*a priori*" por el investigador o como parte del procedimiento. Estos métodos utilizan la matriz de datos originales para construir el agrupamiento (Balzarini et al., 2015, p. 93).

En el proceso de agrupamiento, al inicio, cada caso es considerado como un grupo en sí mismo, luego se van fusionando, conformando grupos cercanos por similitud, hasta que todos confluyen en un gran grupo. Para el agrupamiento, hay varios métodos como: centroide, vecino más cercano, vecino más lejano, vinculación promedio y el método de Ward. Este último ha sido adoptado intensamente en las investigaciones agropecuarias.

**Método de centroide.**- Comienza uniendo dos observaciones más cercanas, el grupo se sustituye por un centroide, todas las observaciones toman un valor promedio, se vuelve a calcular, se une otro par de observaciones y se recalcula la matriz, y así sucesivamente.

**Método del vecino más cercano.**- Las distancias entre los grupos se calculan tomando las distancias más cercanas en vez del promedio. Es muy sensible a los datos atípicos.

**Método del vecino más lejano.**- Las distancias entre los grupos se calculan tomando las distancias más lejanas de cada grupo.

**Método de la vinculación promedio.**- La distancia entre los grupos se obtiene calculando la distancia promedio entre todos los pares de observaciones que puedan formarse.

**Método de Ward.**- Maximiza la homogeneidad dentro de los grupos y plantea todas las posibles fusiones en cada etapa concreta. Calcula los centroides de los grupos resultantes de las posibles fusiones y luego la distancia al centroide de todas las observaciones del grupo (SC total). La solución elegida es la que tiene menor suma de cuadrados total.

**Ejercicio 101:**

Analizar la calidad organoléptica en función de los métodos de beneficio y de los cultivares de café robusta en la Amazonía ecuatoriana (Duicela et al., 2018, p. 11).

**Datos y análisis:**

Las calificaciones sensoriales de los cafés robustas amazónicos en función de los métodos de beneficio y genotipos cultivados, se exponen en el Cuadro 72. Los promedios más altos de calidad organoléptica, en Sucumbíos y Orellana, se lograron usando el beneficio húmedo enzimático.

Los clones NP-2024 en Sucumbíos y NP-3056 en Orellana tuvieron los mayores promedios de calidad sensorial. La muestra del clon NP 3056, beneficiada con el uso de enzimas pectolíticas, logró 81,92 puntos SCA que se trata de un café especial.

El diagrama de similitudes entre los métodos de beneficio, usando el ACJ, se exponen en el Gráfico 30. Los métodos de beneficio de café robusta por vía seca (natural) y semihúmedo (semilavado) permiten obtener calidades sensoriales similares.

Los cafés lavados preparados con los métodos de beneficio húmedo convencional y húmedo enzimático resultaron similares en calidad organoléptica.

*Cuadro 72. Calidad de taza de genotipos de café robusta, en función de cuatro métodos de beneficio*

<u>Orellana</u>		Genotipos de café robusta				Medias de métodos de beneficio
Métodos de beneficio	Tipos de Café	Genotipo local	NP 2024	NP 3013	NP 3056	
Beneficio por vía húmeda (convencional)	Lavado 1	77,04	77,54	77,96	79,25	77,95
Beneficio húmedo enzimático	Lavado 2	75,29	79,54	77,50	<b>81,92</b>	78,56
Beneficio semihúmedo	Semilavado	77,25	75,75	76,54	78,88	77,11
Beneficio por vía seca	Natural	77,45	76,10	76,10	71,00	75,16
Medias de genotipos		76,76	77,23	77,03	77,76	77,19

<u>Sucumbíos</u>		Genotipos de café robusta				Media de métodos de beneficio
Métodos de beneficio	Tipo de Café	Genotipo local	NP 2024	NP 3013	NP 3056	
Beneficio por vía húmeda (convencional)	Lavado 1	76,67	78,88	78,75	77,54	77,96
Beneficio húmedo enzimático	Lavado 2	78,75	78,83	78,17	77,29	78,26
Beneficio semihúmedo	Semilavado	73,21	77,00	78,00	76,13	76,09
Beneficio por vía seca	Natural	79,00	78,70	76,90	78,21	78,20
Media de genotipos		76,91	78,35	77,96	77,29	77,63

Los clones de café robusta NP 3056 y NP 2024 mostraron similares calidades de taza. El clon NP 3013 y los genotipos locales (híbridos de origen desconocido) resultaron similares en calidad sensorial. La similitud en calidad de taza en función de los genotipos se indica en el Gráfico 31.

El análisis de la interacción entre territorios y métodos de beneficio, usando la técnica de ACJ, permitió determinar que tienen similitud en la calidad organoléptica, excepto el método de beneficio por vía seca que resultó diferente con respecto a los métodos de beneficio (Gráfico 32).

Un aspecto fundamental en la interpretación de ACJ es el grado de confiabilidad en la matriz de datos, que se mide a través del Coeficientes de correlación cofenética, que en el dendrograma indica las distancias que mejor reflejan las diferencias verdaderas. En el presente ejercicio, en los tres ACJ, los coeficientes de correlación cofenética fueron: 0,888, 0,714 y 0,915, respectivamente; por lo tanto, los resultados son confiable.

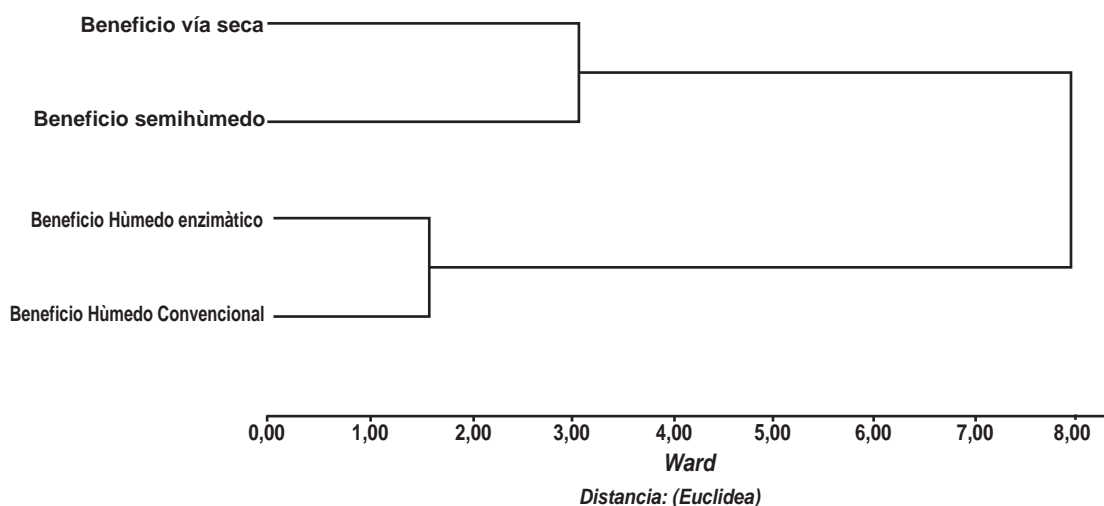


Gráfico 30. Calidad de taza de los cafés robustas en función de los métodos de beneficio

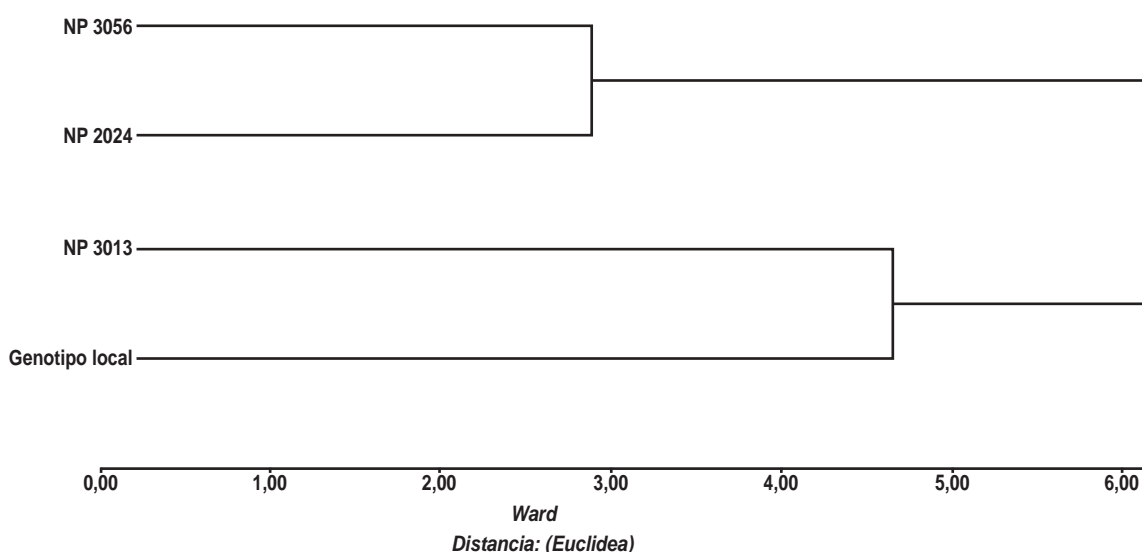


Gráfico 31. Similitud en calidad de taza en función de los genotipos de café robusta

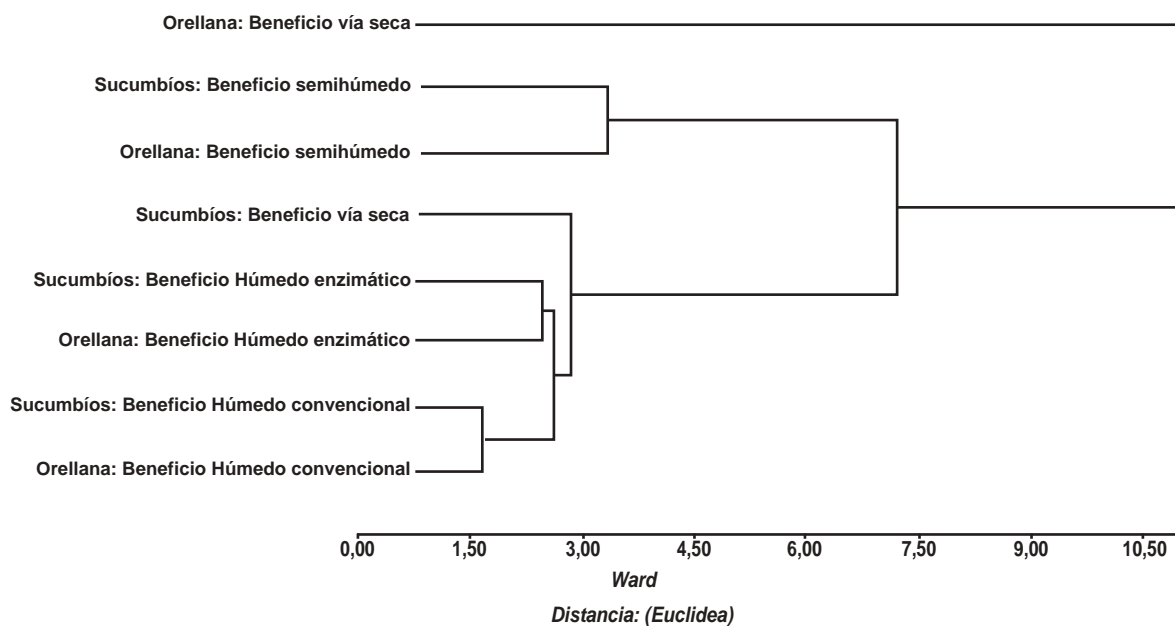


Gráfico 32. Similitud en taza de los cafés robustas de Orellana y Sucumbíos, según métodos de beneficio

### Decisión

En las provincias Orellana y Sucumbíos, los métodos de beneficio húmedo convencional (lavados) en café robustas dieron calidades sensoriales similares. Los cafés naturales (vía seca) dieron tazas de menor puntaje sensorial comparados con los cafés lavados preparados con los métodos húmedo convencional y húmedo enzimático. Los coeficientes de Correlación cofenética permiten afirmar que los resultados, en los tres análisis, son confiables.



## CAPÍTULO V: INVESTIGACIÓN RELACIONAL

La investigación relacional es un método de investigación “no experimental” en el cual se mide la intensidad de la asociación estadística entre dos variables, sin esperar encontrar relaciones de dependencia. En este nivel investigativo, se amplían los estudios descriptivos de las poblaciones y muestras hacia la búsqueda de relaciones entre las variables de interés. Las hipótesis en la investigación relacional se enfocan a determinar concordancias y asociaciones estadísticas entre las variables de interés, en circunstancias en que no hay manipulación de factores.

Cuando los datos son cuantitativos y continuos y las series de datos tienen distribución normal, se hacen correlaciones y regresiones con finalidad explicativa y predictiva. En este caso, el modelo lineal explica la relación entre dos variables, una dependiente (Y) y otra independiente (X), que se expresa en el modelo general:  $Y = f(X)$ . Para explicar la dependencia de (Y) en función de varias variables independientes selectas  $(X_1, X_2, \dots, X_n)$ , se usa el modelo general:  $Y = f(X_1 + X_2 + \dots + X_n)$ .

En la investigación cualitativa, los datos no son cuantitativos, se analizan datos nominales y ordinales, aunque se puede combinar con datos cuantitativos, cuando el investigador tiene interés de explicar si hay o no relación entre las variables:  $X_1, X_2, \dots, X_n$ . De esto trata la investigación relacional. A partir de los estudios relacionales se pueden desarrollar modelos predictivos solo que haya evidencia de alta repetibilidad.

En este nivel investigativo, como ejemplos, se respondería a las preguntas de investigación: ¿Hay relación entre el grado de escolaridad y los ingresos?, ¿El aroma del café se asocia con la altitud de la zona de cultivo?, ¿Se cumple la teoría mendeliana, de que en F<sub>2</sub>, la frecuencia de plantas con flores rojas es 75 % y frecuencia de las plantas con flores blancas es 25 %? (Un progenitor con dominancia completa para flores rojas).

Las técnicas usadas en este nivel investigativo incluyen el análisis de frecuencias en dos direcciones, el Coeficiente de correlación de Pearson, el Coeficiente de correlación por rangos de Spearman, la prueba  $\chi^2$  y la prueba de Kappa de Cohen.

### 5.1 ANÁLISIS DE FRECUENCIAS EN DOS SENTIDOS

Cuando en una investigación se fija como objetivo describir el comportamiento de una variable en una población o muestra, se inicia con el análisis de frecuencias univariado (en un sentido), para lo cual, la serie de datos se agrupa en clases, se elabora la tabla de frecuencias y se representa en un histograma. Pero si el objetivo es relacionar dos variables simultáneamente, se debe realizar el análisis de frecuencias en dos sentidos o de dos variables (dos direcciones).

Para el análisis de frecuencias en dos sentidos, se organiza en forma lógica las series de datos, agrupadas en clases, verticalmente y horizontalmente, luego se añaden las frecuencias absolutas, absolutas acumuladas, relativas y relativas acumuladas. Complementariamente a la tabla de frecuencias, se elaboran las representaciones gráficas como: histogramas, polígonos de frecuencias y ojivas, por cada variable en estudio. Esta técnica es muy útil para asociar variables de interés y tomar decisiones rápidas de selección de genotipos, por ejemplo, de alta producción y de porte bajo.



**Ejercicio 102:**

Determinar la relación entre altura de la planta y producción de café oro (g.planta<sup>-1</sup>) a partir de los datos de una muestra n = 40 plantas de café robusta, con base al análisis de frecuencias en dos sentidos.

**Datos y análisis:**

Los datos se organizan en clases en los dos sentidos: en columnas la variable “altura de planta” (cm) clasificada en seis categorías y en las filas la variable producción de café oro (g.planta<sup>-1</sup>) organizada en cinco categorías. En el Cuadro 73, se indica el análisis de frecuencias en dos sentidos, destacándose que hay 4 plantas con menos de 300 cm de altura y 5 plantas que producen más de 1100 g.planta<sup>-1</sup>. No se encontró plantas de porte bajo y alta producción simultáneamente, en este muestreo.

*Cuadro 73. Análisis de frecuencias absolutas (FA), relativas (FR) y acumuladas (FAA y FRA) de las variables altura de planta y producción, en una muestra aleatoria de 40 plantas de café robusta*

		Altura de planta (cm)						FA	FAA	FR (%)	FRA (%)
		≤ 300	301-350	351-400	401-450	451-500	> 500				
Producción (g.planta <sup>-1</sup> )	> 1100				I	II	II	5	5	13	13
	801 - 1100			II	III	IIII		10	15	25	38
	501 - 800		II	IIII	IIII	I		15	30	38	75
	201 - 500	III	III	I				7	37	18	93
	≤ 200	I	II					3	40	8	100
FA		4	7	9	10	8	2	40		100	
FAA		4	11	20	30	38	40				
FR (%)		10	18	23	25	20	5	100			
FRA (%)		10	28	50	75	95	100				

**Decisión:**

La asociación de las frecuencias indica una tendencia: a mayor altura de planta más producción.

**5.2 COEFICIENTE DE CORRELACIÓN**

Los estudios correlacionales se fundamentan en el modelo:  $X_1 = f(X_2)$  y los estudios de causalidad en modelo:  $Y = f(X)$ . En estas dos circunstancias se puede usar el Coeficiente de correlación momento producto  $r$  de Pearson (técnica paramétrica) y el Coeficiente de correlación por rangos de Spearman (técnica no paramétrica).

El Coeficiente de correlación de Pearson ( $r$ ) se usa cuando los datos son cuantitativos con tamaño de muestra adecuado y las series de datos tienen distribución normal. La correlación por rangos de Spearman ( $r_s$ ) no exige el cumplimiento de las condiciones indicadas para  $r$ . En la práctica,  $r_s$ , se puede usar para relacionar series de datos cuantitativos, cualitativos o mixtos.

La interpretación de los dos Coeficientes de correlación es equivalente. Una forma de valorar la correlación es mediante un dispersograma (Gráfico 33), donde se pueden distinguir las siguientes situaciones: hay correlación positiva ( $r \rightarrow + 1$ ), hay correlación negativa ( $r \rightarrow - 1$ ) o no hay correlación ( $r \rightarrow 0$ ).

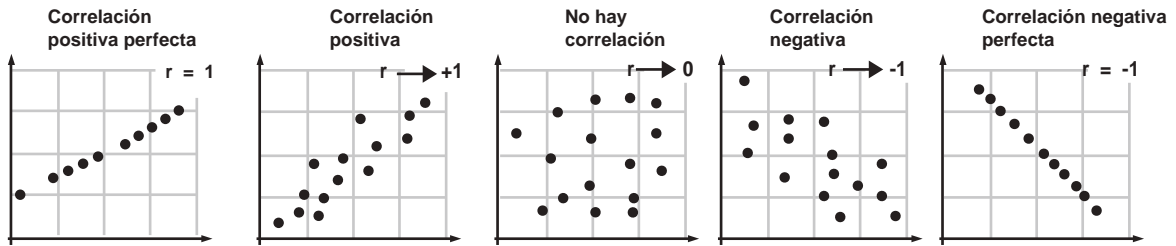


Gráfico 33. Diagramas de los tipos de correlaciones entre variables y su aproximación al coeficiente  $r$

### 5.2.1 COEFICIENTE DE CORRELACIÓN DE PEARSON

El Coeficiente de correlación momento producto de Pearson ( $r$ ) es una técnica paramétrica que mide el grado de relación entre dos variables, variando de  $-1$  a  $+1$ .

La fórmula más usada para calcular el Coeficiente de correlación es la siguiente:

$$r = \frac{Cov_{XY}}{\sqrt{S_X^2(S_Y^2)}} \quad [96]$$

$$r = \frac{Cov_{XY}}{(S_X)(S_Y)} \quad [97]$$

La covarianza se calcula con la fórmula:

$$Cov_{xy} = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{n - 1} \quad [98]$$

Las varianzas de las variables X e Y se calculan con las fórmulas pertinentes:

La hipótesis estadística se conjetura como:

$H_0$  = No hay correlación significativa

$H_1$  = Si hay correlación significativa

La regla de decisión es la siguiente:

Si  $r$  calculada  $<$   $r$  crítico: Aceptar  $H_0$

Si  $r$  calculada  $\geq$   $r$  crítico: Aceptar  $H_1$

Los valores críticos de  $r$  se obtienen en la tabla  $r$  de Pearson o usando Excel, considerando el nivel de significación ( $\alpha$ ) y los grados de libertad ( $n - 2$ ). En este caso,  $n$  representa en número de pares de datos observados.

En Excel, el coeficiente de correlación  $r$  se obtiene con las siguientes funciones:

=COEF.DE.CORREL(matriz1;matriz2).

=PEARSON(matriz1;matriz2).

La prueba de significación de  $r$  se realiza con la prueba  $t$ <sup>18</sup>, usando la fórmula:

$$t_{calculada} = \frac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} \quad [99]$$

Dónde:

$t$  = Estadístico

$r$  = Coeficiente de correlación

$n$  = Número de pares de datos

<sup>18</sup> En el Anexo 5 se indica el procedimiento de lectura en la tabla  $t$  de Student para obtener los valores críticos

El valor  $t_{\text{crítico}}$  para un nivel de significación ( $\alpha$ ) con grados de libertad  $n - 2$ , se obtiene en la tabla de valores críticos  $t$  de Student o usando Excel.

En Excel, para obtener el valor crítico de  $t$  se usa la función:

=DISTR.T.INV(probabilidad; grados de libertad).

El valor  $p$  de  $t$  se obtiene directamente en Excel con la función:

=DISTR.T(X; grados de libertad; colas)

### Ejercicio 103:

Calcular la correlación lineal de Pearson para las variables altura de los cafetos ( $X_1$ ) y el número de pares de ramas ( $X_2$ ), con un nivel de confianza del 95 %.

#### Datos y análisis:

Los datos de altura de planta y número de pares ramas se exponen en el Cuadro 74.

Cuadro 74. Altura y número de pares de ramas en cafetos en una muestra aleatoria de nueve plantas

Variables	Muestra aleatoria de plantas de café								
	1	2	3	4	5	6	7	8	9
Altura de planta (cm)	230	200	180	210	205	245	260	195	190
Nº pares de ramas	24	20	17	25	19	26	24	19	20

En el Cuadro 75 se exponen los datos de las variables  $X_1$  y  $X_2$  y los cálculos de potencias y productos que sirven de base para aplicar las fórmulas pertinentes.

Cuadro 75. Datos de las variables altura de planta y número de ramas, cuadrados y productos

	Altura de planta		Número de pares de ramas		Producto
	$X_1$	$X_1^2$	$X_2$	$X_2^2$	$X_1(X_2)$
	230	52.900	24	576	5520
	200	40.000	20	400	4000
	180	32.400	17	289	3060
	210	44.100	25	625	5250
	205	42.025	19	361	3895
	245	60.025	26	676	6370
	260	67.600	24	576	6240
	195	38.025	19	361	3705
	190	36.100	20	400	3800
$\Sigma$	1915	413.175	194	4264	41.840

#### Cálculos de las varianzas de $X_1$ y $X_2$ y covarianza

$$S^2_{X_1} = \frac{413175 - \frac{(1915)^2}{9}}{8} = 713,19$$

$$S^2_{X_2} = \frac{4264 - \frac{(194)^2}{9}}{8} = 10,28$$

$$\text{COV}_{X_1 \times X_2} = \frac{\sum 41840 - \frac{(1915)(194)}{9}}{8} = 70,14$$

$$r = \frac{\text{Cov}_{X_1 \times X_2}}{\sqrt{S_{X_1}^2(S_{X_2}^2)}} \rightarrow r = \frac{70,14}{\sqrt{713,19(10,28)}} = 0,819 \rightarrow r_{\text{calculada}} = 0,819$$

**Valores críticos de r:**

Los valores críticos de  $r$  se obtienen en la tabla de Pearson considerando un nivel de significación ( $\alpha = 0,05$ ) y los grados de libertad ( $GL = n - 2$ )  $\rightarrow r_{(0,05 \text{ y } GL = 7)} = 0,666$

**Hipótesis estadística:**

$H_0$  = No hay correlación significativa

$H_1$  = Si hay correlación significativa

**Regla de decisión:**

Si  $r_{\text{calculada}} < r_{0,05}$ : Aceptar  $H_0$

Si  $r_{\text{calculada}} \geq r_{0,05}$ : Aceptar  $H_1$

**Interpretación:**

$$r_{\text{calculada}} = 0,819 > r_{(0,05 \text{ y } GL = 7)} = 0,666 *$$

Si hay correlación estadísticamente significativa con el 95 % de confianza.

**Prueba de significación de r**

$$t_{\text{calculada}} = \frac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} \rightarrow t_{\text{calculada}} = \frac{0,819\sqrt{(9-2)}}{\sqrt{(1-0,819^2)}} \rightarrow t_{\text{calculada}} = \frac{0,819(2,646)}{0,574} = 3,78$$

$$t_{\text{calculada}} = 3,78 > t_{(0,05 \text{ y } GL = n - 2)} = 2,365 (**)$$

$$t_{\text{calculada}} = 3,78 > t_{(0,01 \text{ y } GL = n - 2)} = 3,499 (**)$$

El valor  $p$  de  $t$ , en Excel, se obtiene con la función:

$$=\text{DISTR.T}(X;\text{grados\_de\_libertad\_colas})$$

Aplicando al ejercicio:

$$=\text{DISTR.T}(t_{\text{calculada}} = 3,78; GL = 7; 2 \text{ colas}) \rightarrow p \text{ de } t = 0,0069 \sim 0,69 \% \text{ (probabilidad de equivocarse insignificante)}$$

$$\text{El nivel de confianza con la prueba } t = (1 - 0,0069) = 0,9931 \sim 99,31 \%$$

**Decisión:**

Si hay correlación positiva y altamente significativa con el 99,31 % de confianza.

**5.2.2 COEFICIENTE DE CORRELACIÓN POR RANGOS DE SPEARMAN**

El Coeficiente de correlación de Spearman ( $r_s$ ) es una técnica no paramétrica que mide el grado de asociación entre dos variables, que debe usarse principalmente cuando el tamaño de muestra es pequeño, no se han verificado los supuestos de normalidad de las distribuciones en las dos series de datos a relacionar y cuando los datos provienen de mediciones en escala ordinal (Manterola et al., 2018, p. 684; Mondragón, 2014, p. 100).

El Coeficiente  $r_s$  puede variar desde  $-1$  hasta  $+1$ . El procedimiento se basa en la transformación de los datos originales a rangos (posiciones), en orden ascendente. Si hay

pocos datos, la transformación a rangos se realiza por simple inspección. Cuando las series de datos tienen alto número de observaciones y, además, existen valores repetidos (ligas), la transformación a rangos se puede realizar usando la función de Excel:

=JERARQUIA.MEDIA(número, referencia,[orden]).

En el caso de usar Excel, el “número” corresponde a la celda donde se ubica el primer dato a transformar (p.e. B2), la referencia es la matriz de la serie original (p.e. B2:B10) y el orden es la indicación ascendente (0) o descendente (1).

La fórmula para calcular el Coeficiente de correlación de Spearman<sup>19</sup> es la siguiente:

$$r_s = 1 - \frac{6 \sum d_i^2}{n^3 - n} \quad [100]$$

Dónde:

$r_s$  = Estadístico de Spearman

$\sum d_i^2$  = Sumatoria de las diferencias de rangos al cuadrado

$n$  = Número de pares de datos

La hipótesis estadística se formula como:

$H_0$  = No hay correlación significativa entre variables.

$H_1$  = Si hay correlación significativa entre variables.

La regla de decisión es la siguiente:

Si  $r_s$  calculada <  $r_s$  crítico: Se acepta  $H_0$ . No hay correlación significativa

Si  $r_s$  calculada  $\geq r_s$  crítico: Se acepta  $H_1$ . Si hay correlación significativa

#### Ejercicio 104:

Calcular la correlación por rangos de Spearman para las variables altura de planta (cm) y número de ramas, en una muestra de  $n = 9$  cafetos, cuyos datos se indican a continuación:

#### Datos y análisis:

Los datos de las variables altura de planta ( $X_1$ ) y Número de pares de ramas ( $X_2$ ), la transformación a rangos y cálculo de las diferencias, se indican en el Cuadro 76.

Cuadro 76. Transformación a rangos de los datos altura de planta y pares de ramas y diferencias de rangos

N	Altura de planta (cm)		Número de pares de ramas		d	d <sup>2</sup>
	$X_1$	Rango $X_1$	$X_2$	Rango $X_2$		
1	230	7	24	6,5	0,5	0,25
2	200	4	20	4,5	-0,5	0,25
3	180	1	17	1	0	0
4	210	6	25	8	-2,0	4,0
5	205	5	19	2,5	2,5	6,25
6	245	8	26	9	-1,0	1,0
7	260	9	24	6,5	2,5	6,25
8	195	3	19	2,5	0,5	0,25
9	190	2	20	4,5	-2,5	6,25
					$\Sigma$	24,5

Nota: La diferencia  $d = \text{Rango } X_1 - \text{Rango } X_2$

<sup>19</sup> En el Anexo 2 se expone la tabla parcial de los valores críticos de la correlación por rangos de Spearman

$$r_s = 1 - \frac{6 \sum d_i^2}{n^3 - n} \quad \rightarrow \quad r_s = 1 - \frac{6(24,5)}{9^3 - 9} \quad \rightarrow \quad r_s = 0,796$$

**Valor crítico:**  $r_s (\alpha = 0,05 \text{ y } n = 9) = 0,60$

**Hipótesis estadística:**

$H_0$  = No hay correlación significativa entre  $X_1$  y  $X_2$

$H_1$  = Si hay correlación significativa entre  $X_1$  y  $X_2$

**Regla de decisión:**

Si  $r_s$  calculada  $<$   $r_s$  valor crítico: Aceptar  $H_0$ .

Si  $r_s$  calculada  $\geq$   $r_s$  valor crítico: Aceptar  $H_1$ .

**Interpretación:**

$r_s$  calculada = 0,796  $>$   $r_s$  crítico = 0,60 \*: Aceptar  $H_1$ . Si hay correlación significativa.

**Decisión:**

Hay correlación positiva y significativa entre  $X_1$  y  $X_2$  con el 95 % de confianza.

### 5.3 PRUEBAS DE CONCORDANCIA

Las pruebas de concordancia permiten valorar la confiabilidad de las mediciones realizadas por diferentes observadores o métodos, identificar las posibles fuentes de error en los datos y mejorar la calidad de la investigación. Un uso común de las pruebas de concordancia es la medición del grado de acuerdo entre frecuencias observadas y esperadas, por ejemplo, en la teoría mendeliana.

#### 5.3.1 PRUEBA CUADRADA

La prueba  $\chi^2$  (Chi cuadrada o Ji cuadrada) es una técnica no paramétrica que permite medir la discrepancia entre un valor observado y un valor esperado (teórico) y en caso de haber, determinar si es atribuible a la causalidad (hay significación estadística) o a la casualidad (no hay significación). Esta herramienta es usada especialmente en análisis de conteos (Gómez & Gómez, 1983, p. 461-466).

La fórmula general para calcular Chi cuadrado<sup>20</sup> es la siguiente:

$$\chi^2 = \frac{(O - E)^2}{E} \quad [101]$$

Dónde:

$\chi^2$  = Estadístico Chi cuadrado

O = Valor observado

E = Valor esperado

La hipótesis se formula como:

$H_0$ : Observado = Esperado. Si hay concordancia

$H_1$ : Observado  $\neq$  Esperado. No hay concordancia

La regla de decisión se plantea de la siguiente manera:

Si  $\chi^2$  calculada  $<$   $\chi^2$  valor teórico: Aceptar  $H_0$

<sup>20</sup> En el Anexo 4 se indica el procedimiento de lectura en la tabla Chi cuadrado para obtener los valores críticos

Si  $\chi^2_{\text{calculada}} \geq \chi^2_{\text{valor teórico}}$ : Aceptar  $H_1$

Grados de libertad (GL) = (Columnas - 1)(Hileras - 1) = (C - 1)(H - 1)

Cuando hay GL= 1, se usa la fórmula con la corrección de Yates:

$$\chi^2 = \frac{\sum[(O - E) - 0,5]^2}{E} \quad [102]$$

Las pruebas  $\chi^2$  se usan en genética cualitativa, donde se busca determinar la concordancia entre frecuencias observadas y esperadas en progenies derivadas de cruzamientos entre progenitores con caracteres conocidos (Robles, 1986, p. 83). Como ejemplo, en el cruzamiento entre un progenitor ♂ homocigótico dominante con dominancia completa, que tiene semillas amarillas (AA) y un progenitor ♀ de semillas verdes que es homocigoto recesivo (aa), en la gametogénesis se originan gametos haploides (A, a) cuyos alelos se recombinan y en F1 forman individuos heterocigótico Aa (100 % de semillas amarillas). Las plantas heterocigóticas “Aa”, al autofecundarse cambian las frecuencias génicas en F2, a la relación:  $\frac{1}{4} AA + \frac{2}{4} Aa + \frac{1}{4} aa$ , que fenotípicamente corresponden a:  $\frac{3}{4}$  semillas amarillas +  $\frac{1}{4}$  semillas verdes (Figura 27).

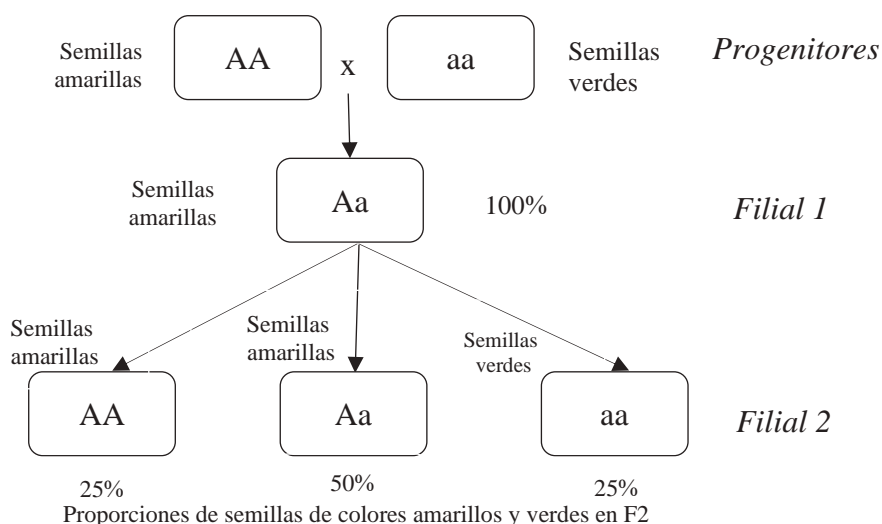


Figura 27. Frecuencias de semillas amarillas y verdes en F2 derivadas del cruce monohíbrido

### Ejercicio 105:

Se realizó un cruzamiento entre una planta con semillas de color amarillo con dominancia completa (AA) y una planta con semillas de color verde, recesivo (aa), originando en F1 el 100 % de plantas con semillas amarillas. Estas plantas se auto fecundaron, y se observó que en F2, había 200 plantas que daban semillas amarillas y 50 semillas verdes. ¿Hay concordancia con la teoría mendeliana de que la proporción de amarillas y verdes es 3 : 1?..

### Datos y análisis:

Del cruce entre progenitores de semillas amarillas (homocigótico dominante) y verdes (homocigótico recesivo), en F2 se observaron 200 plantas con semillas amarillas y 50 plantas con semillas verdes. Según la teoría mendeliana, lo esperado es  $\frac{3}{4}$  de semillas amarillas y  $\frac{1}{4}$  de semillas verdes.

**Cálculos:**

Si  $N = 250$ , el 75 % corresponde a 188 plantas que deberían tener semillas amarillas ( $\frac{3}{4} N$ ) y 62 plantas de semillas verdes ( $\frac{1}{4} N$ ).

Cálculo de Chi cuadrado para las plantas con semillas amarillas:

$$\chi^2 = \frac{\Sigma[(O - E) - 0,5]^2}{E} \rightarrow \chi^2 = \frac{[(200 - 188) - 0,5]^2}{188} = 0,70$$

Cálculo de Chi cuadrado para las plantas con semillas verdes:

$$\chi^2 = \frac{\Sigma[(O - E) - 0,5]^2}{E} \rightarrow \chi^2 = \frac{[(50 - 62) - 0,5]^2}{62} = 2,52$$

Cálculo de Chi cuadrado para todas las plantas:

$$\chi^2 = \frac{\Sigma[(O - E) - 0,5]^2}{E} \rightarrow \chi^2 = \frac{[(200 - 188) - 0,5]^2}{188} + \frac{[(50 - 62) - 0,5]^2}{62} = 3,22$$

$\chi^2 = 3,22$

**Valores críticos:**

$$\chi^2_{0,05} = 3,841 \dots \dots \dots \text{ y } \chi^2_{0,01} = 6,635$$

**Hipótesis:**

$H_0$ : Frecuencia observadas = frecuencias esperadas (Si hay concordancia)

$H_1$ : Frecuencias observadas  $\neq$  frecuencias esperadas (No hay concordancia)

**Regla de decisión:**

Si  $\chi^2_{calculada} < \chi^2_{valor\ crítico}$ : Aceptar  $H_0$

Si  $\chi^2_{calculada} \geq \chi^2_{valor\ crítico}$ : Aceptar  $H_1$

**Interpretación:**

$$\chi^2_{calculada} = 3,23 < \chi^2_{0,01} = 6,635: \text{ Aceptar } H_0$$

**Decisión:**

Si hay concordancia entre las frecuencias observadas y esperadas, de acuerdo a la teoría mendeliana.

**Ejercicio 106:**

Determinar si los resultados obtenidos en F2, en el color y rugosidad de las semillas, tienen concordancia con la teoría mendeliana, donde las proporciones esperadas del cruce dihíbrido, son: 9 : 3 : 3 : 1 (Martínez y Sáenz, 2003, p. 25). En el ensayo se observaron 556 plantas, de las cuales 315 tuvieron semillas lisas y amarillas, 108 lisas y verdes, 101 semillas rugosas y amarillas y 32 rugosas y verdes.

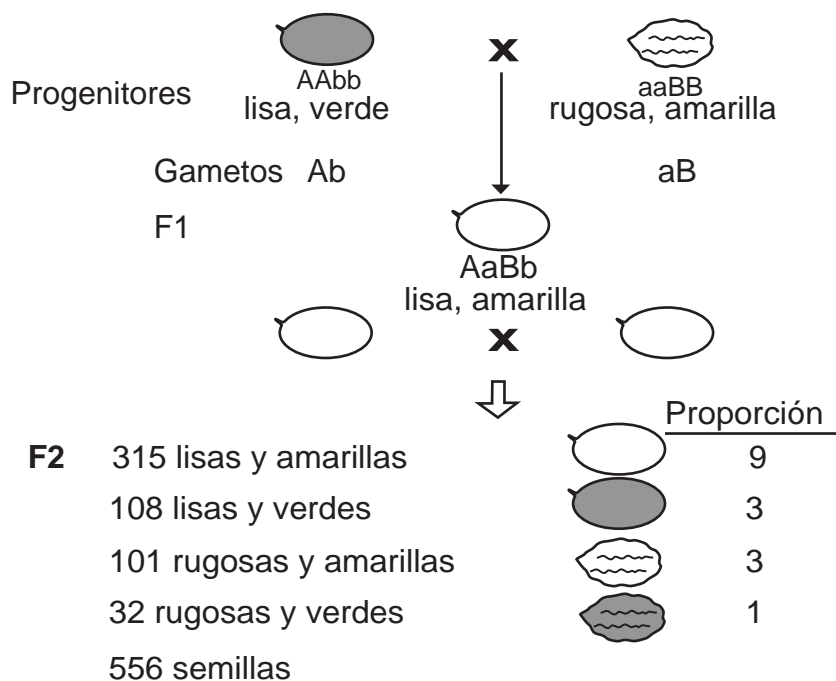
**Datos y análisis:**

El diagrama del cruce dihíbrido, entre progenitores ( $\sigma$  y  $\varphi$ ) con características de color y rugosidad de la semilla y las frecuencias observadas de plantas en F2, se indica en la Figura 28.

**Cálculo de las frecuencias esperadas (E):**

Los cálculos de las frecuencias observadas y esperadas de plantas con clasificación de las semillas, según la forma y color, se calcula de la forma siguiente:





Fuente: Martínez y Sáenz (2003, p. 25)

Figura 28. Frecuencias observadas en F2 de semillas derivadas de un cruce dihíbrido

$$E_{\text{categoría}} = \frac{\text{Proporción mendeliana}}{\text{Total de proporciones}} (\text{Total observado})$$

$$E_{\text{lisas y amarillas}} = \frac{9}{16} (556) = 313$$

$$E_{\text{lisas y verdes}} = \frac{3}{16} (556) = 104$$

$$E_{\text{rugosas y amarillas}} = \frac{3}{16} (556) = 104$$

$$E_{\text{rugosas y verdes}} = \frac{1}{16} (556) = 35$$

En el Cuadro 77, se expone los datos de 556 plantas donde se observaron: 315 lisas y amarillas, 108 lisas y verdes, 101 rugosas y amarillas y 32 rugosas y verdes. En el referido cuadro se indica, además, el cálculo del estadístico Chi-cuadrado.

Cuadro 77. Frecuencias observadas y esperadas de plantas con semillas clasificadas por forma y color y cálculo de  $\chi^2$

Pantas en F2	Semillas lisas y amarillos	Semillas lisas y verdes	Semillas rugosas y amarillas	Semillas rugosas y verdes	Total
Proporción mendeliana	9	3	3	1	16
Plantas observadas (O)	315	108	101	32	556
Plantas esperadas (E)	313	104	104	35	556
(O - E)	2	4	-3	-3	--
$\chi^2 = \frac{(O - E)^2}{E}$	0,02	0,13	0,10	0,22	<b>0,47</b>

**Hipótesis estadísticas:**

H<sub>0</sub>: Frecuencias observadas = frecuencias esperadas

H<sub>1</sub>: Frecuencias observadas ≠ frecuencias esperadas

**Regla de decisión:**

$\chi^2_{\text{calculado}} < \chi^2_{\text{crítico}}$ : Aceptar H<sub>0</sub>

$\chi^2_{\text{calculado}} > \chi^2_{\text{crítico}}$ : Aceptar H<sub>1</sub>

**Valores críticos:**

Los valores críticos de Chi cuadrado se obtienen en la tabla específica o usando Excel, con la función:

=PRUEBA.CHI.INV(probabilidad; grados de libertad).

Los grados de libertad (GL) = (Columnas - 1)(Hileras - 1) = (4 - 1)(2 - 1) = 3

**Interpretación:**

$\chi^2_{\text{calculado}} = 0,47 < \chi^2_{0,05} = 7,815$  : Aceptar H<sub>0</sub>

El valor *p* de Chi cuadrado se obtiene en Excel con la función:

=PRUEBA.CHI(rango real; rango esperado)

*p* = 0,9254 ~ 92,5 %

**Decisión:**

Las frecuencias observadas de plantas, en F<sub>2</sub>, derivadas del cruce dihíbrido para los caracteres forma y color de las semillas, tienen concordancia con las frecuencias esperadas 9 : 3 : 3 : 1, según la teoría mendeliana (*p* > 0,05).

5.3.2 COEFICIENTE DE CONTINGENCIA DE PEARSON

Este coeficiente (C) es una medida de la intensidad de la relación entre dos o más variables cualitativas, que varía desde 0 hasta 1, basado en la comparación de las frecuencias observadas y esperadas de la prueba Chi cuadrado. En la práctica, el coeficiente C es complementario. El valor informativo de Chi cuadrado es reducido debido a que su límite superior depende de los valores que adquieren las variables y del tamaño de la población.

Un valor de C que tiende a 1 significa que la asociación estadística se aproxima al máximo esperado y cuando C tiende a 0 significa que no hay asociación estadística. En base de este concepto, se usa el coeficiente C para estimar la certidumbre de la prueba estadística.

La fórmula para calcular el Coeficiente de contingencia de Pearson es la siguiente:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \quad [103]$$

Dónde:

C = Coeficiente de contingencia de Pearson

$\chi^2$  = Valor de Chi-cuadrado calculado

n = Tamaño de muestra

La certidumbre de la prueba estadística (CPE), se calcula con la fórmula:

$$CPE. = 1 - \sqrt{\frac{\chi^2}{\chi^2 + n}} = 1 - C \quad [104]$$

**Ejercicio 107:**

A partir de los datos de preferencias electorales, donde:  $n = 545$  y  $\chi^2 = 125,4$ , calcular el Coeficiente de contingencia y la certidumbre de la prueba estadística.

**Datos y análisis:**

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{125,4}{125,4 + 545}} = 0,432$$

La certidumbre de la prueba estadística (CPE) =  $1 - C = 0,57 \sim 57 \%$ .

**Interpretación:**

La prueba  $\chi^2$ , en la valoración de las cualidades de los candidatos, indicó que había diferencias significativas ( $p = 0,0001$ ), siendo el Coeficiente de Contingencia,  $C = 0,432$  que significa una mediana intensidad de asociación de atributos.

**Decisión:**

La asociación estadística de los atributos de los candidatos está en un nivel mediano y la certidumbre de la prueba es del 57 %.

**Ejercicio 108:**

En el experimento de Mendel de los caracteres forma y color de las semillas en F2, siendo la muestra  $n = 556$  y el valor  $\chi^2 = 0,470$ . ¿Cuál es el Coeficiente de contingencia y qué significado tiene?

**Datos y análisis:**

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{0,470}{0,470 + 556}} \rightarrow C = 0,029$$

**Interpretación:**

El Coeficiente de contingencia,  $C = 0,029$  lo cual indica que tiende a cero, por lo tanto, se justifica plenamente la decisión de aceptar  $H_0$ .

La confianza en la prueba estadística,  $CPE = 1 - C = 0,97 \sim 97 \%$

**Decisión:**

El Coeficiente de contingencia (C) indica una asociación estadística de baja intensidad. La certidumbre o confianza en la prueba estadística (CPE) es 97 %.

**5.3.3 COEFICIENTE DE CONCORDANCIA DE KAPPA DE COHEN**

El Coeficiente de Kappa de Cohen es una medida estadística que ajusta el efecto del azar a la proporción observada para elementos cualitativos (Manterola et al., 2018, p. 681). La variable de respuesta es calificada por evaluadores autorizados a tomar decisiones y generar acuerdos y desacuerdos. Si hay pleno acuerdo:  $K = 1$ , si no hubiera acuerdo:  $K = 0$ .

La fórmula para calcular el Coeficiente de concordancia Kappa de Cohen es:

$$K = \frac{AO - AE}{1 - AE} \quad [105]$$

Dónde:

$K$  = Coeficiente de Kappa de Cohen

AO = Acuerdos observados

AE = Acuerdos esperados

Las hipótesis estadísticas se formula como:

$H_0$ : Acuerdos observados = Acuerdos esperados

$H_1$ : Acuerdos observados  $\neq$  Acuerdos esperados

La regla de decisión, propuesta por Manterola et al. (2018, p. 681), es la siguiente:

< 0,01 No hay acuerdo

0,01 - 0,20 Ninguna a escaso acuerdo

0,21 - 0,40 Regular o razonable acuerdo

0,41 - 0,60 Moderado acuerdo

0,61 - 0,80 Substantial acuerdo

0,81 - 1,00 Acuerdo casi perfecto

### Ejercicio 109:

Un grupo de 50 estudiantes presentan una solicitud con sustento académico para acceder a una beca. Las propuestas son analizadas por los evaluadores a y b que califican con un "Sí" o un "No". Determinar si hay concordancia entre los dos evaluadores.

#### Datos y análisis:

En el Cuadro 78, se indica las decisiones de los evaluadores a y b.

Cuadro 78. Evaluación a 50 aspirantes a becas estudiantiles por dos evaluadores

		Evaluador b		Total
		Sí	No	
Evaluador a	Sí	20	5	25
	No	10	15	25
Total		30	20	50

#### Cálculos:

$$\text{Acuerdos observados: } AO = \frac{20 + 15}{50} = 0,70$$

$$\text{Acuerdos esperados: } AE = \frac{20 + 15}{50} = 0,70$$

$$aa_{si} = \frac{25}{50} = 0,50 \quad \rightarrow aa_{no} = 1,0 - 0,50 = 0,50$$

$$bb_{si} = \frac{30}{50} = 0,60 \quad \rightarrow bb_{no} = 1,0 - 0,60 = 0,40$$

$$(aa_{si})(bb_{si}) = 0,50(0,60) = 0,30$$

$$(aa_{no})(bb_{no}) = 0,50(0,40) = 0,20$$

$$\text{Acuerdos esperados: } AE = 0,30 + 0,20 = 0,50$$

$$K = \frac{AO - AE}{1 - AE} = \rightarrow \frac{0,70 - 0,50}{1 - 0,50} = \frac{0,20}{0,50} = 0,40 \quad \rightarrow K = 0,40$$

#### Interpretación:

Si  $K = 0,40$  varía de 0,21 a 0,40 la concordancia es regular.

#### Decisión:

La concordancia entre evaluadores es  $K = 0,40$  equivalente a regular.

5.3.4 PRUEBA CHI CUADRADA PARA PROPORCIONES

La prueba  $\chi^2$  también se usa para probar la homogeneidad de las proporciones, determinando si dos o más poblaciones tienen la misma distribución, como en los casos: hombres vs mujeres, antes vs después, vivos vs muertos, sanos vs enfermos. La variable de respuesta es categórica con dos o más clases. El procedimiento para este análisis  $\chi^2$  es similar al indicado para las pruebas de la independencia.

La hipótesis estadísticas se formulan como:

H<sub>0</sub>: Las proporciones son iguales

H<sub>1</sub>: Las proporciones son diferentes

La regla de decisión se plantea como:

Si  $\chi^2_{calculada} < \chi^2_{valor\ crítico}$ : Aceptar H<sub>0</sub>

Si  $\chi^2_{calculada} \geq \chi^2_{valor\ crítico}$ : Aceptar H<sub>1</sub>

**Ejercicio 110:**

Determinar si las proporciones de pulgones muertos, después de ser tratados con tres dosis de un insecticida (20 ppm, 40 ppm y 80 ppm), son estadísticamente iguales o hay diferencias estadísticas significativas. Para la prueba se tomaron tres muestras aleatorias de 100 pulgones (n = 300 pulgones) y se asignó los tratamientos al azar (dosis de insecticida).

**Datos y análisis:**

Las frecuencias observadas se organizan en una tabla de contingencia 3 x 2, en las columnas se distribuyen las tres clases, según las dosis del insecticida del ensayo: baja (20 ppm), media (40 ppm) y alta (60 ppm), en las hileras se organizan los datos de insectos vivos y muertos (variable experimental), además, en forma secuencial, se exponen las frecuencias esperadas y el correspondiente cálculo de los valores de Chi cuadrada (Cuadro 79).

Cuadro 79. Frecuencias observadas y esperadas de insectos vivos y muertos en función de las dosis de insecticida y cálculo de Chi cuadrada

Frecuencia Observada (O)	Dosis (ppm)			Total de Insectos
	20 (Baja)	40 (Media)	60 (Alta)	
Muertos	32	59	92	183
Vivos	68	41	8	117
Total	100	100	100	300
Frecuencia Esperada (E)	Dosis			Total de Insectos
	20 (Baja)	40 (Media)	60 (Alta)	
Muertos	61	61	61	183
Vivos	39	39	39	117
Total	100	100	100	300
Cálculo de Chi cuadrado	Dosis			$\Sigma$
	20 (Baja)	40 (Media)	60 (Alta)	
Muertos	$\frac{(32 - 61)^2}{61}$	$\frac{(59 - 61)^2}{61}$	$\frac{(92 - 61)^2}{61}$	29,6
Vivos	$\frac{(68 - 39)^2}{39}$	$\frac{(41 - 39)^2}{39}$	$\frac{(8 - 39)^2}{39}$	46,3
$\Sigma \chi^2$	35,35	0,17	40,40	75,91

Fuente: Di Rienzo et al. (2009, p. 268)

**Valores críticos de Chi cuadrado:**

$$\chi_{0,05}^2 = 5,99 \text{ y } \chi_{0,01}^2 = 9,21$$

$\chi_{calculado}^2 = 75,91^{**}$ . El valor de  $\chi_{calculado}^2$  es mayor que los valores críticos.

**Hipótesis estadísticas:**

H<sub>0</sub>: Las proporciones de insectos vivos y muertos son estadísticamente iguales

H<sub>1</sub>: Las proporciones de insectos vivos y muertos son estadísticamente diferentes

**Regla de decisión:**

Si  $\chi_{calculada}^2 < \chi_{valor\ crítico}^2$ : Aceptar H<sub>0</sub>

Si  $\chi_{calculada}^2 \geq \chi_{valor\ crítico}^2$ : Aceptar H<sub>1</sub>

**Interpretación:**

$\chi_{calculado}^2 = 75,91 > \chi_{0,01}^2 = 9,21$ : Aceptar H<sub>1</sub>. Las proporciones de insectos vivos y muertos son estadísticamente diferentes. Hay efecto de los insecticidas.

**Decisión:**

La proporción de insectos muertos depende de las aplicaciones de insecticida. Para comparar la dosis de insecticida se realiza una prueba *post hoc*.

### 5.3.5 COMPARACIÓN DE ATRIBUTOS CON CHI CUADRADA

La comparación de atributos con Chi cuadrada es utilizada para analizar datos discretos en diversas áreas, como la epidemiología, la sociología, la mercadotecnia, la investigación de mercado y el análisis de preferencias, entre ellas, el político. En el caso de haber asociación estadística significativa, se puede afirmar que hay relación entre los atributos analizados, mientras que una falta de asociación significativa indica que no hay evidencia suficiente para afirmar que hay una relación.

La prueba  $\chi^2$  se usa para analizar tablas de contingencia, posibilitando desglosar los atributos y compararlos, en base a las preferencias diferenciadas sobre los grupos o individuos de interés, aplicable a estudios exploratorios, descriptivos y relacionales.

**Ejercicio 111:**

En una encuesta a n = 545 ciudadanos, sobre sus preferencias electorales de cinco candidatos a una dignidad de elección popular, se procura identificar las cualidades más relevantes de cada uno de ellos con relación a los demás.

**Datos:**

En el Cuadro 80, se detalla los resultados de n = 545 encuestas de opinión sobre cuatro atributos de cinco candidatos a una dignidad de elección popular.

El objetivo es valorar la percepción de la ciudadanía con respecto a los atributos de los candidatos.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 80. Cualidades de los candidatos percibidas por la ciudadanía en una encuesta*

Frecuencias Observadas		Candidatos					Total
		Alva	Bernal	Castro	Díaz	Estrada	
Cualidades	Integridad	33	69	34	10	1	147
	Experiencia	14	12	8	5	1	40
	Comunicación	24	58	91	40	3	216
	Formación académica	11	19	41	62	9	142
Total de encuestados		82	158	174	117	14	545

Frecuencias Esperadas		Candidatos					Total
		Alva	Bernal	Castro	Díaz	Estrada	
Cualidades	Integridad	22	43	47	32	4	147
	Experiencia	6	12	13	9	1	40
	Comunicación	32	63	69	46	6	216
	Formación académica	21	41	45	30	4	142
Total de encuestados		82	158	174	117	14	545

Chi cuadrado		Candidatos					Total
		Alva	Bernal	Castro	Díaz	Estrada	
Cualidades	Integridad	5,4	<b>16,3</b>	3,6	<b>14,7</b>	2,0	42
	Experiencia	<b>10,6</b>	0,0	1,8	1,5	0,0	14
	Comunicación	2,2	0,3	<b>7,0</b>	0,9	1,2	12
	Formación académica	5,0	<b>11,9</b>	0,4	<b>32,6</b>	7,9	58
Aceptación		23,2	<b>28,6</b>	12,8	<b>49,7</b>	11,1	<b>125,4</b>

**Interpretación:**

El mejor candidato por el atributo integridad es Bernal (16,3) seguido de Díaz (14,7). Por formación académica, Díaz es más valorado (32,6) que Bernal (11,9). En general, Díaz tiene mayor aceptación (49,7) comparado con Bernal (28,6). En el candidato Alva se reconoce que tiene más experiencia (10,6) y Castro una mejor comunicación (7,0). Los candidatos Estrada y Castro no son reconocidos como buenos candidato, pues tienen valores  $\chi^2$  de 11,1 y 12,8, respectivamente, que son los más bajos.

**Decisión:**

El candidato Díaz tiene mayor formación académica y, en general, una mayor aceptación.

## CAPÍTULO VI: INVESTIGACIÓN EXPERIMENTAL

La experimentación es un nivel investigativo que trata del análisis de la causalidad, en busca de explicar las relaciones causa-efecto de los fenómenos biológicos, agrosociales y ambientales. En este nivel investigativo se manipula de forma planificada, no deliberada, los factores y niveles en estudio, con arreglo a los probables orígenes de la variación. Se puede estudiar una muestra o varias muestras simultáneamente, organizados en experimentos, donde se miden las variables de respuesta concretas, en función de los objetivos de la investigación.

Los investigadores, generalmente, desarrollan variedades de cultivos vegetales de altos rendimientos, prueban los efectos de los abonos sobre la producción o de los plaguicidas en el control de problemas fitosanitarios; determinan las óptimas densidades poblacionales, así como, de láminas y frecuencias de riego. En otras áreas del conocimiento, comparan dietas en la alimentación de los animales, eficiencia de los sistemas de producción, métodos de poscosecha y aseguramiento de la calidad e inocuidad de los alimentos, efecto de la presión de visitantes en escenarios turísticos, comparan modelos de gestión administrativa, valoran los impactos de las alternativas de las tecnologías de la información y comunicación o identifican métodos de mitigación de agua, aire y suelo contaminados.

Las variables de respuesta están en función de los objetivos de la investigación, y básicamente se orientan a maximizar los rendimientos, a minimizar los costos, a optimizar el uso de los insumos y recursos.

### 6.1 ORGANIZACIÓN DEL EXPERIMENTO

En la investigación experimental se debe cumplir ciertas condiciones para que los resultados tengan validez estadística y se puedan realizar las inferencias pertinentes, de otro modo se estaría incurriendo en prácticas no deseables que darían resultados no confiables que podrían significar un elevado costo para la sociedad.

#### 6.1.1 FACTORES Y NIVELES

Los aspectos fundamentales de la experimentación son: la identificación del factor (es) y niveles en estudio, el diseño de los tratamientos, el diseño experimental, las condiciones de la unidad experimental, las variables de respuesta y los procedimientos de análisis estadístico. El factor en estudio es la causa de las variaciones en las mediciones. ¿Cuál es el factor que determina la variación en la producción de cacao?, ¿Los genotipos tienen efecto en la producción?, ¿Los contenidos de nitrógeno en el suelo tienen efectos sobre la producción?, ¿La dotación de agua de riego al cultivo tiene efecto en la producción?

#### **Ejercicio 112:**

Diseñar los tratamientos para un experimento de cinco genotipos.

Si el factor A se refiere a genotipos en cinco niveles, entonces se describe:

A1 = variedad testigo, A2 = variedad nueva, A3 = variedad de la competencia, A4 = híbrido experimental y A5 = híbrido de la competencia.

La priorización de los factores genotipos [A] y densidades poblacionales [B], conlleva a reformular el diseño de los tratamientos (Cuadro 81).



**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 81. Factores y niveles para cultivares (A) y densidades poblacionales (B)*

Factores en estudio	Niveles				
	1	2	3	4	5
[A] Cultivares	Variedad Criolla [A1]	Variedad nueva [A2]	Variedad competencia [A3]	Híbrido experimental [A4]	Híbrido de competencia [A5]
[B] Densidades (plantas.ha <sup>-1</sup> )	40000 [B1]	50000 [B2]	60000 [B3]	--	--

En este momento de la planificación, hay dos alternativas: realizar experimentos unifactoriales o un experimento multifactorial (Briceño et al., 2021, p. 62). La decisión del número de experimentos y del grado de complejidad dependerá de la disponibilidad de recursos, de la priorización que realice el investigador y la experiencia en análisis de datos.

**Alternativa 1.-** Experimentos unifactoriales (Cuadro 82).

- Experimento 1.1 “Selección de variedades” [5 tratamientos A con *r* repeticiones]. En este caso se tiene que usar solo una densidad poblacional
- Experimento 1.2 “Comparación de tres densidades poblacionales” [3 tratamientos B con *r* repeticiones]. En este caso, tiene que identificar la variedad que se va a usar en el ensayo.

**Alternativa 2.-** Ensayo A x B = 5 x 3 = 15 tratamientos con *r* repeticiones (Cuadro 83).

*Cuadro 82. Diseño de los tratamientos en dos experimentos unifactoriales*

Ensayo: Selección de cultivares		
Tratamientos	Código	Descripción
1	[A1]	Variedad criolla (usada por los productores)
2	[A2]	Variedad experimental
3	[A3]	Variedad comercial de la competencia
4	[A4]	Híbrido experimental
5	[A5]	Híbrido comercial de la competencia
Ensayo: Comparación de tres densidades poblacionales		
Tratamientos	Código	Descripción
1	[B1]	Densidad baja: 40.000 plantas.ha <sup>-1</sup>
2	[B2]	Densidad media: 50.000 plantas.ha <sup>-1</sup>
3	[B3]	Densidad alta: 60.000 plantas.ha <sup>-1</sup>

*Cuadro 83. Diseño de los tratamientos en un ensayo bifactorial*

Tratamientos	Factor A	Factor B	Código	Descripción de los tratamientos
1	1	1	A1B1	Variedad criolla en densidad baja
2	1	2	A1B2	Variedad criolla en densidad media
3	1	3	A1B3	Variedad criolla en densidad alta
4	2	1	A2B1	Variedad experimental en densidad baja
5	2	2	A2B2	Variedad experimental en densidad media
6	2	3	A2B3	Variedad experimental en densidad alta
7	3	1	A3B1	Variedad comercial en densidad baja
8	3	2	A3B2	Variedad comercial en densidad media
9	3	3	A3B3	Variedad comercial en densidad alta
10	4	1	A4B1	Híbrido experimental en densidad baja
11	4	2	A4B2	Híbrido experimental en densidad media
12	4	3	A4B3	Híbrido experimental en densidad alta
13	5	1	A5B1	Híbrido comercial en densidad baja
14	5	2	A5B2	Híbrido comercial en densidad media
15	5	3	A5B3	Híbrido comercial en densidad alta

6.1.2 UNIDAD EXPERIMENTAL

En un muestreo, la unidad experimental (UE) es el elemento o conjunto de elementos donde se hacen mediciones para generar información de una o más variables. En un ensayo de campo, la UE está conformada por el área útil y un borde. Se aplica un tratamiento a la UE (parcela útil + borde). Las mediciones de las variables de interés se realizan solo en la parcela útil.

Las UE en ensayos de ciclo corto como: arroz, maíz y maní y hortalizas, se definen en áreas (m<sup>2</sup>) y en los cultivos anuales y perennes (plátano, papaya, pitahaya, cacao, café, mango, teca, balsa), se conforman por una planta o por un grupo de plantas organizadas en parcelas. A nivel de laboratorio, las UE lo integran una o varias cajas *petri*, tubos de ensayo o recipientes. En un vivero, pueden conformarlo una maceta o conjunto de macetas, bandejas o fundas con plantas. Las UE, en función de los objetivos, también pueden conformarse por muestras de suelo, de tejidos vegetales o de agua. Las UE con sus formas, tamaños, número de tratamientos y repeticiones determinan el tamaño del experimento (Figura 29).

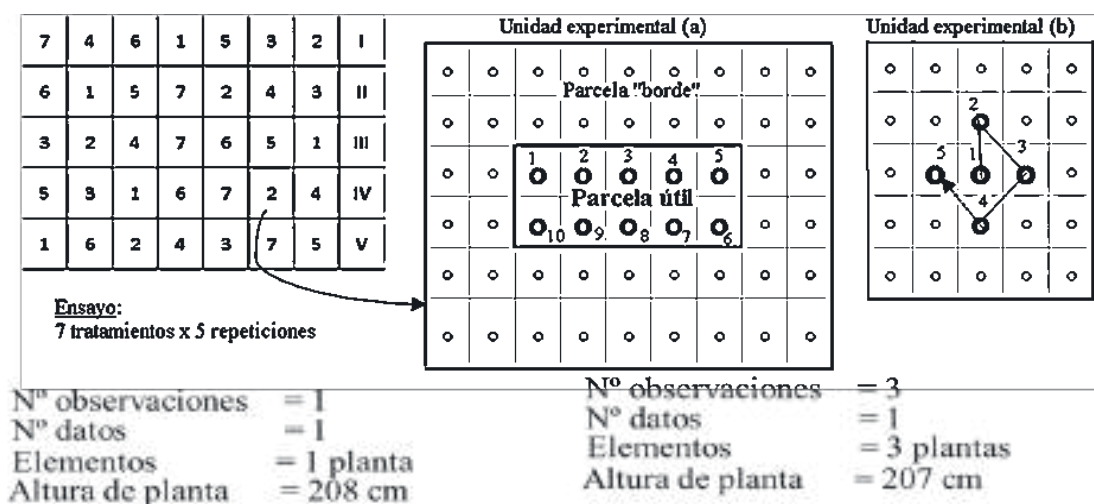


Figura 29. Distribución de las unidades experimentales y algunas formas de la parcela útil

Un aspecto fundamental de la planificación es la definición de los elementos de muestreo. Por ejemplo, si es de interés la variable "altura de planta" (AP), al medir solo en un elemento (planta) se genera un dato, supóngase AP = 208 cm; al medir la altura de tres plantas sucesivas y promediar se tiene un dato AP = 207 cm (Figura 30). Esta media representaría de mejor manera el dato puntual de AP, claro, pero implica un mayor tiempo en la recolección de datos.

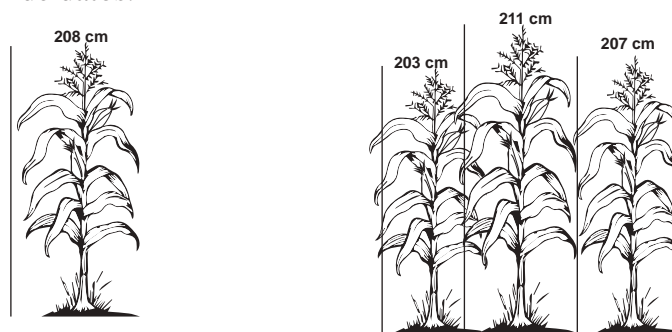


Figura 30. Elementos de muestreo y generación de datos

### 6.1.3 DISEÑOS BÁSICOS

El diseño de un experimento es el procedimiento de planificación de un estudio de causalidad, donde se presupone la identificación de las fuentes de las variaciones sobre una o varias características medibles (variables de respuesta). Una variable de respuesta se mide como efecto de un tratamiento y es el valor que interesa mejorarse. En un diseño experimental se manipulan deliberadamente los factores y niveles que constituyen las variables independientes las cuales provocan efectos sobre las variables dependientes (experimentales) a ser medidas, cuantificadas y analizadas. Los factores en estudio con sus niveles constituyen las fuentes de variación y pueden ser cualitativos o cuantitativos.

Un factor cualitativo se manifiesta en el origen de la variación, siendo de los tipos nominal u ordinal. Por ejemplo en un ensayo de fertilización edáfica, los niveles: testigo F0 (sin fertilización), fertilización orgánica (F1), fertilización química (F2) y fertilización combinada (F3), son variables nominales (cualitativas). Otro ejemplo. Si el factor en estudio es la abonadura con humus de lombriz, los niveles ordinales podrían ser las dosis: baja (B), media (M) y alta (A). Un factor cuantitativo es la indicación de una fuente de variación donde se prueban niveles en intervalo constante. Por ejemplo, en un ensayo sobre fertilización nitrogenada: (T0) testigo 0 kg N, (T1) 50 kg N, (T2) 100 kg N, (T3) 150 kg N y (T4) 200 kg N. En ensayos de producción pecuaria, la unidad experimental que no recibe ningún tratamiento, Condo y Pazmiño (2015, p. 43) lo identifican como un testigo satelital, el mismo que puede adaptarse a experimentos en los diseños DCA y BA.

En todo experimento hay factores controlables y no controlables. Los controlables son aquellos que se manipulan, cambian y adecuan a niveles de operación concretos como son las dosis de fertilizante nitrogenado. Los factores no controlables generan variaciones no pertinentes que se conoce como ruido en las respuestas. Entre estos factores se encuentran la luminosidad, velocidad del viento o altitud de la zona de cultivo. En todo experimento hay errores aleatorios que corresponden a la variabilidad observada no explicable a través de las fuentes de variación especificadas en el diseño. En la experimentación, la aleatorización en la distribución de los tratamientos es clave para reducir los errores y cumplir el supuesto de independencia.

Los diseños experimentales básicos son: diseño completamente aleatorio (DCA), diseño en bloques al azar (BA), diseño en cuadrado latino (CL), diseño en láctice (DL) y el diseño irrestricto al azar (DIA) (Cuadro 84). Para estudiar fertilizantes en distintos niveles se han propuesto diseños factoriales incompletos como el “San Cristóbal” y para estudios multifactoriales complejos se han desarrollado los métodos Taguchi.

*Cuadro 84. Diseños experimentales básicos y sus principales características*

Diseños	Repeticiones	Condición del campo	Restricciones
Diseño completamente al azar (DCA)	Con igual o desigual número de repeticiones	Gradiente en una dirección	No hay restricciones
Diseño en bloques al azar (BA)	Igual número de repeticiones	Gradiente en una dirección	No hay restricciones
Diseño en cuadrado latino (CL)	Igual número de repeticiones	Gradiente en dos direcciones	Hasta 10 tratamientos
Diseño en láctice (DL)	Igual número de repeticiones	Heterogéneo	Alto número de tratamientos
Diseños irrestrictos al azar (DIA)	Muestreos dentro de parcelas	Heterogéneo	Reducido número de tratamientos

La tendencia actual entre investigadores, para tomar decisiones estadísticas, es usar el valor  $p$ , en la perspectiva de indicar el verdadero nivel de confianza, apoyados en uso de programas estadísticos como: INFOSTAT<sup>21</sup>, STATGRAPHICS<sup>22</sup>, MSTAT<sup>23</sup>, NINITAB<sup>24</sup>, SPSS<sup>25</sup>, INFOGEN<sup>26</sup> y XLSTAT<sup>27</sup>.

#### 6.1.4 REGISTRO DE VARIABLES EXPERIMENTALES

Una variable de respuesta o experimental es la característica medible que depende de la manipulación de los factores en estudio y que explica su efecto. Por ejemplo: en el ensayo "Efecto de la fertilización nitrogenada sobre la producción de cacao", las variables experimentales serían: número de mazorcas/árbol/año, peso de la mazorca, índice de mazorca, peso de 100 semillas, producción en gramos/planta/año, producción en kilos/planta/año o rendimiento/hectárea.

Como variable predictiva se conoce a los niveles de un factor causal (variable independiente "X"), asumiendo que un cambio cuantitativo en ella genera cambios en la respuesta de la variable dependiente "Y". Por ejemplo: a mayor dosis de nitrógeno mayor respuesta productiva o a mayor incidencia de moniliasis menor cosecha de cacao. Los dos fenómenos descritos se pueden desarrollar en modelos gráficos o matemáticos, en procura de optimizar la variable de respuesta.

En la investigación agrícola, agrosocial, agroindustrial, agroforestal, biológica y ambiental se debe identificar el ámbito de estudio, las variables, las escalas de medición, las unidades de medida y los códigos pertinentes, si fuese necesario. En agronomía, si el ámbito de estudio es la morfología de la planta hay que recurrir a los descriptores morfológicos y a los protocolos de registros de datos. En estudios de sanidad vegetal hay que identificar los momentos de registro, estado fenológico del cultivo, estadio del ciclo de vida de la plaga, aplicando los métodos específicos para valorar la incidencia, severidad y otras variables e índices biológicos.

En socioeconomía amerita registrar las variables: ingresos familiares, costos unitarios, beneficio neto, rentabilidad, escolaridad, nivel de bienestar y necesidades básicas insatisfechas. En estudios ambientales hay variables de interés como: grado de contaminación física, química y biológica del suelo, grado de deforestación e intensidad de la erosión hídrica. En el ámbito alimentario interesan variables como: composición química, calidad nutritiva, calidad sensorial, incidencia de micotoxinas y residuos de agroquímicos. Para el registro de datos de caracterización morfológica hay que usar los descriptores varietales, como en los casos de café, arroz, fréjol, maíz y sorgo (Muñoz y Fernández, 1993).

#### 6.1.5 ANÁLISIS ESTADÍSTICO

El análisis estadístico es un conjunto de técnicas y métodos utilizados para recopilar, organizar, interpretar y presentar datos con el objetivo de extraer información significativa y tomar decisiones informadas. La estadística se ocupa de recopilar datos numéricos de variables observadas en poblaciones o muestras de interés, y el análisis estadístico proporciona las herramientas para comprender y describir esos datos.

---

<sup>21</sup> <https://www.infostat.com.ar/index.php?mod=page&id=46>

<sup>22</sup> <https://www.statgraphics.com/>

<sup>23</sup> <https://mstat.software.informer.com/5.4/>

<sup>24</sup> <https://www.minitab.com/es-mx/>

<sup>25</sup> <https://ibm-spss-statistics-base.uptodown.com/windows>

<sup>26</sup> <https://www.info-gen.com.ar/>

<sup>27</sup> <https://www.xlstat.com/es/>

El primer paso en el análisis de datos es el cálculo de los estadígrafos con énfasis en las medidas de tendencia central, de dispersión y de forma de la distribución. Un siguiente paso es el análisis de la varianza (ADEVA: abreviación en español), que es un procedimiento estadístico de partición de la varianza total en cada una de las fuentes causales de la variación, se complementa con las comparaciones *post hoc* o de separación de medias. Además, dependiendo del estudio, se realizan los análisis de correlaciones, regresiones, de maximización de la producción, de minimización de costos, de optimización de insumos, análisis multivariado u otros, según los objetivos.

## 6.2 ANÁLISIS DE VARIANZA PARAMÉTRICO

El análisis de la varianza (ADEVA) es una técnica estadística de contraste de hipótesis nula ( $H_0$ ) y la hipótesis alternativa ( $H_1$ ), de un conjunto de tratamientos en base de la prueba F, basada en la desagregación de la varianza total en cada una de las fuentes que lo originan (Little y Hills, 1987, p. 51). Para poder usar el ADEVA paramétrico, las series de datos deben cumplir las condiciones de normalidad, homocedasticidad e independencia.

En el ADEVA, las hipótesis estadísticas se formulan como:

$H_0$ : Las medias de los tratamientos son estadísticamente iguales.

$H_1$ : Al menos un tratamiento es estadísticamente diferente.

Al aceptar  $H_0$  no hay necesidad de comparar las medias, pues no hay evidencias de diferencias significativas ( $p > 0,05$ ). Aceptar  $H_1$  equivale a rechazar  $H_0$  ( $p < 0,05$ ), por lo tanto, surge la necesidad de separar las medias usando las pruebas *post hoc*.

Para calcular la varianza total de la muestra ( $S_X^2$ ), de una variable de interés, se usa la fórmula:

$$S_X^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}$$

Dónde:

$S_X^2$  = Varianza de la variable X (puede representarse con cualquier letra)

$\sum X^2$  = Sumatoria de los cuadrados de cada uno de los valores de X

$(\sum X)^2$  = Cuadrado de la suma total de X

$\frac{(\sum X)^2}{n}$  = Factor de corrección

$\sum X^2 - \frac{(\sum X)^2}{n}$  = Fórmula para calcular la suma de cuadrados

n = Número de observaciones total

n - 1 = Grados de libertad para la varianza muestra

En el análisis de la varianza se desagrega la varianza total en las distintas fuentes de variabilidad especificados en el esquema del ADEVA. El numerador en la fórmula de la varianza corresponde a la suma de cuadrados (SC), que tiene dos términos: sumatoria de cada observación elevado al cuadrado,  $\sum X^2$  (equivale al minuendo) y el cuadrado de la suma de las observaciones dividido para el número de observaciones,  $\frac{(\sum X)^2}{n}$  (equivale al sustraendo), que en términos del ADEVA se conoce como factor de corrección. La suma de cuadrados (SC) dividido para los grados de libertad (n - 1) da como resultado la varianza.

En el Gráfico 34, se expone, como ejemplo, el efecto de tres tratamientos sobre las distribuciones normales de una variable experimental. El T1 (testigo) tiene un promedio inferior a T2 y T3. El tratamiento T3 es numéricamente mayor que T2. Estos datos se pueden representar:  $T3 > T2 > T1$ . La prueba F, con un nivel de significación  $\alpha$ , posibilita comparar las medias de los tratamientos

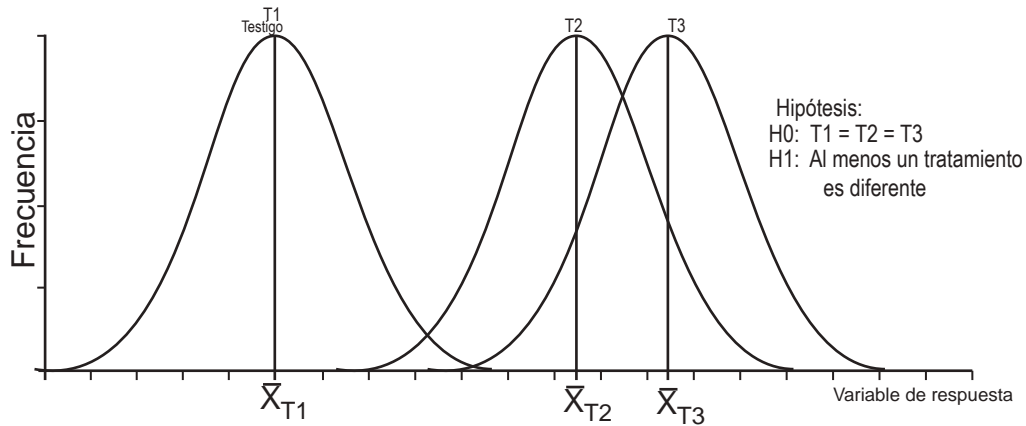


Gráfico 34. Comportamiento de una variable de respuesta como efecto de tres tratamientos

El ADEVA de un experimento está en función de las características del campo experimental y de la creatividad del investigador, siempre cumpliendo las condiciones y procedimientos del método científico. Los diseños experimentales más usados son: diseños completamente aleatorios (DCA), diseños en bloques al azar (BA), cuadrado latino (CL), látice (L), irrestrictos al azar (DIA) y arreglos factoriales.

En el ADEVA se exponen las fuentes de variación según el diseño aplicado, los grados de libertad, la suma de cuadrados, las varianzas o cuadrados medios, los valores de F calculada, los valores críticos de F tomados de la tabla y el valor p de la prueba. La significación estadística puede estar explícita, en una columna adicional o implícita en F calculada con las indicaciones: ns = no hay diferencia significativa ( $p > 0,05$ ), \* = hay diferencia estadística significativa ( $p < 0,05$ ) o \*\* = hay diferencia estadística altamente significativa ( $p < 0,01$ ). En función de estos resultados se decide aceptar o rechazar la hipótesis de nulidad ( $H_0$ ) con un nivel de significación concreto ( $\alpha$ ).

Al aceptar  $H_0$  se está decidiendo que no hay diferencia estadística significativa y al aceptar  $H_1$ , se afirma, con fundamento estadístico, que al menos un tratamiento, un bloque, una localidad, un ciclo de cultivo o cualquier otra fuente de variación resultó estadísticamente diferente. Al aceptar  $H_1$ , el análisis estadístico continúa hacia las pruebas *post hoc*, entre ellas: una prueba t, la diferencia mínima significativa (DMS) o las pruebas Duncan, Tukey, Newman-Keuls o comparaciones ortogonales, entre otras.

Complementariamente se calcula coeficiente de variación (CV) con la fórmula:

$$CV(\%) = \frac{\sqrt{CME}}{\bar{Y}} (100) \quad [106]$$

Dónde:

CV = Coeficiente de variación

CME = Cuadrado medio del error (varianza del error)

$\bar{Y}$  = Promedio de la variable en estudio



El CV denota la representatividad de la media general ( $\bar{Y}$ ) y el grado de distribución normalizada de los errores, siendo apropiado el valor más bajo, como indicativo de que la media es representativa y que la distribución se aproxima a la normalidad. Un experimento no debería tener CV máximos de 33 %. Cuando los CV son muy altos, se deduce que hay fuertes sesgos en la distribución y probablemente existen valores atípicos u “Outlier”.

Autores como Vásquez y Caballero (2011) no recomiendan emplear el coeficiente de variación como indicador de la variabilidad experimental debido a su inconsistencia (p. 43).

### 6.2.1 EXPERIMENTO EN DISEÑO COMPLETAMENTE AL AZAR

El diseño completamente al azar (DCA) se basa en la asignación aleatorizada de los tratamientos en las unidades experimentales (parcelas, individuos, grupos, jaulas, animales, árboles). El DCA se usa cuando los campos experimentales son homogéneos como los establos, mesones de laboratorio, invernaderos y viveros, cámara bioclimática o terrenos sin gradientes. Este tipo de diseño se utiliza en una variedad de experimentos, particularmente cuando hay interés en determinar el efecto promedio de los tratamientos, sin embargo, no es adecuado cuando hay gradientes o factores de ruido.

Al utilizar un DCA, los análisis estadísticos continúan con la separación de medias de los tratamientos y las inferencias sobre la población en general. Este diseño experimental se aplica en dos circunstancias: los tratamientos tienen igual número repeticiones o desigual número de repeticiones. Los orígenes de la variación en un DCA se atribuyen a los efectos de los tratamientos y al error experimental, según el modelo siguiente:

$$Y_{ij} = \mu + t_i + \epsilon_{ij} \quad i = 1,2,3,\dots,t, \quad j = 1,2,3,\dots,n$$

Dónde:

$Y_{ij}$  = Variable de respuesta en i-ésimo tratamiento y j-ésima observación

$\mu$  = Valor promedio del experimento

$t_i$  = Efecto del tratamiento

$\epsilon_{ij}$  = Error experimental

El esquema del análisis de varianza de un DCA se indica en el Cuadro 85.

Cuadro 85. Referencias para diseñar experimentos totalmente al azar

Fuentes de variación	Grados de libertad en DCA con igual número de repeticiones	Grados de libertad en DCA con desigual número de repeticiones
Tratamientos (t)	t - 1	t - 1
Error experimental	t(r - 1)	n - t
Total	r(t) - 1	n - 1

Un experimento en DCA, cuando cumple la aleatorización, tiene ventajas como: flexibilidad, sencillez del análisis y amplia aplicabilidad en ensayos unifactoriales o multifactoriales, aun cuando se encuentre desbalanceado (Corral, 2019, p. 214). Las desventajas del DCA se asocian a la pérdida de sensibilidad cuando el número de tratamientos es elevado, el número de repeticiones es reducido o resulta difícil mantener la uniformidad del campo experimental. Una medición es "sensible" cuando los cambios en una variable independiente generan cambios en las variables de respuesta.

6.2.1.1 EXPERIMENTO DCA DESBALANCEADO

El diseño DCA con desigual número de observaciones se aplica cuando los campos experimentales son homogéneos. Este es un diseño experimental con número de observaciones desbalanceado (desigual), ya sea porque se planeó de esa manera o a causa de las “parcelas perdidas”. En este último caso, significa que no hay un dato de un tratamiento en un bloque concreto a causa de un agente externos o por la limpieza de datos “atípicos”.<sup>28</sup>

El cálculo de las sumas de cuadrados comprende:

Factor de corrección:  $FC = \frac{(\sum X)^2}{n}$

Suma de cuadrados total:  $SCTotal = \sum X^2 - \frac{(\sum X)^2}{n} \rightarrow SC\ Total = \sum X_i^2 - FC$

Suma de cuadrados de los tratamientos:  $SC\ t = \frac{\sum t^2}{n_i} - FC$

Dónde:

$n$  = Número de observaciones total

$n_i$  = Número de observaciones por tratamiento.

Suma de cuadrados del error,  $SCE = SC\ Total - SC\ t$

**Ejercicio 113:**

Determinar el efecto de cuatro dietas para la cría y multiplicación de broca del café a nivel de laboratorio<sup>29</sup>. La variable experimental es el número de larvas.recipiente<sup>-1</sup>.

**Datos y análisis:**

En el cuadro 86, se exponen los datos del número de larvas de broca del café por recipiente, en un ensayo donde se probaron de cuatro dietas artificiales, donde el número de observaciones fue desbalanceado.

Cuadro 86. Larvas de broca del café por recipiente en función de cuatro dietas artificiales

Dietas en la alimentación de la broca del café				
<u>D1</u>	<u>D2</u>	<u>D3</u>	<u>D4</u>	
Número de larvas de broca por recipiente				
45	35	34	41	
46	33	34	41	
49		35	44	
44		34	43	
		33	41	
			42	
			44	
			41	
			41	
n1 = 4	n2 = 2	n3 = 5	n4 = 9	n = 20
$\Sigma D1 = 17$	$\Sigma D2 = 68$	$\Sigma D3 = 170$	$\Sigma D4 = 378$	$\Sigma X = 800$
$\bar{X}D1 = 46$	$\bar{X}D2 = 34$	$\bar{X}D3 = 34$	$\bar{X}D4 = 42$	$\bar{X} = 40$

<sup>28</sup> La limpieza de datos atípicos se realiza en base criterios de dispersión

<sup>29</sup> Información del Laboratorio de control biológico de la Estación Experimental Tropical Pichilingue, INIAP. 1989.



**Hipótesis:**

H<sub>0</sub>: La sobrevivencia de brocas son con las distintas dietas es estadísticamente igual

H<sub>1</sub>: Al menos una dieta es estadísticamente diferente

**Regla de decisión:**

Si F calculada < F crítico: Aceptar H<sub>0</sub>

Si F calculada ≥ F crítico: Aceptar H<sub>1</sub>

**Cálculo de las sumas de cuadrados:**

Factor de corrección:

$$FC = \frac{\Sigma(X)^2}{n} = \frac{(800)^2}{n} = 32.000$$

Suma de Cuadrados Total:

$$SC \text{ Total} = \Sigma X^2 - FC$$

$$SCT = (45)^2 + (46)^2 + (49)^2 + \dots + (41)^2 + (41)^2 - 32.000 = 464$$

Suma de Cuadrados de los Tratamientos:

$$SCt = \Sigma \left( \frac{t^2}{n_i} \right) - FC$$

$$SC \text{ t} = \frac{(184)^2}{4} + \frac{(68)^2}{2} + \frac{(170)^2}{5} + \frac{(378)^2}{9} - 32.000 = 432$$

Suma de Cuadrados del Error Experimental:

$$SCE = SC \text{ Total} - SC \text{ t} \rightarrow SCE = 464 - 432 = 32$$

En el Cuadro 87 se indica el resultado del ADEVA.

Cuadro 87. Resultado del ADEVA en un experimento en DCA desbalanceado

FV	GL	SC	CM	F <sub>calculada</sub>	F <sub>0,05</sub>	F <sub>0,01</sub>	p	Variación %
Dietas	t - 1 = 3	432	144	72**	3,24	5,29	0,0001	93,1
Error experimental	n - t = 16	32	2					6,9
Total	n - 1 = 19	464						100

Nota: \*\* hay diferencia altamente significativa entre tratamientos (p < 0,0001)

$$CV = \frac{\sqrt{CME}}{\bar{X}} (100) \rightarrow CV(\%) = \frac{\sqrt{2}}{40} (100) = 3,54 \%$$

La media general es representativa del ensayo y la variabilidad es reducida.

**Interpretación:**

Como F calculada > F<sub>0,01</sub>: Aceptar H<sub>1</sub>. Al menos una dieta es estadísticamente diferente con 99 % de confianza. La contribución de las dietas en la sobrevivencia de las brocas es 93,1 %. Y el CV(%) = 3,54 %.

**Decisión:**

De las dietas probadas, al menos una es estadísticamente diferente (p < 0,01).

6.2.1.2 EXPERIMENTO DCA BALANCEADO

Los experimentos DCA con igual número de observaciones por tratamiento se usan en terrenos uniformes, invernaderos, laboratorios, cámaras bioclimáticas y viveros.

**Cálculos de las sumas de cuadrados:**

Factor de corrección:  $FC = \frac{(\sum X)^2}{n}$

Suma de cuadrados total:  $SCTotal = \sum X^2 - \frac{(\sum X)^2}{n}$

SC Total =  $\sum Xi^2 - FC$

Suma de cuadrados de los tratamientos:

SC t =  $\frac{\sum t^2}{r} - FC$

Suma de cuadrados del error experimental:

SCE = SC Total – SC t

**Ejercicio 114:**

Realizar el ADEVA del ensayo “Efecto de las frecuencias de riego sobre el peso del plátano (kg.racimo<sup>-1</sup>): t1 = riego diario, t2 = riego cada dos días y t3 = riego cada tres días.

**Datos y análisis:**

Los datos del ensayo se indican en el Cuadro 88. El ensayo se caracteriza por tener: t = 3, n = r = 20, r(t) = 60.

Cuadro 88. Peso de los racimos en función de los tratamientos de frecuencias de riego

Tratamientos	Racimos y peso unitario en kg																				Σ	X̄
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
t1	35	36	38	34	39	42	45	37	35	37	42	46	40	39	39	34	39	33	37	45	772	38,6
t2	36	39	43	47	40	38	37	34	34	34	35	32	32	37	36	44	41	41	34	35	749	37,5
t3	29	29	26	33	24	26	26	25	25	23	21	26	29	29	23	28	30	29	27	26	534	26,7
																						34,25

**Hipótesis estadísticas:**

H<sub>0</sub>: Las medias de peso por efecto del riego son estadísticamente iguales

H<sub>1</sub>: Las medias de peso por efecto del riego son estadísticamente diferentes

**Regla de decisión:**

Si F calculada < F crítico: Aceptar H<sub>0</sub>

Si F calculada ≥ F crítico: Aceptar H<sub>1</sub>

**Cálculo de las sumas de cuadrados:**

Factor de corrección:

$FC = \frac{(\sum X)^2}{rt} = (2055)^2/60 = 70.384$

Suma de cuadrados total:

$SCT = \sum X^2 - FC = (35)^2 + (36)^2 + \dots + (26)^2 - 70.386 = 2477$

Suma de cuadrados de los tratamientos:

$$SC\ t = \frac{\sum t^2}{r} - FC = \frac{(772)^2 + (749)^2 + (534)^2}{20} - 70.386 = 2208$$

Suma de cuadrados del error experimenta:

$$SCE = SCT - SC\ t = 2477 - 2.208 = 269$$

Los resultados del ADEVA se indican en el Cuadro 89.

Cuadro 89. ADEVA del ensayo de frecuencias de riego

Fuente de variación	GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p de F	Variación (%)
Frecuencias de riego	2	2208	1103,98	233,7**	3,16	5,00	0,000001	89,1
Error experimental	57	269	4,72					10,9
Total	59	2477	41,99					100,0

Nota: \*\* = hay diferencia estadística altamente significativa para las frecuencias de riego (p < 0,01)

La proporción de la variación correspondiente a cada fuente se calcula relacionando las SC parciales/SC Total. La variación en frecuencias de riego del 89,1 % indica que la respuesta productiva se debe prevalentemente a este factor.

$$CV(\%) = \frac{\sqrt{CME}}{Media} (100) = \frac{\sqrt{4,72}}{34,25} (100) = 6,14 \%$$

El CV(%) = 6,14 % indica que la media es representativa del ensayo y que la dispersión de los datos es adecuada para cultivos anuales.

**Interpretación:**

F calculada > F<sub>0,01</sub>: Se acepta H<sub>1</sub> (p < 0,01).

**Decisión:**

Las frecuencias de riego influyen significativamente sobre el peso de racimos de plátano (p < 0,01).

**Ejercicio 115:**

Determinar si la ganancia de peso (GP en libras) en ovinos de una raza Z y de igual edad, depende de las dietas aplicadas durante 100 días (Little y Hills, 1987, p. 54).

**Datos y análisis:**

La ganancia de peso en función de las dietas en ovinos se indica en el Cuadro 90.

Cuadro 90. Alternativas de alimentación de las ovejas y ganancias de peso

	Alternativas de alimentación de ovejas				
	t1	t2	t3	t4	
	Dieta convencional	Dieta A	Dieta B	Dieta C	
	47	50	57	54	
	52	54	53	65	
	62	67	69	74	
	51	57	57	59	
∑ Tratamientos	212	228	236	252	928
$\bar{X}$ GP en libras	53	57	59	63	58

Nota: t = 4, r = 4, rt = 16

**Hipótesis estadísticas:**

H<sub>0</sub>: t<sub>1</sub> = t<sub>2</sub> = t<sub>3</sub> = t<sub>4</sub>. Las dietas no influyen en la ganancia de peso

H<sub>1</sub>: Al menos una dieta probada es estadísticamente diferente

**Regla de decisión:**

Si F calculada < F crítica: Aceptar H<sub>0</sub>

Si F calculada ≥ F crítica: Aceptar H<sub>1</sub>

**Cálculos de las sumas de cuadrados:**

Factor de corrección:

$$FC = \frac{(\sum X)^2}{rt} = \frac{(928)^2}{16} = 53.824$$

Suma de cuadrados total:

$$SCT = \sum X^2 - FC = (47)^2 + (52)^2 + \dots + (59)^2 - 53.824 = 854$$

Suma de cuadrados de los tratamientos:

$$SC t = \frac{\sum t^2}{r} - FC = \frac{212^2 + 228^2 + 236^2 + 252^2}{16} - 53.824 = 208$$

Suma de cuadrados del error experimental:

$$SCE = SCT - SC t = 854 - 208 = 646$$

En el Cuadro 91 se expone el ADEVA del experimento.

Cuadro 91. Resultados del ADEVA en ensayo DCA balanceado

Fuente de variación	GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p	Variación (%)
Tratamientos	3	208	69,3	1,29 ns	3,49	5,95	0,323	24,4
Error experimental	12	646	53,8					75,6
Total	15	854						100

Nota: ns = No hay diferencia estadística significativa entre tratamientos

$$CV(\%) = \frac{\sqrt{CME}}{\bar{y}} (100) = \frac{\sqrt{53,8}}{58} (100) = 12,65 \%$$

El CV tiende a ser bajo, por tanto, la media es representativa del experimento. La contribución de la variación de los tratamientos a la variación total es 24,4 % y del error es 75,6 %. Cabe revisar el cumplimiento de los supuestos que exige la estadística paramétrica.

**Interpretación:**

Sobre los tratamientos: F calculada < F<sub>0,05</sub>: Se acepta H<sub>0</sub>. Las dietas no influyen sobre la ganancia de peso en los ovinos. ¿Hay confianza en la afirmación: t<sub>4</sub> = 63 libras es estadísticamente igual a t<sub>1</sub> = 53 libras?, ¿10 libras de diferencia entre t<sub>4</sub> y t<sub>1</sub> no es significativa? (se estará incurriendo en el error tipo I).

**Decisión:**

Las medias de los tratamientos son estadísticamente iguales. No hace falta comparar las medias. Si tuviera duda (incertidumbre) del resultado: t<sub>4</sub> = t<sub>1</sub>, podría realizar el análisis de varianza por rangos de Kruskal y Wallis.

6.2.2 EXPERIMENTO CON DISEÑO EN BLOQUES AL AZAR

Se usa cuando en el campo experimental se constata una gradiente en una dirección, distribuyéndose en forma perpendicular a la gradiente y los tratamientos se asignan aleatoriamente. Los bloques con gradiente en una dirección pueden tratarse de una ligera pendiente (Figura 31), del lindero con un camino o un borde de árboles.

**Experimento de 5 tratamientos y 4 bloques**

<b>I</b>	<b>2</b>	<b>1</b>	<b>3</b>	<b>5</b>	<b>4</b>
<b>II</b>	<b>5</b>	<b>3</b>	<b>4</b>	<b>2</b>	<b>1</b>
<b>III</b>	<b>1</b>	<b>4</b>	<b>5</b>	<b>3</b>	<b>2</b>
<b>IV</b>	<b>3</b>	<b>2</b>	<b>1</b>	<b>4</b>	<b>5</b>

Figura 31. Distribución de unidades experimentales en un diseño de Bloques al azar

Un objetivo del diseño en bloques al azar (BA) es mantener la variabilidad tan pequeña como sea posible, entre las unidades experimentales que están dentro del bloque y maximizar las diferencias entre bloques. Si no hubiera diferencias significativas entre bloques entonces el diseño no contribuye a detectar las diferencias entre tratamientos.

El modelo lineal del diseño en bloques al azar, para el estudio de un factor causal, es:

$$Y = \mu + B_i + t_j + \epsilon_{ij}$$

$$i = 1, 2, \dots t \quad j = 1, 2, \dots B$$

Dónde:

Y = Valor de la variable experimental

$\mu$  = Media general

$B_j$  = Efecto del j-ésimo bloque

$t_i$  = Efecto del i-ésimo tratamiento

$\epsilon_{ij}$  = Errores normales e independientes

Cálculo del “factor de corrección”:

$$FC = \frac{(\sum X)^2}{r(Tt)}$$

Suma de cuadrados total:

$$SC T = \sum X^2 - \frac{(\sum X)^2}{n}$$

$$SC T = \sum X_i^2 - FC$$

Suma de cuadrados de los tratamientos:

$$SC t = \frac{\sum T t^2}{n_i} - FC$$

Suma de cuadrados de los bloques o repeticiones:

$$SCR = \frac{\Sigma Repeticiones^2}{Tratamientos} - FC$$

Suma de cuadrados del error experimental:

$$SCE = SCT - SC_t - SCR$$

### Ejercicio 116:

Realizar el ADEVA del rendimiento de forraje seco (heno), en kg.parcela<sup>-1</sup>, en ensayo de cuatro variedades de sorgo con cinco repeticiones (Reyes, 1980, p. 135).

#### Datos y análisis:

En el Cuadro 92, se indican los resultados de producción por parcela de un ensayo de 4 variedades de sorgo en 5 repeticiones.

Cuadro 92. Producción por parcela en 4 variedades de sorgo en experimento en bloques al azar

Variedades (V)	Bloques					$\Sigma V$	Media V (kg.parcela <sup>-1</sup> )
	I	II	III	IV	V		
Honey	32,30	34,00	34,30	35,00	36,50	172,10	34,42
Milk Maker	33,30	33,00	36,30	36,80	34,50	173,90	34,78
Silo King	30,80	34,30	35,30	32,30	35,80	168,50	33,70
Beef Builder	29,30	26,00	29,80	28,00	28,80	141,90	28,38
$\Sigma$	125,70	127,30	135,70	132,10	135,6	656,40	--
$\bar{X}$	31,43	31,83	33,93	33,03	33,9		32,82

Nota: t = 4, r = 5, r(t) = 20

#### Hipótesis estadísticas sobre los bloques:

H<sub>0</sub>: No hay efecto de los bloques sobre el rendimiento de heno

H<sub>1</sub>: Al menos en un bloque hay diferencia significativa en el rendimiento de heno

#### Hipótesis estadísticas sobre los tratamientos:

H<sub>0</sub>: No hay diferencia en los rendimientos de heno, entre las variedades de sorgo

H<sub>1</sub>: Al menos una variedad de sorgo es estadísticamente diferente

#### Regla de decisión:

Si F calculada < F crítica: Aceptar H<sub>0</sub>

Si F calculada ≥ F crítica: Aceptar H<sub>1</sub>

#### Cálculos de las sumas de cuadrados:

Factor de corrección:

$$FC = \frac{(\Sigma X)^2}{r(t)} = \frac{(656)^2}{20} = 21.543$$

Suma de cuadrados total:

$$SCT = \Sigma X^2 - FC = (32,30)^2 + (33,30)^2 + \dots + (28,80)^2 - 21.543 = 182,17$$

Suma de cuadrados de los tratamientos:

$$SC\ t = \frac{\sum t^2}{r} - FC = \frac{(172,1)^2 + (173,9)^2 + (168,5)^2 + (141,9)^2}{5} - 21.543 = 134,45$$

Suma de cuadrados de los bloques:

$$SC\ R = \frac{\sum R^2}{t} - FC = \frac{(125,7)^2 + (127,3)^2 + (135,7)^2 + (132,1)^2 + (135,6)^2}{4} - 21.543 = 21,46$$

Suma de cuadrados del error experimental:

$$SCE = SCT - SC\ t - SCR = 182,17 - 134,45 - 21,46 = 26,26$$

El análisis de varianza se expone en el Cuadro 93.

Cuadro 93. ADEVA del experimento de variedades de sorgo forrajero

Fuente de variación	GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p	Variación (%)
Bloques	4	21,46	5,40	2,45 ns	3,26	5,41	0,1030	11,8
Tratamientos	3	134,45	44,80	20,48 **	3,49	5,95	0,00005	73,8
Error experimental	12	26,26	2,20					14,4
Total	19	182,17						100

Nota: ns = no hay diferencia estadística significativa, \*\* = hay diferencia estadística altamente significativa

$$CV(\%) = \frac{\sqrt{CME}}{\bar{y}} (100) = \frac{\sqrt{2,20}}{32,82} (100) = 4,51 \%$$

La media general es representativa del ensayo y la dispersión tiende a ser reducida. Los tratamientos (variedades) explican el 73,8 % de la variación total.

### Interpretación:

Sobre los bloques.-  $F_{calculada} < F_{0,05}$ : Se acepta  $H_0$ . No hubo efecto significativo de los bloques en los rendimientos de heno.

Sobre los tratamientos.-  $F_{calculada} > F_{0,01}$ : Se acepta  $H_1$ . Al menos una de las variedades de sorgo tiene diferencia estadística altamente significativa en el rendimiento de heno.

### Decisión:

No hubo efecto de los bloques sobre el rendimiento de heno. Al menos una de las cuatro variedades de sorgo es estadísticamente diferente ( $p < 0,01$ ). Se debe realizar las pruebas *post hoc* para la separación de medias.

## 6.2.3 EXPERIMENTO CON DISEÑO EN CUADRADO LATINO

Un cuadrado latino (CL) es una matriz de tamaño  $t^2$  conformado por C columnas y H hileras, donde los tratamientos (t) se distribuyen aleatoriamente y deben estar presentes en todas las columnas y en todas las hileras: Columnas C = hileras H = tratamientos t.

El diseño CL es muy eficiente con un máximo de 10 tratamientos, cuando hay evidencia de gradientes en dos direcciones (Figura 32). La desventaja de este diseño es la reducción de los grados de libertad del error y el análisis tiende a ser más complejo que los diseños DCA y BA. Como restricción del modelo está la condición de que todos los tratamientos deben aparecer en cada columna y en cada hilera.

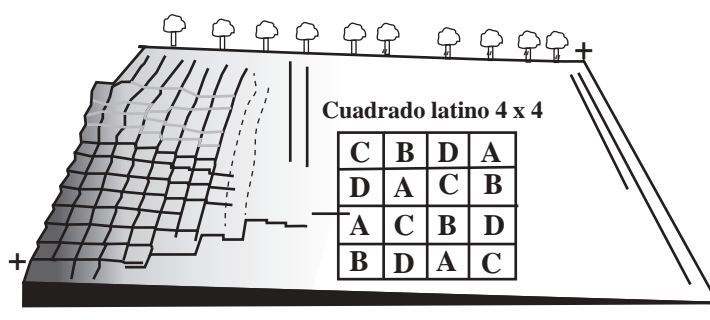


Figura 32. Diagrama de un experimento en cuadrado latino 4 x 4

El esquema del análisis de la varianza es el siguiente:

Fuente de variación	Grados de libertad
Columnas	$t - 1$
Hileras	$t - 1$
Tratamientos	$t - 1$
Error experimental	$(t - 1)(t - 2)$
Total	$t^2 - 1$

El modelo estadístico es el siguiente:

$$Y_{ijk} = \mu + C_j + H_i + t_k + \varepsilon_{ikj}$$

$i = 1, 2, 3...t \quad j = 1, 2, 3...t \quad k = 1, 2, 3...t \quad C = H = t$

Dónde:

$Y_{ijk}$  = Variable de respuesta en la  $ijk$ -ésima unidad experimental

$\mu$  = Media general de la variable de respuesta

$C_j$  = Efecto de la  $j$ -ésima columna

$H_i$  = Efecto de la  $i$ -ésima hilera

$t_k$  = Efecto de los  $k$ -ésimo tratamientos

$\varepsilon_{ikj}$  = Error experimental asociado a la  $ijk$ -ésima unidad experimental

**Cálculos de las sumas de cuadrados:**

Factor de corrección:

$$FC = \frac{\sum(X)^2}{t^2}$$

Suma de Cuadrados Total:

$$SCT = \sum Xi^2 - FC$$

Suma de cuadrados de las columnas (C): Representan a la gradiente en una dirección.



$$SC \text{ Columnas} = \frac{\sum \text{Columnas}^2}{t} - FC$$

Suma de cuadrados de las hileras (H): Corresponde a la gradiente en la otra dirección.

$$SC \text{ Hileras} = \frac{\sum \text{Hileras}^2}{t} - FC$$

Suma de cuadrados de los tratamientos:

$$SC \text{ t} = \frac{\sum t^2}{c} - FC$$

Suma de Cuadrados del Error:

$$SCE = SC \text{ Total} - SC \text{ Columnas} - SC \text{ Hileras} - SC \text{ Tratamientos.}$$

**Ejercicio 117:**

En gallinas ponedoras, se probó el efecto de cuatro dietas A, B, C y T en un diseño CL 4<sup>2</sup>, asignando al azar cuatro jaulas con 10 ponedoras de la misma edad y origen genético. Las ponedoras de cada jaula fueron alimentadas durante cuatro períodos de 24 días con cada una de las dietas. Se registraron los datos de peso de los huevos en los últimos 12 días de cada período, para eliminar cualquier efecto remanente de la dieta anterior (Corral, 2019, p. 258).

En el Cuadro 94 se indica las características del experimento en cuadrado latino de cuatro dietas a gallinas ponedoras (A, B, C y T), en 4 jaulas (columnas) y 4 períodos (hileras) donde se mide el peso de los huevos

Cuadro 94. Distribución de las jaulas y tratamientos en cuatro períodos de estudio

		Columnas: <u>Jaulas</u>			
		I	II	III	IV
Hileras: <u>Períodos</u>	1	T	C	B	A
	2	C	B	A	T
	3	B	A	T	C
	4	A	T	C	B

**Datos y análisis:**

Los resultados del ensayo, promediando el peso de los huevos de las columnas (jaulas) e hileras (períodos), se exponen en el Cuadro 95. Los promedios de las dietas (tratamientos) y de las jaulas (columnas) se indica en el Cuadro 96.

Cuadro 95. Pesos de los huevos y cálculo de sumas y medias de columnas (jaulas) e hileras (períodos)

		Columnas: <u>Jaulas</u>				Σ períodos (g)	X̄ períodos (g)
		I	II	III	IV		
Hileras: <u>Períodos</u>	1	T = 47,7	C = 41,9	B = 58,3	A = 50,5	198,4	49,6
	2	C = 44,6	B = 60,2	A = 54,1	T = 52,2	211,1	52,8
	3	B = 57,3	A = 59,7	T = 51,1	C = 47,2	215,3	53,8
	4	A = 55,2	T = 49,8	C = 45,3	B = 61,4	211,7	52,9
Σ jaulas (g)		204,8	211,6	208,8	211,3	836,5	--
X̄ jaulas (g)		51,2	52,9	52,2	52,8	--	52,3

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

Cuadro 96. Datos de los pesos de los huevos y cálculo de sumas y promedios de dietas y jaulas

		Columnas: <u>Jaulas</u>				Σ dietas (g)	$\bar{X}$ dietas (g)
		I	II	III	IV		
Tratamientos: <u>Dietas</u>	A	55,2	59,7	54,1	50,5	219,5	54,9
	B	57,3	60,2	58,3	61,4	237,2	59,3
	C	44,6	41,9	45,3	47,2	179,0	44,8
	T	47,7	49,8	51,1	52,2	200,8	50,2
Σ jaulas (g)		204,8	211,6	208,8	211,3	836,5	
$\bar{X}$ jaulas (g)		51,2	52,9	52,2	52,8		52,3

**Hipótesis estadísticas sobre las columnas (jaulas):**

H<sub>0</sub>: Los pesos de los huevos en función de las jaulas resultaron iguales

H<sub>1</sub>: Al menos una de las jaulas tiene diferencia estadística significativa

**Hipótesis estadísticas sobre las hileras (períodos):**

H<sub>0</sub>: Los pesos de los huevos en función de los períodos de estudio resultaron iguales

H<sub>1</sub>: Al menos uno de los períodos tiene diferencia estadística significativa

**Hipótesis estadísticas sobre los tratamientos:**

H<sub>0</sub>: Los pesos de los huevos con las cuatro dietas resultaron estadísticamente iguales

H<sub>1</sub>: Al menos una de las cuatro dietas probadas tiene diferencia estadística significativa

**Regla de decisión:**

Si  $F_{calculada} < F_{crítica}$ : Aceptar H<sub>0</sub>

Si  $F_{calculada} \geq F_{crítica}$ : Aceptar H<sub>1</sub>

**Cálculos de las sumas de cuadrados:**

Factor de corrección:

$$FC = \frac{(\sum X)^2}{t^2} = \frac{(836)^2}{16} = 43,73$$

Suma de cuadrados total:

$$SCT = \sum X^2 - FC = (47,7)^2 + (41,9)^2 + \dots + (61,4)^2 - 43,73 = 547,2$$

Suma de cuadrados de las columnas (jaulas):

$$SC C = \frac{\sum \text{Columnas}^2}{t} - FC \rightarrow SC C = \frac{(204,8)^2 + (211,6)^2 + (208,8)^2 + (211,3)^2}{4} - 43,73 = 7,42$$

Suma de cuadrados de las hileras (períodos):

$$SC H = \frac{\sum \text{Hileras}^2}{t} - FC \rightarrow SC H = \frac{(198,4)^2 + (211,1)^2 + (215,3)^2 + (211,7)^2}{4} - 43,73 = 40,9$$

Suma de cuadrados de los tratamientos (dietas):

$$SC t = \frac{\sum t^2}{c} - FC \rightarrow \frac{(219,5)^2 + (237,2)^2 + (179,0)^2 + (200,8)^2}{4} - 43,73 = 468,2$$

Suma de cuadrados del error experimental:

$$SCE = SCT - SC C - SC H - SC t = 547,2 - 7,42 - 40,9 - 468,2 = 30,7$$

El ADEVA del ensayo en CL se indica en el Cuadro 97

Cuadro 97. ADEVA del diseño en cuadrado latino

Fuente de variación		GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p	Variación (%)
Jaulas (Columnas)	t - 1	3	7,4	2,47	0,48 ns	4,76	9,78	0,706	1,4
Períodos (Hileras)	t - 1	3	40,9	13,6	2,67 ns	4,76	9,78	0,141	7,5
Dietas (Tratamientos)	t - 1	3	468,2	156,1	30,52 **	4,76	9,78	0,0001	85,6
Error experimental	(t - 1)(t - 2)	6	30,7	5,1					5,6
Total	t <sup>2</sup> - 1	15	547						100

Nota: ns = no hay diferencia estadística significativa, \*\* = hay diferencia altamente significativa

$$CV(\%) = \frac{\sqrt{CME}}{Media} (100) = \frac{\sqrt{5,1}}{52,3} (100) = 4,3 \%$$

La media general es representativa del ensayo y la variación entre unidades experimentales es reducida. La contribución de las dietas a la variación total es de 85,6 %.

**Interpretación:**

Sobre las jaulas:

F calculada < F<sub>0,05</sub>: Se acepta H<sub>0</sub>. No hay efecto de las jaulas sobre la ganancia de peso.

Sobre los períodos:

F calculada < F<sub>0,05</sub>: Se acepta H<sub>0</sub>. No hay efecto de los períodos sobre la ganancia de peso.

Sobre las dietas:

F calculada > F<sub>0,01</sub>: Se acepta H<sub>1</sub>. Los pesos promedios de los huevos, en al menos una de las cuatro dietas probadas, resultó estadísticamente diferente.

**Decisión:**

Hay efectos de las dietas sobre la ganancia de peso de los huevos en gallinas ponedoras (p < 0,0001). Cabe realizar una prueba *post hoc* para comparar las medias.

6.2.4 EXPERIMENTO CON DISEÑO EN LÁTICE

El diseño en látice (LT) forma un grupo de diseños conocidos como “bloques incompletos” dentro de las repeticiones. Este diseño resulta más eficiente comparado con los otros cuando hay alta heterogeneidad del campo experimental y alto número de tratamientos. El diseño LT desagrega la varianza total en una fuente adicional que corresponde a los bloques incompletos, conceptualmente diferente al de repeticiones.

En un campo heterogéneo no se logra homogeneidad entre las unidades experimentales, por lo tanto, al dividir las repeticiones en bloques incompletos se reduce la variabilidad dentro de cada unidad experimental. Este diseño tiene especial importancia en mejora genética y en ensayos de mezclas de ingredientes activos. La restricción del LT es que el número de tratamientos debe ser un cuadrado perfecto.

El modelo estadístico de diseño LT es el siguiente:

$$Y_{ijk} = \mu + B_{j(k)} + t_{j(k)} + \epsilon_{ijk}$$

Dónde:

$Y_{ijk}$  = Observación cualquiera

$\mu$  = Valor de la media general

$B_{j(k)}$  = Efecto del j-ésimo bloque alojado en la k-ésima repetición

$i$  = Efecto del i-ésimo tratamiento ( $i = 1, \dots, t$ )

$\epsilon_{ijk}$  = Efecto del error experimental relacionado al i-ésimo tratamiento y k-ésimo bloque alojado en la j-ésima repetición.

***Esquema del análisis de varianza:***

Fuente de variación	Grados de libertad
Repeticiones (R)	$R - 1$
Tratamientos (t)	$t - 1$
Bloques incompletos (BI)	$2(BI - 1)$
Error experimental	$[R(t) - 1] - (R - 1) - (t - 1) - 2(BI - 1)$
Total	$R(t - 1)$

***Cálculos de las sumas de cuadrados:***

Factor de corrección:  $FC = \frac{(\sum X)^2}{R(t)}$

Suma de Cuadrados Total:  $SC T = \sum X^2 - FC$

SC Repeticiones:  $SC R = \frac{\sum R^2}{t} - FC$

SC Bloques incompletos:  $SC BI = \frac{\sum BI^2}{t} - FC$

SC Tratamientos:  $SC t = \frac{\sum t^2}{R}$

Suma de Cuadrados del Error:  $SCE = SC Total - SC R - SC BI - SC t$

**Ejercicio 118:**

Realizar el ADEVA de un ensayo con diseño en Látice, donde se prueba 25 variedades de maíz (V), midiendo la variable rendimiento de grano en  $kg.parcela^{-1}$  (Lagos y Criollo, 2001).

***Datos y análisis:***

Los resultados del experimento del diseño en látice LT, midiendo el rendimiento en  $kg.parcela^{-1}$ , se presentan en el Cuadro 98. Los cinco bloques incompletos están aleatorizados, en las repeticiones I y II. Dentro de cada bloque se distribuyen al azar cinco parcelas y cada una de ellas contiene una variedad experimental para calcular la suma y media de los rendimientos por bloque incompleto.

En el Cuadro 99, se indican los promedios por variedad y repetición, las sumas, medias y las varianzas, además, se incluye el valor escalar al máximo (EAM). En el Cuadro 100, se resume el rendimiento, de las variedades por bloques incompletos, alojados en las repeticiones.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 98. Promedios de rendimiento (kg.parcela<sup>-1</sup>) de 25 variedades de maíz en diseño LT 5 x 5*

Bloque incompleto		Repetición I					Σ BI	Media BI (kg.parcela <sup>-1</sup> )
3	Parcela	1	2	3	4	5		
	Variedades	11	12	14	15	13		
	<i>Rendimiento</i>	<i>1120</i>	<i>361</i>	<i>355</i>	<i>129</i>	<i>1019</i>	<i>2984</i>	<i>597</i>
5	Parcela	10	9	8	7	6		
	Variedades	21	25	22	24	23		
	<i>Rendimiento</i>	<i>154</i>	<i>604</i>	<i>166</i>	<i>75</i>	<i>377</i>	<i>1375</i>	<i>275</i>
4	Parcela	11	12	13	14	15		
	Variedades	18	19	17	20	16		
	<i>Rendimiento</i>	<i>501</i>	<i>629</i>	<i>500</i>	<i>198</i>	<i>128</i>	<i>1955</i>	<i>391</i>
1	Parcela	20	19	18	17	16		
	Variedades	5	1	4	2	3		
	<i>Rendimiento</i>	<i>1452</i>	<i>530</i>	<i>789</i>	<i>103</i>	<i>689</i>	<i>3563</i>	<i>713</i>
2	Parcela	21	22	23	24	25		
	Variedades	8	10	6	7	9		
	<i>Rendimiento</i>	<i>1157</i>	<i>1078</i>	<i>643</i>	<i>965</i>	<i>1304</i>	<i>5146</i>	<i>1029</i>
						15023	601	
Bloque incompleto		Repetición II					Σ BI	Media BI (kg.parcela <sup>-1</sup> )
5	Parcela	30	29	28	27	26		
	Variedades	10	20	15	5	25		
	<i>Rendimiento</i>	<i>747</i>	<i>187</i>	<i>407</i>	<i>348</i>	<i>265</i>	<i>1954</i>	<i>391</i>
2	Parcela	31	32	33	34	35		
	Variedades	17	2	7	22	12		
	<i>Rendimiento</i>	<i>375</i>	<i>458,2</i>	<i>407</i>	<i>223</i>	<i>118</i>	<i>1581</i>	<i>316</i>
3	Parcela	40	39	38	37	36		
	Variedades	18	3	13	8	23		
	<i>Rendimiento</i>	<i>474</i>	<i>353</i>	<i>797</i>	<i>613</i>	<i>263</i>	<i>2500</i>	<i>500</i>
1	Parcela	41	42	43	44	45		
	Variedades	1	6	11	16	21		
	<i>Rendimiento</i>	<i>491</i>	<i>796</i>	<i>536</i>	<i>145</i>	<i>298</i>	<i>2266</i>	<i>453</i>
4	Parcela	50	49	48	47	46		
	Variedades	4	9	14	19	24		
	<i>Rendimiento</i>	<i>120</i>	<i>486</i>	<i>441</i>	<i>620</i>	<i>45</i>	<i>1712</i>	<i>342</i>
						10013	401	

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 99. Promedio de rendimiento de las variedades por repeticiones; suma, promedio y varianza, valor escalar al máximo (EAM) y priorización de selección.*

Tratamientos	R1	R2	Suma	Media	Varianza	EAM	Prioridad
V1	530	491	1021	511	776	56	11
V2	103	458	561	281	63137	31	18
V3	689	353	1042	521	56582	57	10
V4	789	120	908	454	223881	50	13
V5	1452	348	1801	900	609132	99	3
V6	643	796	1439	719	11682	79	7
V7	965	407	1371	686	155849	75	8
V8	1157	613	1769	885	147914	97	5
V9	1304	486	1790	895	333908	98	4
V10	1078	747	1825	912	54599	100	1
V11	1120	536	1656	828	170265	91	6
V12	361	118	480	240	29525	26	20
V13	1019	797	1816	908	24553	100	2
V14	355	441	796	398	3745	44	16
V15	129	407	536	268	38531	29	19
V16	128	145	273	137	136	15	24
V17	500	375	875	437	7781	48	14
V18	501	474	976	488	355	54	12
V19	629	620	1248	624	40	68	9
V20	198	187	385	192	53	21	23
V21	154	298	451	226	10411	25	21
V22	166	223	388	194	1627	21	22
V23	377	263	640	320	6527	35	17
V24	75	45	121	60	438	7	25
V25	604	265	869	434	57613	48	15
$\Sigma R$	15023	10013	25036				
$\bar{x} R$	601	401	1001	501			

*Cuadro 100. Promedios de rendimiento de grano los bloques incompletos alojados en las repeticiones*

Bloques incompletos	R1	R2	$\Sigma BI$
1	3563	2266	5829
2	5146	1581	6726
3	2984	2500	5484
4	1955	1712	3668
5	1375	1954	3329
$\Sigma R$	15023	10013	25036

Nota: R = 2, V = 25, R(V) = 50, BI = 5 y Parcelas = 5

**Hipótesis estadísticas sobre las repeticiones:**

H<sub>0</sub>: Los rendimientos en las repeticiones son estadísticamente iguales

H<sub>1</sub>: Los rendimientos en las repeticiones son estadísticamente diferentes

**Hipótesis estadísticas sobre los bloques incompletos:**

H<sub>0</sub>: Los rendimientos en los bloques incompletos son estadísticamente iguales

H<sub>1</sub>: Los rendimientos en los bloques incompletos son estadísticamente diferentes

**Hipótesis estadísticas sobre los tratamientos:**

H<sub>0</sub>: Los rendimientos de las variedades de maíz son estadísticamente iguales

H<sub>1</sub>: Al menos una variedad de maíz tiene rendimientos estadísticamente diferente

**Regla de decisión:**

Si  $F_{calculada} < F_{crítico}$ : Aceptar H<sub>0</sub>

Si  $F_{calculada} \geq F_{crítico}$ : Aceptar H<sub>1</sub>

**Cálculo de sumas de cuadrados:**

Factor de corrección:

$$FC = \frac{\Sigma(X)^2}{R(t)} \rightarrow FC = \frac{(25036)^2}{50} = 12.536.176$$

Suma de Cuadrados Total:

$$SC T = \Sigma X^2 - FC$$

$$SCT = (530)^2 + (491)^2 + (103)^2 + \dots + (604)^2 + (265)^2 - 12.536.176 = 5.646.145$$

SC Repeticiones:

$$SC R = \frac{\Sigma R^2}{t} - FC$$

$$SC R = \frac{(15.023)^2 + (10.013)^2}{25} - 12.536.176 = 502.072$$

SC Bloques incompletos:

$$SC BI = \frac{\Sigma BI^2}{R(\text{Parcelas})} - FC$$

$$SC BI = \frac{(5829)^2 + (6726)^2 + (5484)^2 + (3668)^2 + (3329)^2}{10} - 12.536.176 = 846.717$$

SC Variedades:

$$SC V = \frac{\Sigma t^2}{R} - FC$$

$$SC V = \frac{(1021)^2 + (561)^2 + \dots + (120)^2 + (869)^2}{2} - 12.536.176 = 3.637.085$$

Suma de Cuadrados del Error: SCE = SC T – SC R – SC V – SC BI

$$SCE = 5.646.145 - 502.072 - 3.637.085 - 846.717 = 660.271$$

En el Cuadro 101, se expone el ADEVA del experimento en látice:

Cuadro 101. ADEVA del experimento en Látice de 25 variedades

Fuente de variación	GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p	Variación (%)
Repeticiones	1	502.072	502072	12,17**	4,498	8,53	0,003	8,9
Bloques incompletos BI	8	846.717	105840	2,56ns	2,591	3,89	0,052	15,0
Variedades	24	3.637.085	151545	3,67**	2,235	3,18	0,005	64,4
Error experimental	16	660.271	41267					11,7
Total	49	5.646.145						100

Nota: ns = no hay diferencia estadística significativa, \*\* hay diferencia estadística altamente significativa

$$CV(\%) = \frac{\sqrt{CME}}{\bar{y}} (100) = \frac{\sqrt{41266,94}}{500,72} (100) = 40,6 \%$$

El CV de 40,6 % tiende a ser elevado lo cual indica que hay una alta dispersión.

**Interpretación:**

Sobre las repeticiones: F calculada > F<sub>0,01</sub>: Se acepta H<sub>1</sub> con el 99% de confianza. Hay diferencia estadística altamente significativa entre repeticiones (p = 0,003).

Sobre los bloques incompletos: F calculada < F<sub>0,05</sub>: Se acepta H<sub>0</sub>. Los promedios de rendimientos en los bloques incompletos resultaron estadísticamente iguales (p = 0,052).

Sobre los tratamientos: F calculada > F<sub>0,01</sub>: Se acepta H<sub>1</sub> con el 99% de confianza. Al menos una de las variedades de maíz resulta estadísticamente diferentes (p = 0,005).

**Decisión:**

Hay diferencias estadísticas entre variedades de maíz (p < 0,01). El 64,4 % de la variación se explica por el factor genético. El EAM indica que la mejor variedad es V10.

6.2.5 EXPERIMENTO CON DISEÑO IRRESTRICTO AL AZAR

El diseño irrestricto al azar (DIA) es aquel donde se asigna aleatoriamente los tratamientos a las unidades experimentales. La aleatorización en la asignación de tratamientos es una de las condiciones para que tenga validez el análisis de varianza (Melo, et al., 2020, p.23). Este diseño se basa muestreos dentro de las parcelas y se recomienda usar cuando se desconoce el factor causante del “ruido” o no es posible controlarlo (Navarro y Vargas, 2015, p. 3).

En un DIA, las unidades experimentales se asignan aleatoriamente posibilitando controlar los sesgos y maximizar la validez interna de los resultados (Navarro y Vargas, 2015, p. 3). El DIA se usa en experimentación con especies forestales y otros cultivos perennes, en laboratorio, en viveros e invernaderos o en ensayos con animales, donde las unidades experimentales no pueden organizarse.

Los ensayos se diseñan, principalmente, para probar dosis de abonos edáficos o foliares, dosis de bioestimulantes o de productos hormonales, mezclas de sustancias en dietas o cambios en concentraciones de sustancias componentes de alimentos para animales. Se aplica en ensayos balanceados o desbalanceados, no es recomendable cuando el número de tratamientos en elevado o se evidencie la acción de factores ambientales que provocan alteraciones no uniformes.

El modelo estadístico es el siguiente:

$$Y_{ij} = \mu + t_{j(k)} + \epsilon_{ij} + \lambda_{ijk}$$

Dónde:

$Y_{ijk}$  = Observación cualquiera

$\mu$  = Valor de la media general

$t_i$  = Efecto del i-ésimo tratamiento (i =1, ...,t)

$\epsilon_{ijk}$  = Efecto del error experimental

$\lambda_{ijk}$  = Efecto del error de los muestreos

**Esquema del análisis de varianza:**

Fuente de variación	Grados de libertad
Tratamientos (t)	t – 1
Error en parcelas /tratamiento	t(n – 1)
Error del muestreo en parcela	n.t(m – 1)
Error experimental total	t(nm – 1)
Total	n.m.t – 1

Nota: t = tratamientos, n = Parcelas por tratamiento, m = muestras por unidad experimental

**Sumas de cuadrados en los experimentos DIA:**

Factor de corrección:

$$FC = \frac{(\sum X)^2}{n}$$

Suma de Cuadrados Total:  $SCT = \sum X^2 - FC$

SC Tratamientos:  $SC t = \frac{\sum t^2}{Muestras} - FC$

Suma de Cuadrados del Error experimental:  $SCE = SC Total - SC Tratamientos$



La SCE se desagrega en sus partes: SCE parcela.tratamiento y SCE muestra.parcela

Suma de cuadrados del error en parcela.tratamiento:

SCE Par.t = Minuendo del total – Minuendo de parcela. tratamiento

Suma de cuadrados del error en Muestra.parcela:

$$SCE \text{ Muestra.Parcela} = \frac{\Sigma(\text{Par.t})^2}{n.m} - \frac{\Sigma t^2}{n.m}$$

Suma de cuadrados del error:

SCE = SCE Muestra.Parcela + SCE Parcela.t

### Ejercicio 119:

Realizar el ADEVA en un experimento DIA donde se probaron tres tratamientos de fertilización edáfica en fréjol, a nivel de invernadero, en cuatro macetas, tomando dos muestras foliares, en distintos momentos para determinar el contenido de magnesio en las hojas (en ppm) mediante análisis foliar.

#### Datos y análisis:

En el Cuadro 102, se indica los datos promedios del contenido de magnesio (Mg) como respuesta a la fertilización edáfica, a nivel de invernadero, en dos muestras tomadas en cuatro macetas cada una.

Cuadro 102. Promedios de magnesio en hojas de fréjol en función de tres tratamientos de fertilización edáfica. Diagrama de un experimento en cuadrado latino 4 x 4

Tratamientos: Fertilización edáfica	Muestras	Macetas				Σ tratamientos por muestra
		1	2	3	4	
t1	1	3,3	3,5	4,1	3,9	14,8
	2	3,5	3,6	3,7	3,8	14,6
t2	1	5,0	4,4	3,8	4,7	17,9
	2	3,9	4,6	5,1	4,2	17,8
t3	1	8,0	7,9	7,0	7,8	30,7
	2	8,1	8,0	7,4	7,0	30,5

Nota: n = 4 parcelas (macetas), m = 2 muestras, t = 3 tratamientos, n.m.(t) = 12 unidades experimentales

#### Hipótesis estadísticas sobre los tratamientos:

H<sub>0</sub>: El contenido de Mg en las muestras foliares de fréjol resultaron iguales

H<sub>1</sub>: El contenido de Mg en al menos un tratamiento es estadísticamente diferente

#### Regla de decisión:

Si F calculada < F crítico: Aceptar H<sub>0</sub>

Si F calculada ≥ F crítico: Aceptar H<sub>1</sub>

En el Cuadro 103, se resume los promedios de magnesio en las hojas de fréjol en función de los tres tratamientos de fertilización edáfica.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 103. Medias de Mg en follaje de fréjol como respuesta a tres alternativas de fertilización edáfica*

Tratamientos	Muestras /parcela /t	Σ P1	Σ P2	Σ P3	Σ P4	Σ t	$\bar{X} t$ (ppm de Mg)
t1	m 1 y 2	6,8	7,1	7,8	7,7	29,4	3,68
t2	m 1 y 2	8,9	9,0	8,9	8,9	35,7	4,46
t3	m 1 y 2	16,1	15,9	14,4	14,8	61,2	7,65
$\Sigma t$						126,3	--
$\bar{X} t$						--	5,26

Resumen de las características del ensayo:

Σ t	$\bar{X} t$	t	Parcelas (n)	Muestras (m)	UE: t(n)m	Parcela.muestra (n.m)
126,3	5,26	3	4	2	24	8

***Cálculo de las sumas de cuadrados en los experimentos DIA:***

Factor de corrección:

$$FC = \frac{(X)^2}{n.m.(t)} \rightarrow FC = \frac{(126,3)^2}{24} = 664,65$$

Suma de Cuadrados Total:

$$SC T = \Sigma X^2 - FC$$

$$SCT = (3,3)^2 + (3,5)^2 + (4,1)^2 + \dots + (7,4)^2 + (7,0)^2 - 664,65 = 739,03 - 664,65 = 74,4$$

SC Tratamientos:

$$SC t = \frac{\Sigma t^2}{n(m)} - FC$$

$$SC t = \frac{(29,4)^2 + (35,7)^2 + (61,2)^2}{8} - 664,65 = 735,54 - 664,65 = 70,88$$

Suma de Cuadrados del Error experimental: SCE = SC Total – SC Tratamientos:

$$SCE = 74,4 - 70,88 = 3,49$$

En este diseño, la SCE se desagrega en sus componentes: SCE n.(t) y SCE n.(m)

Suma de cuadrados del error en parcela.tratamiento:

$$SCE Par.t = \Sigma X^2 - \frac{\Sigma(Par.t)^2}{m}$$

$$SCE Par.t = 739,03 - \frac{(6,8)^2 + (7,1)^2 + \dots + (14,4)^2 + (14,8)^2}{2}$$

$$SCE Par.t = 739,03 - 736,92 = 2,11$$

Suma de cuadrados del error en Muestra.parcela:

$$SCE Muestra.Parcela = \frac{\Sigma(Par.t)^2}{n'm} - \frac{\Sigma t^2}{n'm} \rightarrow = 736,92 - \frac{(29,4)^2 + (35,7)^2 + (61,2)^2}{2(4)}$$

$$SCE Muestra.Parcela = 736,92 - 735,54 = 1,38$$

$$Se\ verifica: SCE = SCE\ Muestra.Parcela + SCE\ Parcela.t = 2,11 + 1,38 = 3,49$$

**Resultados del análisis de varianza:**

FV	GL	Cuadrados de X	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p	Variación (%)
Tratamientos	2	739,03	70,88	35,44	213,0**	3,47	5,78	0,00001	95,30
Error experimental	21		3,49	0,17					4,70
<i>Error en parcela/tratamiento</i>	(12)	736,92	739,03 - 736,92 = 2,11	0,176					
<i>Error en muestra/parcela</i>	(9)	735,54	736,92 - 735,54 = 1,38	0,153					
Total	23		74,38						100,00

Nota: \*\* Hay diferencia estadística altamente significativa entre tratamientos

$$CV(\%) = \frac{\sqrt{CME}}{\bar{y}} (100) = \frac{\sqrt{0,17}}{5,26} (100) = 7,83 \%$$

El CV reducido indica que la media es representativa del experimento y que la dispersión de los datos es adecuada. Los tratamientos explican el 95,30 % de la variación.

**Interpretación:**

F calculada > F<sub>0,01</sub>: Se acepta H<sub>1</sub>.

**Decisión:**

El contenido de Mg en las muestras foliares de fréjol, en al menos un tratamiento, es estadísticamente diferente con el 99 % de confianza. Se debe realizar una prueba *post hoc*.

6.2.6 EXPERIMENTOS FACTORIALES

Un experimento de un factor (A) con varios niveles (A<sub>1</sub>, A<sub>2</sub>,...A<sub>n</sub>) es de tipo unifactorial. Por ejemplo: “Probar la fertilización nitrogenada en cuatro dosis (niveles) sobre la producción de biomasa” (variable de respuesta). El modelo teórico: Y = f(X) indica que la producción de biomasa depende del fertilizante nitrogenado (N). La pregunta de investigación puede formularse como: ¿Cuál es la dosis óptima de N para obtener la máxima producción por hectárea?

Los experimentos unifactoriales se adaptan a cualquiera de los diseños: DCA, BA, CL y LT siendo evidente que los resultados conducen a identificar la causalidad de los fenómenos e introducir en el análisis predictivo. Con los experimentos factoriales se pueden probar más de dos factores simultáneamente. Los programas computacionales facilitan los análisis multifactoriales, determinando los efectos de cada factor y de sus interacciones.

En un experimento factorial completo, los niveles de un factor se combinan con cada uno de los niveles de los otros factores. Un cambio en un nivel de un factor (variable independiente) generará una respuesta diferenciada en la variable dependiente. En un estudio factorial incompleto no se combinan todos los niveles de los factores en estudio, solo se seleccionan una parte con base a criterios matemáticos como la ortogonalidad. Los experimentos factoriales más usados son:

**Factoriales simples.**- Son del tipo A x B, donde se prueban dos factores en distintos niveles. Por ejemplo: densidad poblacional x láminas de riego. La densidad a probarse puede ser de varios niveles: A<sub>1</sub> y A<sub>2</sub>. Las láminas de riego pueden probarse en tres niveles: L<sub>1</sub>, L<sub>2</sub> y L<sub>3</sub>. Al ser un ensayo A en dos niveles x B en tres niveles se trataría de un experimento 2 x 3 = 6 combinaciones.

**Factoriales  $A^k$ .**- Esta nomenclatura indica que se prueban:  $k$  = factores en  $A$  = niveles. Un diseño  $(2)^2$  corresponde a dos factores en dos niveles cada uno, por lo tanto se genera 4 combinaciones o tratamientos. Un factor  $A$  = densidad poblacional con dos niveles ( $A_1, A_2$ ) y un factor  $B$  como láminas de riego, también con dos niveles ( $L_1, L_2$ ) origina un experimento de cuatro tratamientos:  $A_1L_1, A_1L_2, A_2L_1, A_2L_2$ . En un Diseño  $(2)^3$  donde se prueban tres factores por dos niveles cada uno: Factor  $A$  = densidad poblacional ( $A_1, A_2$ ), factor  $B$  = láminas de riego ( $B_1$  y  $B_2$ ) y factor  $C$  = Fertilización ( $C_1$  y  $C_2$ ) genera ocho tratamientos:  $A_1B_1C_1, A_1B_1C_2, A_1B_2C_1, A_1B_2C_2, A_2B_1C_1, A_2B_1C_2, A_2B_2C_1$  y  $A_2B_2C_2$ .

**Arreglos factoriales  $A^k + 1$ .**- En estos experimentos de  $k$  factores con  $A$  niveles se adiciona un testigo, que puede referirse a la tecnología local del productor.

**Arreglos factoriales  $A^k + 1 + 1$ .**- En estos experimentos, al factorial  $A^k$ , se adicionan dos testigos, uno puede ser el absoluto (+1) y el otro un testigo local.

### 6.2.6.1 EXPERIMENTO BIFACTORIAL

Un experimento bifactorial se basa en la función:  $Y = f(X_1, X_2)$ . En esta función intervienen dos factores causales (variables independientes  $X_1$  y  $X_2$ ), que se conjetura tienen influencia sobre una variable de respuesta (variable dependiente  $Y$ ).

El modelo  $Y = f(X_1, X_2)$  se operativiza, por ejemplo, en la función:  $PR = f(V, F)$ .

Dónde:

PR = Producción de maíz (kilos. parcela<sup>-1</sup>)

V = Variedades

F = Alternativas de fertilización (kg.ha<sup>-1</sup>).

Se formulan las siguientes preguntas: ¿Cuál es la variedad de mayor producción?, ¿Cuál es la alternativa de fertilización que permite obtener la mayor producción?

#### Ejercicio 120:

Realizar el análisis de varianza de un experimento BA con arreglo factorial  $2^2$ , donde se prueban dos variedades y dos alternativas de fertilización, en 5 repeticiones, y la variable de respuesta es producción medida en kg.parcela<sup>-1</sup>.

#### Datos y análisis:

En el Cuadro 104, se presentan los resultados del experimento de campo.

Cuadro 104. Datos de producción del experimento bifactorial VF

Tratamientos	Factor V Código	Factor F Variedad	Fertilización	Repeticiones					Suma t	Media (kg.parcela <sup>-1</sup> )
				I	II	III	IV	V		
V1F1	1	1		11	12	14	10	11	58	11,6
V1F2	1	2		13	15	12	13	10	63	12,6
V2F1	2	1		14	13	14	16	15	72	14,4
V2F2	2	2		15	17	16	18	17	83	16,6
	Suma R			53	57	56	57	53	276	
	Media R			13,25	14,25	14,00	14,25	13,25		13,8

Nota: V = 2, F = 2, VF =  $t = 4$ , R = 5, V.F.R =  $2(2)(5) = 20 = n$

#### Hipótesis sobre las repeticiones:

H<sub>0</sub>: Las repeticiones no tienen efecto significativo sobre la producción

H<sub>1</sub>: En al menos una repetición hay un efecto de las repeticiones

**Hipótesis estadísticas sobre los tratamientos:**

H<sub>0</sub>: No hay diferencia estadística entre tratamientos

H<sub>1</sub>: En al menos un tratamiento es estadísticamente diferente

**Hipótesis estadísticas sobre las variedades:**

H<sub>0</sub>: La producción de las dos variedades es estadísticamente igual

H<sub>1</sub>: Una de las dos variedades probadas es estadísticamente diferente

**Hipótesis estadísticas sobre la fertilización:**

H<sub>0</sub>: No hay efecto de la fertilización

H<sub>1</sub>: Una de las alternativas de fertilización es estadísticamente diferente

**Regla de decisión:**

Si F calculada < F crítica: Aceptar H<sub>0</sub>

Si F calculada ≥ F crítica: Aceptar H<sub>1</sub>

**Cálculo de la suma del efecto de las variedades:**

Variedad	Fertilización		Σ variedades	X̄ Variedades
	1	2		
1	58	63	121	60,5
2	72	83	155	77,5
Σ fertilización	130	146	276	--
X̄ fertilización	65	72		69

**Interacción las variedades x fertilización:**

Variedad	Fertilización		X̄ variedades
	F1	F2	
V1	11,6	12,6	12,1
V2	14,4	16,6	15,5
X̄ fertilización	13,0	14,6	13,8

Nota: La combinación V2F2 muestra la tendencia de mejor respuesta

**Cálculo de sumas de cuadrados:**

$$\text{Factor de corrección (FC)} = \frac{\Sigma Y^2}{VFR} = \frac{(276)^2}{20} = 3.809$$

$$\text{Suma de Cuadrados Total (SC T)} = \Sigma Y^2 - FC$$

$$\text{SC Total} = [(11)^2 + (12)^2 + (14)^2 + \dots + (18)^2 + (17)^2] - 3.809 \rightarrow \text{SC T} = 105,2$$

$$\text{SC R} = \frac{\Sigma R^2}{VF} - FC \rightarrow \text{SC R} = \frac{53^2 + 57^2 + 56^2 + 57^2 + 53^2}{4} - 3.809 \rightarrow \text{SC R} = 4,2$$

$$\text{SC t} = \frac{\Sigma t^2}{R} - FC \rightarrow \text{SC t} = \frac{58^2 + 63^2 + 72^2 + 83^2}{5} - 3.809 \rightarrow \text{SC t} = 72,4$$

$$\text{SC V} = \frac{\Sigma V^2}{RF} - FC \rightarrow \text{SC V} = \frac{121^2 + 155^2 + 72^2 + 83^2}{10} - 3.809 \rightarrow \text{SC V} = 57,8$$

$$\text{SC F} = \frac{\Sigma F^2}{RV} - FC \rightarrow \text{SC F} = \frac{130^2 + 146^2}{10} - 3.809 = 12,8$$

$$\text{SC (VF)} = \text{SC t} - \text{SC V} - \text{SC F} \rightarrow \text{SC(VF)} = 72,4 - 57,8 - 12,8 = 1,8$$

Suma de Cuadrados del Error:  $SCE = SC\ T - SC\ R - SC\ t$

$$SCE = 105,2 - 4,2 - 72,4 = 28,6$$

**Resultados del análisis de varianza:**

Los resultados del ADEVA del experimento, se indica en el Cuadro 105.

Cuadro 105. Análisis de la varianza del experimento

Fuente de variación	GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p	Variación (%)
Repeticiones (R)	4	4,20	1,05	0,44 ns	3,259	5,412	0,777	4,0
Tratamientos (t)	3	72	24,13	10,13 **	3,490	5,953	0,001	--
Variedades (V)	(1)	(58)	58	24,25 **	4,747	9,330	0,000	55,1
Fertilización(F)	(1)	(12)	12	5,37 *	4,747	9,330	0,039	11,4
Interacción VF	(1)	(2)	2	0,76 ns	4,747	9,330	0,402	1,9
Error Experimental	12	28,6	2,38					27,2
Total	19	105,2						

Nota: ns = no hay diferencia estadística significativa, \* = hay diferencia estadística significativa, \*\* hay diferencia estadística altamente significativa

$$CV(\%) = \frac{\sqrt{CME}}{\bar{x}} (100) = \frac{\sqrt{2,38}}{13,8} (100) = 11,2 \%$$

El CV de 11,2 % indica que la media es representativa del ensayo y la dispersión adecuada. La contribución del factor genético a la variación es 55,1 %, de la fertilización 11,4 % y del error 27,2 %.

**Interpretación:**

Sobre las repeticiones:

$F_{calculada} < F_{0,05}$ : Se acepta  $H_0$ . Las repeticiones no tienen efecto sobre la producción.

Sobre las variedades:

$F_{calculada} > F_{0,01}$ : Se acepta  $H_1$ . Al menos una variedad es estadísticamente diferente.

Sobre fertilización:

$F_{calculada} > F_{0,01}$ : Se acepta  $H_1$ . Al menos una alternativa de fertilización es estadísticamente diferente.

Sobre la interacción variedades x fertilización:

$F_{calculada} < F_{0,05}$ : Se acepta  $H_0$ . No hay interacción estadísticamente significativa

**Decisión:**

Hay diferencias estadísticas altamente significativas entre variedades ( $p < 0,01$ ) y hay diferencias significativas entre niveles de fertilización ( $p < 0,05$ ).

6.2.6.2 EXPERIMENTO TRIFACTORIAL

El término “arreglo factorial” se refiere a la combinación de los factores que se planean comparar, por tanto, en un ensayo trifactorial se prueban tres factores.

**Ejercicio 121:**

Realizar el análisis de la varianza de la producción de forraje seco ( $t \cdot ha^{-1}$ ) de cuatro genotipos de sorgo (G), probados en dos ciclos de cultivo (C) y una aplicación de bioestimulante foliar (Es) durante el crecimiento (Reyes, 1980, p. 103).

**Datos y análisis:**

Los datos del experimento del comportamiento de cuatro genotipos en dos ciclos de cultivo y con la aplicación de bioestimulante se exponen en el Cuadro 106.

*Cuadro 106. Producción de forraje de cuatro genotipos de sorgo en dos ciclos de cultivo y aplicación de bioestimulante*

Ciclos	Genotipos	Bioestimulante	Bloques						$\bar{x}$ (t.ha <sup>-1</sup> )	
			1	2	3	4	5	6		
Invierno	F1	Con	23,2	21,7	21,0	21,8	22,9	20,5	21,9	
		Sin	22,3	18,7	22,4	19,4	23,3	22,0	21,4	
	F2	Con	17,3	20,9	20,4	18,2	19,7	19,6	19,4	
		Sin	16,0	17,6	14,0	15,5	15,7	17,9	16,1	
	Progenitor A	Con	14,3	17,0	15,2	13,8	13,4	13,3	14,5	
		Sin	18,1	17,8	21,5	17,5	20,0	18,4	18,9	
	Progenitor B	Con	16,2	18,6	19,6	19,0	21,3	18,9	18,9	
		Sin	16,9	11,7	16,2	15,0	19,0	18,3	16,2	
	Verano	F1	Con	19,7	13,7	18,1	21,5	17,1	20,4	18,4
			Sin	20,4	16,6	17,7	19,0	21,4	19,0	19,0
		F2	Con	20,1	18,8	19,6	17,9	21,7	23,4	20,3
			Sin	17,5	16,4	17,9	17,2	16,0	17,0	17,0
Progenitor A		Con	14,9	15,3	17,6	17,7	16,3	15,3	16,2	
		Sin	14,3	14,3	13,7	15,3	13,8	16,3	14,6	
Progenitor B		Con	17,1	14,5	15,5	16,4	19,5	18,2	16,9	
		Sin	16,0	17,4	14,5	15,6	20,3	16,3	16,7	
$\bar{x}$ Bloques			17,8	16,9	17,8	17,6	18,8	18,4	17,9	
Bloques (B)		Ciclos (C)	Genotipos (G)	Bioestimulantes (Es)		Tratamientos		<i>n</i>		
6		2	4	2		C(G)(Es) = 16		96		

**Hipótesis estadísticas:**

Las hipótesis estadísticas para cada fuente de variación se planean como una matriz, ajustada al esquema del análisis de varianza.

**Matriz de hipótesis estadísticas por fuente de variación**

Fuentes de variación	Hipótesis nula (H <sub>0</sub> )	Hipótesis alternativa (H <sub>1</sub> )
Bloques (B)	No hay diferencia entre bloques	Si hay diferencia estadística entre bloques
Genotipos (G)	No hay diferencia entre genotipos	Al menos un genotipo es estadísticamente diferente
Ciclos (C)	No hay diferencia estadística entre ciclos	Hay diferencia estadística significativa entre ciclos
Bioestimulantes (Es)	No hay efecto significativo por la aplicación de bioestimulantes	Si hay efecto significativo por la aplicación de bioestimulantes
IG x C	No hay interacción significativa genotipos x ciclos de cultivo	Si hay interacción significativa genotipos x ciclos de cultivo
IG x Es	No hay interacción genotipos x bioestimulantes	Si hay interacción significativa genotipos x aplicación de bioestimulantes
IC x Es	No hay interacción significativa ciclos x aplicación de bioestimulantes	Si hay interacción significativa ciclos x aplicación de bioestimulantes
IC x G x Es	No hay interacción significativa ciclos x genotipos x bioestimulantes	Si hay interacción significativa ciclos x genotipos x aplicación de bioestimulantes

**Regla de decisión:**

Si  $F_{\text{calculada}} < F_{\text{crítico}}$ : Aceptar  $H_0$

Si  $F_{\text{calculada}} \geq F_{\text{crítico}}$ : Aceptar  $H_1$

En el Cuadro 107, se indican los valores desglosados de las sumas y medias de genotipos y bloques, con ajuste al diseño en bloques al azar. En el Cuadro 108, se exponen las sumas y medias de genotipos y bioestimulantes. En el Cuadro 109, se presentan las sumas y medias de genotipos y ciclos de cultivo. En el Cuadro 110, se indican las sumas y medias de bioestimulantes y ciclos de cultivo. En el Cuadro 111, se observan las sumas y medias genotipos, ciclos de cultivo y bioestimulantes.

*Cuadro 107. Sumas y medias de genotipos y bloques del experimento*

Genotipos	Bloques						$\Sigma$ genotipos	$\bar{X}$ genotipos (t.ha <sup>-1</sup> )
	1	2	3	4	5	6		
F1	85,6	70,7	79,2	81,7	84,7	81,9	483,8	20,2
F2	70,9	73,7	71,9	68,8	73,1	77,9	436,3	18,2
Progenitor 1	61,6	64,4	68,0	64,3	63,5	63,3	385,1	16,0
Progenitor 2	66,2	62,2	65,8	66,0	80,1	71,7	412,0	17,2
$\Sigma$ bloques	284,3	271,0	284,9	280,8	301,4	294,8	1717,2	
$\bar{X}$ de bloques	17,8	16,9	17,8	17,6	18,8	18,4		17,9

*Cuadro 108. Sumas y medias de genotipos y bioestimulantes*

Genotipos	Con bioestimulante	Sin bioestimulante	$\Sigma$ genotipos	$\bar{X}$ Genotipos (t.ha <sup>-1</sup> )
F1	241,6	242,2	483,8	20,2
F2	237,6	198,7	436,3	18,2
Progenitor 1	184,1	201,0	385,1	16,0
Progenitor 2	214,8	197,2	412,0	17,2
$\Sigma$ Bioestimulantes	878,1	839,1	1717,2	--
$\bar{X}$ Bioestimulantes	18,3	17,5		17,9

*Cuadro 109. Sumas y medias de genotipos y ciclos de cultivo*

Genotipos	Ciclo 1	Ciclo 2	$\Sigma$ genotipos	$\bar{X}$ Genotipos (t.ha <sup>-1</sup> )
F1	259,2	224,6	483,8	20,2
F2	212,8	223,5	436,3	18,2
Progenitor 1	200,3	184,8	385,1	16,0
Progenitor 2	210,7	201,3	412,0	17,2
$\Sigma$ ciclos	883,0	834,2	1717,2	
$\bar{X}$ ciclos	18,4	17,4		17,9



**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 110. Sumas y medias de bioestimulantes y ciclos de cultivo*

Bioestimulantes	Ciclo 1	Ciclo 2	Σ bioestimulantes	$\bar{X}$ bioestimulantes
Con bioestimulante	447,8	430,3	878,1	18,3
Sin bioestimulante	435,2	403,9	839,1	17,5
Σ ciclos	883,0	834,2	1717,2	
$\bar{X}$ ciclos	18,4	17,4		17,9

*Cuadro 111. Sumas y medias de genotipos, ciclos de cultivo y bioestimulantes*

Genotipos	Ciclo 1		Ciclo 2		Σ genotipos	$\bar{X}$ genotipos
	Con bioestimulante	Sin bioestimulante	Con bioestimulante	Sin bioestimulante		
F1	131,1	128,1	110,5	114,1	483,8	20,2
F2	116,1	96,7	121,5	102,0	436,3	18,2
Progenitor 1	87,0	113,3	97,1	87,7	385,1	16,0
Progenitor 2	113,6	97,1	101,2	100,1	412,0	17,2
Σ ciclos x bioestimulantes	447,8	435,2	430,3	403,9	1717,2	
$\bar{X}$ ciclos x bioestimulantes	18,7	18,1	17,9	16,8		17,9

**Cálculos de la suma de cuadrados:**

$$FC = \frac{(\sum X)^2}{n} \rightarrow FC = \frac{(1717,2)^2}{96} = 30.716,4$$

Suma de cuadrados total:  $SC T = \sum X^2 - FC$

$$SC T = [(14,3)^2 + (18,1)^2 + (16,2)^2 + \dots + (19,0)^2 + (23,4)^2 + (17,0)^2] - 30.716,4 = 678,9$$

Suma de cuadrados de bloques:  $SC \text{ bloques} = \frac{\sum \text{bloques}^2}{C(Es)(G)} - FC$

$$SC B = \frac{(284,4)^2 + (271)^2 + (284,9)^2 + (280,8)^2 + (301,4)^2 + (294,8)^2}{16} - 30.716,4 = 35,7$$

Suma de cuadrados de genotipos:  $SC G = \frac{\sum G^2}{B(C)(Es)} - FC$

$$SC G = \frac{[(483,8)^2 + (436,3)^2 + (385,1)^2 + (412,0)^2]}{6(2)(2)} - 30.716,4 = 219,7$$

Suma de cuadrados de ciclos:  $SC C = \frac{\sum C^2}{B(Es)(G)} - FC$

$$SC C = \frac{(883,0)^2 + (834,2)^2}{48} - 30.716,4 = 24,8$$

Suma de cuadrados de bioestimulantes:  $SC Es = \frac{\sum Es^2}{B(C)(G)} - FC$

$$SC Es = \frac{(878,1)^2 + (839,1)^2}{6(2)(4)} - 30.716,4 = 15,8$$

Suma de cuadrados de la interacción Ciclos x bioestimulantes:

$$SC C. Es = \frac{\sum (C.Es)^2}{B(G)} - FC - SCC - SC Es$$

$$SC C. Es = \frac{(447,8)^2 + (435,2)^2 + (430,3)^2 + (8403,9)^2}{24} - 30.716,4 - 24,8 - 15,8$$

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

$$SC\ C.\ G = \frac{(259,2)^2 + (212,8)^2 + (200,3)^2 + (210,7)^2 + (224,6)^2 + (223,5)^2 + (184,8)^2 + (201,3)^2}{12} - 30.716,4 - 24,8 - 219,7$$

$$SC\ C.G = 31004,4 - 30716,4 - 24,8 - 219,7 = 43,5$$

Suma de cuadrados de la interacción Genotipos x Bioestimulantes:

$$SC\ G.\ Es = \frac{\Sigma(G.Es)^2}{BC} - FC - SC\ G - SC\ Es$$

$$SC\ G.\ Es = \frac{(241,6)^2 + (237,6)^2 + (184,1)^2 + (214,8)^2 + (242,2)^2 + (198,7)^2 + (201,0)^2 + (197,2)^2}{12} - 30.716,4 - 219,7 - 15,8 = 71,1$$

Suma de cuadrados de la interacción Ciclos x Genotipos x Bioestimulantes:

$$SC\ C.\ G.\ Es = \frac{\Sigma C.G.Es^2}{B} - FC - SC\ C - SC\ G - SC\ Es$$

$$SC\ C.\ G.\ Es = \frac{(131,1)^2 + (128,1)^2 + (110,5)^2 + (114,1)^2 + (116,1)^2 + (96,7)^2 + \dots + (101,2)^2 + (100,1)^2}{6} - 30716,4 - 24,8 - 219,7 - 15,8$$

$$SC\ C.G.Es = 31157,1 - 30716,4 - 24,8 - 219,7 - 15,8 = 180,4$$

SC Error experimental: SC E = SC T – SC Bloques – SC Tratamientos

$$SC\ E = 678,9 - 35,7 - 558,4 = 84,9$$

En el Cuadro 112, se indica el ADEVA del experimento.

*Cuadro 112. Análisis de la varianza del ensayo de cuatro genotipos de sorgo forrajero en dos ciclos de cultivo y aplicación de bioestimulantes*

Fuente de variación	GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p valor	Variación (%)
Bloques (B)	5	35,7	7,14	6,31 **	2,34	3,27	0,00006	5,3
Tratamientos (t)	15	558,4	37,22	32,90 **	1,80	2,29	0,00000	--
Genotipos (G)	3	219,7	73,23	64,73 **	2,73	4,05	0,00000	32,4
Ciclos (C)	1	24,8	24,80	21,92 **	3,97	6,99	0,00001	3,7
Bioestimulantes (Es)	1	15,8	15,80	13,97 **	3,97	6,99	0,00036	2,3
G x C	3	43,5	14,50	12,82 **	2,73	4,05	0,00000	6,4
G x Es	3	72,1	24,03	21,24 **	2,73	4,05	0,00000	10,6
C x Es	1	2,1	2,05	1,81 ns	3,97	6,99	0,18232	0,30
C x G x Es	3	180,4	60,13	53,15 **	2,73	4,05	0,00000	26,6
Error Experimental	75	84,9	1,13					12,5
Total	95	678,9						100

Nota: ns = no hay diferencia estadística significativa, \*\* = Hay diferencia estadística significativa

$$CV(\%) = \frac{\sqrt{CME}}{Media} (100) = \frac{\sqrt{1,13}}{17,9} (100) = 5,94 \%$$

**Interpretación:**

En genotipos (G): F<sub>calculada</sub> = 64,73 > F<sub>0,05</sub> = 2,73 y F<sub>0,01</sub> = 4,05: Se acepta H<sub>1</sub>. Hay diferencia estadística significativa entre genotipos (p < 0,01).

Otras fuentes de variación: Se evidenció diferencias altamente significativas en casi todas las fuentes de variación (p < 0,001), excepto en la interacción C x Es (p = 0,1823).

La variación calculada para genotipos es 32,4 % que corresponde al parámetro genético “heredabilidad”, que se define como la proporción atribuible al factor genético en la variación fenotípica (Gordón y Camargo, 2015, p. 57).

La fórmula para calcular la heredabilidad en sentido amplio es la siguiente:

$$H(\%) = \frac{SC\ G}{SCT} (100) \quad [107]$$

Por ejemplo: Si en un ADEVA, la SC Total = 678,9 y la SC Variedades = 219,7 la heredabilidad se calcula como:

$$H(\%) = \frac{SC\ G}{SCT} (100) = \frac{219,7}{678,9} (100) = 32,4 \%$$

Además, este experimento proporciona información de los progenitores (P1 y P2) y de las progenies (F1 y F2), lo que posibilita calcular la heterosis con la fórmula sugerida por Robles (1986, p. 214):

$$\text{Heterosis} = \frac{F_1 - \frac{P_1 + P_2}{2}}{\frac{P_1 + P_2}{2}} (100) \quad [108]$$

Por ejemplo: usando las medias de los cuatro genotipos de sorgo, en el rendimiento de heno, se calcula la heterosis en 21,4 %.

$$\text{Heterosis} = \frac{F_1 - \frac{P_1 + P_2}{2}}{\frac{P_1 + P_2}{2}} (100) \rightarrow \text{Heterosis} = \frac{20,2 - \frac{16 + 17,2}{2}}{\frac{16 + 17,2}{2}} (100) = 21,4 \%$$

**Decisión:**

Hay diferencia estadística significativa entre genotipos ( $p < 0,01$ ). El 32,4 % de la variabilidad en la producción de forraje se explica por efecto de los genotipos.

6.2.6.3 EXPERIMENTO EN PARCELA DIVIDIDA

Las condiciones del campo experimental y los factores en estudio exigen ser creativos en el diseño, como es el caso de las parcelas divididas o subdivididas que son arreglos factoriales ordenados en “parcelas grandes” y “parcelas chicas, acopladas a los diseños básicos: DCA o BA. El criterio fundamental para usar un diseño de parcelas divididas es que haya un factor de difícil aleatorización.

Los factores que no puedan aleatorizar conformarán las parcelas grandes. Entre estos factores se encuentran: sistemas de riego, sistemas agroforestales o silvopastoriles, densidades poblacionales en cultivos perennes, sistemas de labranza y fumigaciones aéreas, entre otros. Las parcelas grandes corresponden al factor A. Estas se fraccionan en parcelas chicas sobre las cuales, en forma aleatorizada, se integra el factor B. En estas condiciones, cada unidad experimental se conforma de una combinación de los factores A x B. Para el análisis de datos, un diseño en parcela dividida tiene un enfoque de arreglo factorial, con una desagregación del error experimental para las parcelas grandes y para las parcelas chicas.

**Esquema del análisis de la varianza:**

Fuente de variación	Grados de libertad
Bloques (r)	r - 1
Factor A	A - 1
Error experimental de Factor A	(r - 1)(A - 1)
Parcelas grandes (A)	r.A - 1
Factor B	(B - 1)
Interacción AB	(A - 1)(B - 1)
Error experimental de Factor B	A(r - 1)(B - 1)
Subparcelas (B)	r.A.(B - 1)
Tratamientos (t)	A.B - 1
Error experimental total	(r - 1)(t - 1)
Total	r.A.B - 1

**Distribución en el campo de un experimento en parcela dividida:**

En un experimento en parcelas divididas, donde se prueban los factores A y B, cada uno en tres niveles, significa que se estudian nueve tratamientos, que al planearlo en 5 repeticiones, significa que se requiere 45 unidades experimentales (Figura 33).

<b>I</b>	A3B3	A3B1	A3B2	A2B2	A2B1	A2B3	A1B1	A1B3	A1B2
<b>II</b>	A1B1	A1B2	A1B3	A3B1	A3B2	A3B3	A2B3	A2B1	A3B2
<b>III</b>	A2B3	A2B1	A2B2	A1B2	A1B1	A1B3	A3B3	A3B2	A3B1
<b>IV</b>	A3B3	A3B2	A3B1	A2B1	A2B2	A2B3	A1B3	A1B2	A1B1
<b>V</b>	A1B1	A1B3	A1B2	A3B2	A3B1	A3B3	A2B1	A2B3	A2B2

Figura 33. Diagrama de un experimento en parcelas divididas de dos factores por tres niveles

**Ejercicio 122:**

Realizar el análisis de varianza del ensayo (3)<sup>2</sup> con arreglo en parcelas divididas en el diseño en BA con cinco repeticiones donde se prueba el efecto de los métodos de riego (A) y la densidad poblacional (B) sobre la producción de cacao (quintales.ha<sup>-1</sup>).

**Datos y análisis:**

Los datos del experimento factorial se expone en el Cuadro 113.

Cuadro 113. Promedios de producción de cacao de los tratamientos desglosados por factores y niveles

Tratamientos	Código	Factor A	Factor B	Repeticiones					Σ t	X̄ t (qq.ha <sup>-1</sup> )
				I	II	III	IV	V		
1	A1B1	1	1	11	12	9	10	11	53	10,6
2	A1B2	1	2	13	15	12	13	10	63	12,6
3	A1B3	1	3	14	13	14	12	11	64	12,8
4	A2B1	2	1	13	14	16	15	17	75	15,0
5	A2B2	2	2	16	13	17	16	15	77	15,4
6	A2B3	2	3	14	16	16	14	11	71	14,2
7	A3B1	3	1	15	14	15	17	16	77	15,4
8	A3B2	3	2	16	18	17	19	16	86	17,2
9	A3B3	3	3	18	19	17	16	18	88	17,6
Σ R				130	134	133	132	125	654	
X̄ R				14,4	14,9	14,8	14,7	13,9		14,5

Nota: A = 3, B = 3. A.B = 9, r = 5, t = 5(9) = 45

**Hipótesis:**

Sobre las repeticiones:

H<sub>0</sub>: Las repeticiones no tienen efecto significativo sobre la producción

H<sub>1</sub>: En al menos una repetición hay un efecto estadísticamente diferente

Sobre el factor A:

H<sub>0</sub>: La producción en los niveles del factor A resultaron estadísticamente iguales.

H<sub>1</sub>: Al menos en un nivel del factor A hubo diferencias significativas

Sobre el factor B:

H<sub>0</sub>: La producción en los niveles del factor B resultaron estadísticamente iguales.

H<sub>1</sub>: Al menos en un nivel del factor B hubo diferencias significativas

**Regla de decisión:**

Si  $F_{\text{calculada}} < F_{\text{crítico}}$ : Aceptar  $H_0$

Si  $F_{\text{calculada}} \geq F_{\text{crítico}}$ : Aceptar  $H_1$

Las sumas y promedios de la producción de cacao se indican en el Cuadro 114. En el Cuadro 115, se presentan las sumas y medias de las parcelas grandes y de las repeticiones.

Cuadro 114. Sumas y promedios del rendimiento en función de las densidades y frecuencias de riego

Frecuencias de riego	Densidad (Factor B)			$\Sigma$ riego	$\bar{X}$ riego (qq.ha <sup>-1</sup> )
	1	2	3		
1	53	63	64	180,0	12,0
2	75	77	71	223,0	14,9
3	77	86	88	251,0	16,7
$\Sigma$ densidad	205,0	226,0	223,0	654,0	
$\bar{X}$ densidad (qq.ha <sup>-1</sup> )	13,7	15,1	14,9		14,5

Nota: Los datos corresponden a 5 repeticiones

Cuadro 115. Sumas y promedios de las parcelas grandes y de las repeticiones

Tratamientos	Código	Factor A	Factor B	Repeticiones o bloques					Sumas t
				I	II	III	IV	V	
1	A1B1	1	1	11	12	9	10	11	53
2	A1B2	1	2	13	15	12	13	10	63
3	A1B3	1	3	14	13	14	12	11	64
Suma parcela grande A1				38	40	35	35	32	
4	A2B1	2	1	13	14	16	15	17	75
5	A2B2	2	2	16	13	17	16	15	77
6	A2B3	2	3	14	16	16	14	11	71
Suma parcela grande A2				43	43	49	45	43	
7	A3B1	3	1	15	14	15	17	16	77
8	A3B2	3	2	16	18	17	19	16	86
9	A3B3	3	3	18	19	17	16	18	88
Suma parcela grande A3				49	51	49	52	50	
Suma de Parcelas Grandes/bloques				130	134	133	132	125	654

**Cálculos de la suma de cuadrados:**

$$FC = \frac{(\Sigma X)^2}{n \text{ total}} \rightarrow FC = \frac{(654)^2}{45} = 9504,8$$

Suma de cuadrados total:  $SC T = \Sigma X^2 - FC$

$$SC T = [(11)^2 + (12)^2 + (9)^2 + \dots + (17)^2 + (16)^2 + (18)^2] - 9504,8 = 281$$

Suma de cuadrados de bloques:  $SC r = \frac{\Sigma r^2}{t} - FC$

$$SC \text{ Repeticiones} = \frac{(130)^2 + (134)^2 + (133)^2 + (132)^2 + (125)^2}{9} - 9504,8 = 5,64$$

Suma de cuadrados del factor A (riego):  $SC \text{ factor A} = \frac{\Sigma A^2}{rB} - FC$

$$SC A = \frac{(180)^2 + (223)^2 + (251)^2}{15} - 9504,8 = 170,5$$

Suma de cuadrados del factor B (densidad):  $SC B = \frac{\Sigma B^2}{rA} - FC$

$$SC B = \frac{(205)^2 + (226)^2 + (223)^2}{15} - 9504,8 = 17,2$$

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

$$\text{Suma de cuadrados de Parcelas Grandes en A: SC PG} = \frac{\Sigma \text{Parcelas A}^2}{B} - FC$$

$$\text{SC Parcelas grandes} = \frac{(38)^2 + (40)^2 + \dots + (43)^2 + (43)^2 + \dots + (49)^2 + \dots + (50)^2}{3} - 9504,8 = 194,53$$

$$\text{Suma de cuadrados de la interacción AB: SC IA. B} = \frac{\Sigma AB^2}{r} - FC - SCA - SCB$$

$$\text{SC IAB} = \frac{[(53)^2 + (63)^2 + (64)^2 + (75)^2 + (77)^2 + (71)^2 + (77)^2 + (86)^2 + (88)^2]}{5} - 9504,8 - 170,5 - 17,2 = 15,1$$

$$\text{SC Error A} = \text{SC Parcelas grandes} - \text{SCA} - \text{SC repeticiones} \rightarrow = 194,53 - 170,5 - 5,64 = 18,36$$

$$\text{SC Error B} = \text{SC T} - \text{SC Parcelas grandes} - \text{SC B} - \text{SC IA.B} \rightarrow = 281 - 194,53 - 17,2 - 15,1 = 54,17$$

$$\text{SC Error experimental total} = \text{SC T} - \text{SC r} - \text{SC A} - \text{SC B} - \text{SC A.B}$$

$$\text{SCE} = 281 - 202,8 - 5,64 - 15,1 = 72,76$$

El ADEVA del ensayo en diseño en parcela dividida, se indica en el Cuadro 116.

*Cuadro 116. Análisis de la varianza del ensayo en parcela dividida*

FV	GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p	Variación (%)
Bloques r	r - 1	4	5,64	1,41	0,58 ns	2,67	2,67	0,682
Factor A (Riego)	A - 1	2	170,5	85,27	34,80 **	3,29	3,29	0,00000001
Error Factor A	(r - 1)(A - 1)	8	18,36	2,29				
Parcelas grandes A	r.A - 1	14	194,5	13,90	5,67 **	2,01	2,01	0,00002
Factor B (Densidad)	(B - 1)	2	17,20	8,60	3,51 **	3,29	3,29	0,04185
Interacción AB	(A - 1)(B - 1)	4	15,07	3,77	1,54 ns	2,67	2,67	0,21495
Error Factor B	A(r - 1)(B - 1)	24	54,40	2,27				
Subparcelas B	r(A)(B - 1)	30	86,67	2,89	1,18 ns	1,82	1,82	0,3232
Error total	(r - 1)(t - 1)	32	78,40	2,45				27,9
Total	r(A)(B) - 1	44	281,2					100

Nota: ns = no hay diferencia estadística significativa, \*\* = Hay diferencia estadística altamente significativa

$$\text{CV}(\%) = \frac{\sqrt{CME}}{\text{Media}} (100) = \frac{\sqrt{2,45}}{14,5} (100) = 10,77 \%$$

El CV representa a la media y evidencia una adecuada dispersión en los datos.

**Interpretación:**

Los factores riego (p < 0,01) y densidad (p < 0,05) tienen efectos directos sobre la producción. El 60,6 % de la variación está explicada por el riego y el 6,1 % por las densidades. La interacción A.B no resultó significativa (p > 0,05).

**Decisión:**

El riego es un factor clave para favorecer el incremento de la producción (p < 0,01) y explica el 60,6% de la variación fenotípica total.

**6.2.7 EXPERIMENTOS COMPLEJOS**

En la investigación agropecuaria se usa frecuentemente diseños de bloques al azar (BA) y completamente al azar (DCA) con arreglos factoriales que incluyen testigos satelitales. Los ensayos característicos de estos grupos son: AB + 1, AB + 1 + 1 y ABC + 1, ABC + 1 + 1. El ADEVA conjuga análisis de varianza parciales, según el objetivo de estudio.

**6.2.7.1 EXPERIMENTO AB + 1**

Un experimento bifactorial del tipo AB + 1, como es el estudio de dos variedades de un cultivo (V = 2) con tres niveles de fertilización (F = 3) más un testigo (variedad criolla

con fertilización del productor), es del tipo  $2 \times 3 + 1$ . Este ensayo se conforma de 7 tratamientos con r repeticiones. El esquema del ADEVA para este tipo de experimentos es el siguiente:

**Esquema del ADEVA en experimento bifactorial AB + 1:**

Fuente de variación	Grados de libertad
Repeticiones (r = 4)	3
Tratamientos (t = 7)	6
Factor variedades (V = 2)	(1)
Factor fertilización (F = 3)	(2)
Interacción VF	(2)
Testigo vs Resto	(1)
Error experimental	18
Total	27

Un experimento trifactorial ABC + 1 como el estudio de dos variedades de un cultivo determinado (V = 2) en dos densidades poblacionales (D = 2) y tres dosis de fertilización (F = 3) más un testigo local (variedad criolla y con la práctica de fertilización del productor), es del tipo:  $2 \times 2 \times 3 + 1$ . El modelo de ADEVA en diseño trifactorial más un testigo absoluto con r repeticiones (por ejemplo r = 4), se indica a continuación:

**Esquema del ADEVA en experimento trifactorial ABC + 1:**

Fuente de variación	Grados de libertad
Repeticiones (r = 4)	3
Tratamientos (t = 13)	12
Factor variedades (V = 2)	(1)
Factor densidades (D = 2)	(1)
Factor fertilización (F = 3)	(2)
Interacción VD	(1)
Interacción VF	(2)
Interacción DF	(2)
Interacción VDF	(2)
Testigo absoluto vs Resto	(1)
Error experimental	36
Total	51

Nota: Se planea comparar el testigo absoluto vs resto de los tratamientos

6.2.7.2 EXPERIMENTO AB + 1 + 1

La prueba de dos variedades de un cultivo determinado (V = 2) con tres niveles de fertilización (F = 3) más un testigo del agricultor (1), más un testigo comercial (variedad y fertilización de la competencia) corresponde al tipo:  $2 \times 3 + 1 + 1$ . Este ensayo es de ocho tratamientos con r repeticiones (por ejemplo: r = 4).

**El esquema del ADEVA en el experimento AB + 1 + 1:**

Fuente de variación	Grados de libertad
Repeticiones (r = 4)	3
Tratamientos (t = 7)	6
Factor variedades (V = 2)	(1)
Factor fertilización (F = 3)	(2)
Interacción VF	(2)
Testigo absoluto vs Resto	(1)
Testigo comercial vs mejor tratamiento	(1)
Error experimental	18
Total	27

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

El experimento trifactorial: dos variedades ( $V = 2$ ), dos densidades ( $D = 2$ ) y tres dosis de fertilización ( $F = 3$ ) más un testigo local y otro testigo comercial se ajusta al arreglo factorial:  $2 \times 2 \times 3 + 1 + 1$ , es un ensayo de 14 tratamientos con  $r$  repeticiones (ejemplo:  $r = 4$ ). El modelo del ADEVA para un diseño trifactorial más dos testigos es el siguiente:

***Esquema del ADEVA en experimento ABC + 1 + 1***

Fuentes de variación	Grados de libertad
Repeticiones ( $r = 4$ )	3
Tratamientos ( $t = 14$ )	13
Factor variedades ( $V = 2$ )	(1)
Factor densidades ( $D = 2$ )	(1)
Factor fertilización ( $F = 3$ )	(2)
Interacción VD	(1)
Interacción VF	(2)
Interacción DF	(2)
Interacción VDF	(2)
<i>Testigo absoluto vs Resto</i>	(1)
<i>Testigo comercial vs mejor tratamiento</i>	(1)
Error experimental	39
<b>Total</b>	<b>55</b>

**Ejercicio 123:**

Realizar el ADEVA de un experimento  $AB + 1$  con arreglo en bloques al azar con cinco repeticiones: tres variedades de maíz ( $V = 3$ ) x tres alternativas de fertilización ( $F = 3$ ) más un testigo del productor, donde se midió la producción de grano ( $\text{kg} \cdot \text{parcela}^{-1}$ ).

**Datos y análisis:**

Los datos del experimento  $AB + 1$  con cinco repeticiones se indica en el Cuadro 117.

*Cuadro 117. Promedios de producción por parcela del experimento  $AB + 1$*

Tratamientos No.	Factor V Código	Factor F Variedad	Fertilización	Bloques					Suma VF	Media VF $\text{kg} \cdot \text{parcela}^{-1}$
				I	II	III	IV	V		
1	V1F1	1	1	11	12	9	10	11	53	10,6
2	V1F2	1	2	13	15	12	13	10	63	12,6
3	V1F3	1	3	14	13	14	12	11	64	12,8
4	V2F1	2	1	13	14	16	15	17	75	15,0
5	V2F2	2	2	16	13	17	16	15	77	15,4
6	V2F3	2	3	14	16	16	14	11	71	14,2
7	V3F1	3	1	15	14	15	17	16	77	15,4
8	V3F2	3	2	16	18	17	19	16	86	17,2
9	V3F3	3	3	18	19	17	16	18	88	17,6
10	T	Testigo local		9	8	10	9	8	44	8,8
Suma de bloques				139	142	143	141	133	698	139,6
Media de bloques				13,9	14,2	14,3	14,1	13,3	69,8	14,0



**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

El análisis de las interacciones “variedades x fertilización” (IVF), permite determinar los efectos principales de variedades y de la fertilización, donde la variedad V3 es mejor, con un promedio de 16,7 kg.parcela<sup>-1</sup> y la mejor alternativa de fertilización es F2 con un promedio de 15,1 kg.parcela<sup>-1</sup> (Cuadro 118). Al examinar las interacciones, se determinó que V3F3 es la mejor combinación, con una producción promedio de 17,6 kg.parcela<sup>-1</sup>, mientras el testigo produjo 9 kg.parcela<sup>-1</sup>, situación que equivale a una diferencia del 96 %.

*Cuadro 118. Análisis de las interacciones V x F sobre la producción de maíz (kg.parcela<sup>-1</sup>)*

		Alternativas de fertilización				Medias de variedades (kg.parcela <sup>-1</sup> )
		Testigo	F1	F2	F3	
Variedades	Testigo	9,0				9,0
	V1		10,6	12,6	12,8	12,0
	V2		15,0	15,4	14,2	14,9
	V3		15,4	17,2	17,6	16,7
Medias de fertilización (kg.parcela <sup>-1</sup> )		9,0	13,7	15,1	14,9	13,98

**Análisis de varianza:**

En los experimentos AB + 1 se deben realizar dos análisis de varianza. El primero, considerando todos los tratamientos, incluido el testigo absoluto (T) (Cuadro 119), incluyendo la suma de cuadrados de repeticiones, tratamientos, error y total. El segundo ADEVA, toma las sumas de cuadrados de las variedades, fertilización y SC I(V F).

La SC tratamientos del experimento = 340,58

Se calcula la suma de cuadrados de variedades (V), fertilización (F) e IVF.

(SC V+ SC F + SC VF) = 202,8.

La SC “testigo vs resto” se calcula por la diferencia = SC total – (SCV+SCF+ SCVF).

SC “testigo vs resto” = 340,58 – 202,8 = 137,78.

El ADEVA del experimento AB + 1, indica que no hay diferencias estadísticas entre repeticiones y hay alta significación estadística por efecto de los tratamientos. Las variedades son estadísticamente diferentes (p < 0,01). Las alternativas de fertilización muestran diferencias significativas (p < 0,05). No se evidencia interacción V x F. El testigo resultó significativamente diferente del resto de tratamientos (p < 0,01).

*Cuadro 119. Análisis de varianza del experimento AB + 1*

Fuente de variación	GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p
Repeticiones	4	4,11	1,57	0,74 ns	2,634	3,890	0,5693
Tratamientos	9	340,6	37,84	17,90 **	2,153	2,946	0,0001
Factor variedades	(2)	170,5	85,27	40,33 **	3,259	5,248	0,0001
Factor fertilización	(2)	17,2	8,60	4,067*	3,259	5,248	0,0256
Interacción V.F	(4)	15,1	3,77	1,78 ns	2,634	3,890	0,1539
Testigo absoluto vs resto	(1)	137,8	137,78	65,16 **	4,113	7,396	0,0001
Error experimental	36	76,1	2,11				
Total	49	423,0					

CV(%) para el experimento AB + 1 = 10,40. CV (%) para el experimento AB = 10,56

Nota: ns = no hay diferencia estadística significativa, \* = hay diferencia estadística significativa, \*\* = hay diferencia estadística altamente significativa

**Decisión:**

Para variedades y fertilización hay diferencias estadísticas. No hay interacción VF. Hay diferencia altamente significativa entre el testigo vs el resto de los tratamientos.

**Ejercicio 124:**

Realizar el ADEVA del experimento AB + 1 + 1 con cinco repeticiones. En el experimento se prueban V = 3 x F = 3 más un testigo local y un testigo comercial, midiendo como variable de respuesta la producción, en kg.parcela<sup>-1</sup>.

**Datos y análisis:**

En el Cuadro 120, se indica los resultados del experimento AB + 1 + 1 y los cálculos de las sumas y medias de los tratamientos VF y de los testigos por cada uno de los bloques.

*Cuadro 120. Promedios de producción de experimento VF + 1 + 1*

Tratamientos No.	Factor V Código	Factor F Variedad	Fertilización	Bloques					Suma VF	Medias VF kg.parcela <sup>-1</sup>
				I	II	III	IV	V		
1	V1F1	1	1	11	12	9	10	11	53	10,6
2	V1F2	1	2	13	15	12	13	10	63	12,6
3	V1F3	1	3	14	13	14	12	11	64	12,8
4	V2F1	2	1	13	14	16	15	17	75	15,0
5	V2F2	2	2	16	13	17	16	15	77	15,4
6	V2F3	2	3	14	16	16	14	11	71	14,2
7	V3F1	3	1	15	14	15	17	16	77	15,4
8	V3F2	3	2	16	18	17	19	16	86	17,2
9	V3F3	3	3	18	19	17	16	18	88	17,6
10	Testigo local			10	8	10	9	8	45	9,0
11	Testigo comercial			14	13	12	11	14	64	12,8
Suma de bloques				154	155	155	152	147	763	153
Media de bloques				14,0	14,1	14,1	13,8	13,4	69,4	13,9

La media general fue 13,9 kg.parcela<sup>-1</sup>. La mayor producción se obtuvo en la variedad V3 con la fertilización F3, en un segundo nivel se encuentra el tratamiento V3F2. La producción en el testigo comercial fue 12,8 kg.parcela<sup>-1</sup>, similar a V1F3 (Cuadro 121).

*Cuadro 121. Cálculo de las medias de la interacción V.F y testigos*

Variedades	F1	F2	F3	Variedad local (+1)	Variedad comercial (+1)	Medias V (kg.parcela <sup>-1</sup> )
V1	10,6	12,6	12,8	-	-	12,0
V2	15,0	15,4	14,2	-	-	14,9
V3	15,4	17,2	<b>17,6</b>	-	-	<b>16,7</b>
Fertilización local	-	-	-	9,0	-	9,0
Fertilización comercial	-	-	-	-	12,8	12,8
Medias F (kg.parcela <sup>-1</sup> )	13,7	15,1	14,9	9,0	12,8	<b>13,9</b>

**Análisis de varianza de ensayos factoriales**

En el Cuadro 122, se consolida el ADEVA de tres análisis específicos: Los grados de libertad (GL) y suma de cuadrados del experimento global de 11 tratamientos (VF + 1 + 1) con cinco repeticiones proporciona la información básica del total, del error experimental, de las repeticiones y del total de los 11 tratamientos.

El segundo análisis proporciona la información del total de los 10 tratamientos (VF + 1). La suma de cuadrados del testigo comercial se calcula por diferencia, entre los valores totales (SC de 11 t - SC de 10 t). La suma de cuadrados del testigo absoluto se calcula por diferencia, entre los totales de los 10 tratamientos (VF + 1) y el total del factorial (VF).

Cuadro 122. Análisis de varianza del experimento VF + 1 + 1

Fuente de variación	GL	SC	CM	F calculada	F 0,05	F 0,01	p	Variación (%)
Repeticiones	4	4,11	1,03	0,48 ns	2,606	3,828	0,75020	0,94
De 11 tratamientos (t +1 +1)	10	346,91	34,69	16,31**	2,077	2,801	0,00001	-
De 10 tratamientos (t + 1 )	9	299,73	33,30	15,66**	2,124	2,888	0,00001	-
Factorial 3 x3 (t)	8	202,8	25,35	11,92**	2,180	2,993	0,00001	-
Variedades (V)	2	170,5	85,27	40,08**	3,232	5,179	0,00001	39,10
Fertilización (F)	2	17,2	8,60	4,04*	3,232	5,179	0,02523	3,94
Interacción V.F	4	15,1	3,77	1,77 ns	2,606	3,828	0,15394	3,46
Testigo absoluto vs factorial	1	96,9	96,92	45,56**	4,085	7,314	0,00001	-
Testigo comercial vs Resto	1	47,2	47,18	22,18**	4,085	7,314	0,00003	-
Error Experimental	40	85,1	2,13					19,51
Total	54	436,1						100

Nota: ns = no hay diferencia estadística significativa, \* = hay diferencia estadística significativa, \*\* hay diferencia estadística altamente significativa

**Interpretación:**

No hay efecto de las repeticiones ni de la interacción variedades x fertilización. Hay diferencia entre variedades (p<0,01) y entre alternativas de fertilización (p<0,05). Hay diferencias entre los testigos y el resto de tratamientos. La contribución a la variación de la fertilización es reducida, solo el 3,94 %, las variedades contribuyen con 39,1 % y el error aporta con el 19,51 %.

**Decisión:**

Al menos una de las variedades es significativamente mejor en producción (p < 0,01). El factor genético contribuye al incremento significativo de la producción. Se recomienda realizar una prueba *post hoc* para separar las medias y determinar la mejor variedad.

**6.3 DISEÑOS FACTORIALES INCOMPLETOS**

**6.3.1 EXPERIMENTO EN DISEÑO “SAN CRISTÓBAL”**

El uso de diseños experimentales en arreglos factoriales es muy usado por los investigadores de todos los campos del conocimiento. Estos diseños tienen la desventaja que al aumentar los factores en estudio o los niveles se incrementa el número de combinaciones (tratamientos), en muchos casos, imposibilitando la conducción en el campo. Un diseño de este tipo es denominado “San Cristóbal”, los otros se conocen como “métodos Taguchi”.

El diseño San Cristóbal fue propuesto por Basilio Rojas (1962) como un diseño factorial incompleto con un arreglo especial que muestra una mayor zona de exploración comparado con un diseño factorial completo, además, tiene la ventaja adicional de reducir los costos de la investigación.

El “San Cristóbal” es un diseño de segundo orden usado para representar la relación entre dos o tres factores (Villasmil et al., 1974, p. 9), como: X = fertilizantes e Y = producción. En el Cuadro 123, se indica el arreglo factorial para un experimento de nitrógeno y fósforo, que se toma como modelo para explicar este diseño.

Cuadro 123. Factores y niveles en estudio del tipo  $(4)^2$

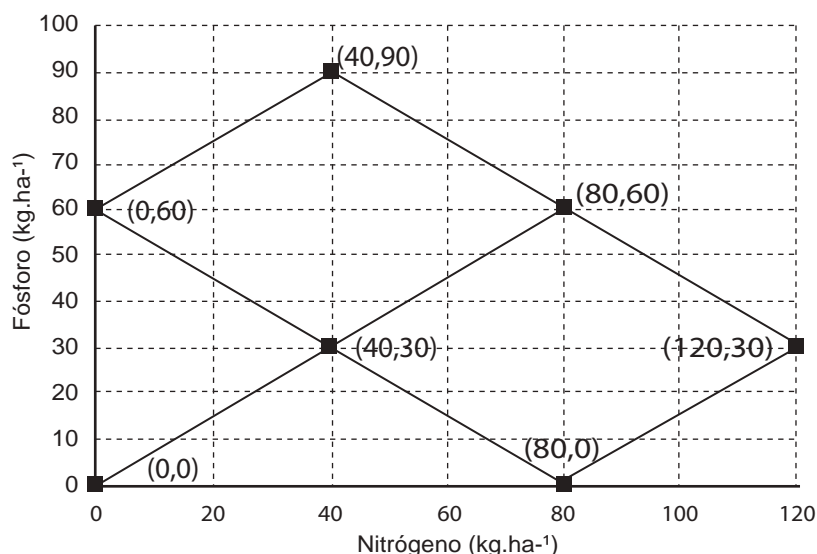
Factores	Niveles			
	1	2	3	4
Dosis de nitrógeno N (Kg.ha <sup>-1</sup> )	0	40	80	120
Dosis de fósforo P <sub>2</sub> O <sub>5</sub> (Kg.ha <sup>-1</sup> )	0	30	60	90

Un ensayo bifactorial  $(4)^2$  tendría 4 niveles de N (N1, N2, N3 y N4) y 4 niveles de P (P1, P2, P3 y P4) que equivale a una combinación de 16 tratamientos. El diseño “San Cristóbal” posibilita reducir el número de tratamientos de 16 a 7 tratamientos, sin perder absolutamente ninguna información.

En un experimento con cuatro repeticiones, por ejemplo, las unidades experimentales requeridas sería 28 y no 64 como en el factorial completo. Esta situación implica una significativa reducción de los costos.

En la Figura 34, se expone la distribución de los tratamientos según el diseño San Cristóbal, donde se exploran los vértices de las combinaciones de los factores y niveles, bajo el supuesto de que se espera una respuesta de segundo orden, no lineal, una expresión característica de estudios con fertilizantes.

La distribución de niveles en los factores N y P<sub>2</sub>O<sub>5</sub> es indicativo de respuesta esperada de segundo orden, basándose en las aristas de la referida figura geométrica. Los datos de producción (g.planta<sup>-1</sup>), en diseños San Cristóbal (Cuadro 124), conllevan directamente al análisis de regresión y no se requiere del ADEVA.



Fuente: Robles (1986, p. 239)

Figura 34. Diseño San Cristóbal para ensayo de dos factores en cuatro niveles

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 124. Producción de un experimento en diseño San Cristóbal de 7 tratamientos con cuatro repeticiones*

Tratamientos	Factores		Repeticiones				Media de tratamientos (g.planta <sup>-1</sup> )
	N	P	1	2	3	4	
1	0	0	10,0	11,0	13,0	10,0	11,0
2	0	60	26,0	22,0	25,0	25,0	24,5
3	80	0	20,0	26,0	23,0	23,0	23,0
4	80	60	26,0	30,0	28,0	31,0	28,8
5	40	30	25,0	32,0	26,0	32,0	28,8
6	120	30	24,0	24,0	29,0	25,0	25,5
7	40	90	26,0	27,0	24,0	27,0	26,0
Medias de repeticiones			22,4	24,6	24,0	24,7	23,9

**Efectos principales:**

La separación de los efectos principales permite conocer las respuestas productivas de las combinaciones de factores y niveles. Con 40 kg.ha<sup>-1</sup> de N se obtiene 27,38 g.planta<sup>-1</sup> y con 30 kg.ha<sup>-1</sup> de P<sub>2</sub>O<sub>5</sub> se obtiene 27,13 g.planta<sup>-1</sup> (Cuadro 125). Este análisis es una aproximación a la identificación de la mejor combinación de fertilizantes y dosis.

*Cuadro 125. Efectos principales de los factores N y P en los distintos niveles en estudio*

	N (kg.ha <sup>-1</sup> )			
	0	40	80	120
Producción gramos.planta <sup>-1</sup>	17,75	<b>27,38</b>	25,88	25,50
	P <sub>2</sub> O <sub>5</sub> (kg.ha <sup>-1</sup> )			
	0	30	60	90
Producción gramos.planta <sup>-1</sup>	17,00	<b>27,13</b>	26,63	26,00

**Interacciones:**

La organización de los datos en una tabla de doble entrada permite valorar las combinaciones de niveles por factor. Si la variable es del tipo “mayor es mejor” se busca el valor más alto y se identifica la mejor combinación más. Se verifica que la mejor respuesta corresponde a la dosis 40 kg.ha<sup>-1</sup> N y 30 kg.ha<sup>-1</sup> P<sub>2</sub>O<sub>5</sub>, logrando 27,3 g.planta<sup>-1</sup> (Cuadro 126).

*Cuadro 126. Interacciones entre los factores N y P (Respuesta en g.planta<sup>-1</sup>)*

Factorial N x P	N0	N40	N80	N120
P0	17,4	22,2	21,4	21,3
P30	22,4	<b>27,3</b>	26,5	26,3
P60	22,2	27,0	26,3	26,1
P90	21,9	26,7	25,9	25,8

En el Gráfico 35, se valora las relaciones entre factores y niveles en estudio, evidenciándose que con N40 hay una tendencia a producciones más altas en comparación de las otras combinaciones.

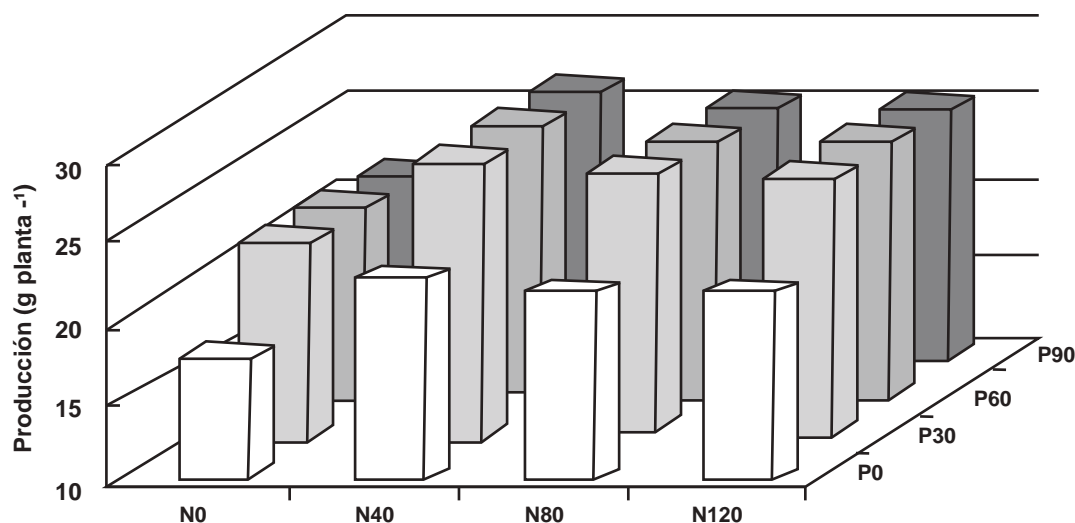


Gráfico 35. Relaciones entre niveles de estudio por los factores nitrógeno y fósforo

El tratamiento N0P0 (testigo) tiene una menor producción que el resto de tratamientos. El comportamiento de N0, N40, N80 y N120 configuran una respuesta no lineal al igual que el fósforo, en las cuatro dosis en estudio. Los experimentos en diseño “San Cristóbal” se recomiendan interpretarlos con la técnica “superficie de respuesta” (Villasmil et al., 1974, p. 8). En el Gráfico 36, se indica la superficie de respuesta, donde se evidencia que la mejor combinación resultó N40 y P30, obteniendo una respuesta de 27,3 g.planta<sup>-1</sup>.

#### Modelos matemáticos N y P:

El modelo matemático se elabora con los datos promedios o con las series completas de datos del experimento<sup>30</sup>. En el ejemplo, la variable dependiente (Y) es la producción y las variables independientes (X1 y X2), constituyen las distintas dosis de los abonos nitrogenado y fosfatado.

En el Gráfico 37 se expone el modelo de la producción en función del nitrógeno:

$$Pr = 21,28 + 0,121 (N) - 0,0008 (N)^2$$

En el Gráfico 38, se indica el modelo de la producción en función de fósforo:

$$Pr = 20,83 + 0,179 (P) - 0,0015(P)^2$$

Dónde:

Pr = Producción en g.planta<sup>-1</sup>

N = Dosis de nitrógeno (kg.ha<sup>-1</sup>)

P = Dosis de fósforo (kg.ha<sup>-1</sup>)

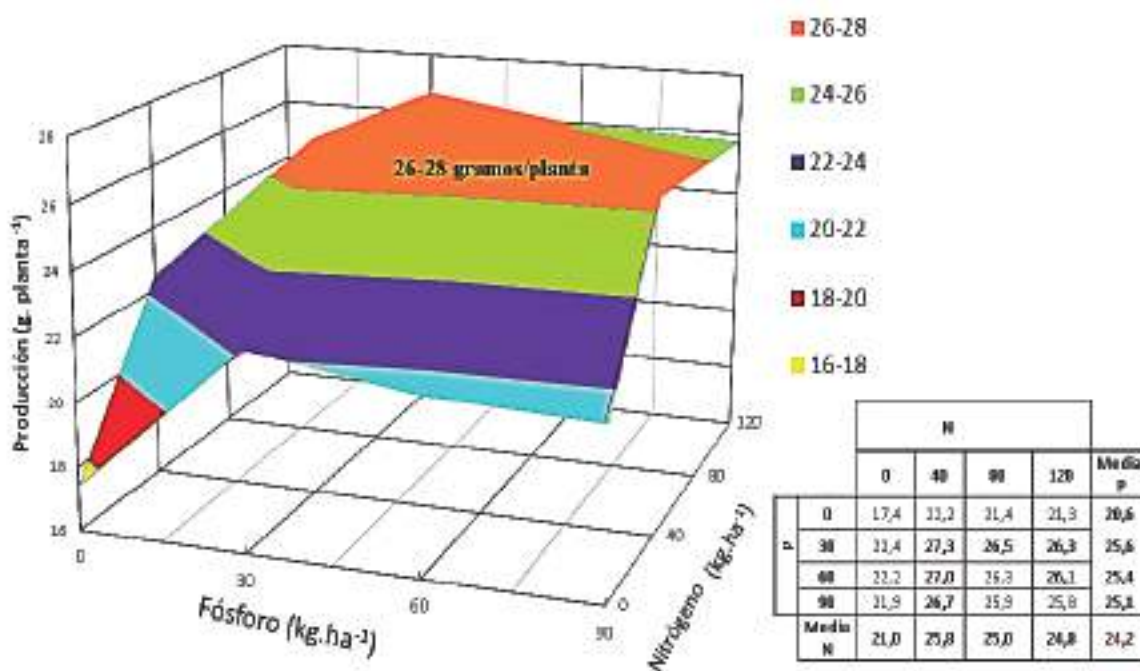


Gráfico 36. Superficie de respuesta a los niveles de nitrógeno y fósforo en la producción

A partir de los modelos matemáticos, apoyado en el cálculo diferencial, se proyecta la maximización de la producción o la optimización de las dosis de insumos con la finalidad de lograr una mejor respuesta biológica o económica.

La superficie de respuesta del ejercicio, permitió identificar combinaciones de los niveles N y P como: N40 con P30, P60 y P90, que posibilitan obtener respuestas de 26 a 28 g.planta<sup>-1</sup>. Conociendo el desempeño productivo, se tiene buscar la combinación N y P económicamente más ventajosa, para lo cual se requiere de información de los costos de los insumos y de los precios de venta del producto (proyectivo) .

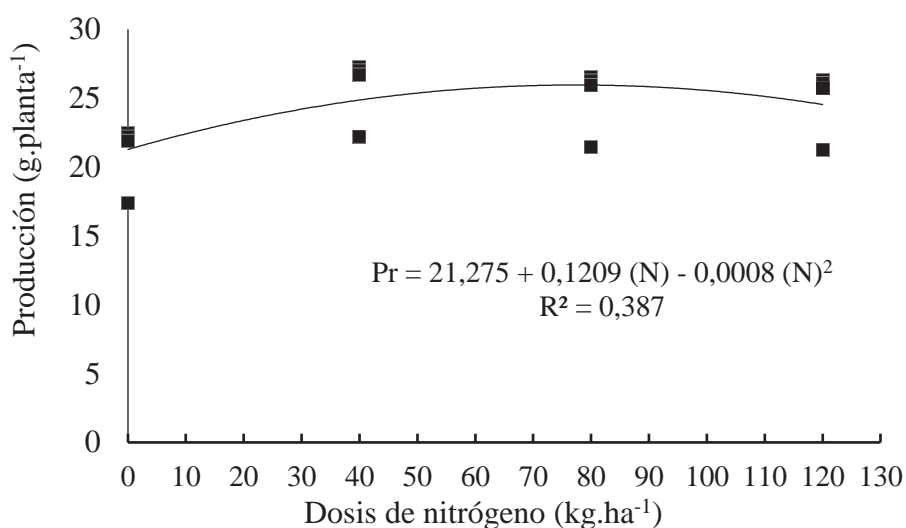


Gráfico 37. Modelo polinómico de grado 2 de la producción en función del nitrógeno

<sup>30</sup> En Excel, a partir del dispersograma y línea de tendencia, se elaboran los modelos gráficos y matemáticos de los tipos lineal, polinómico, exponencial, logarítmico, entre otros.



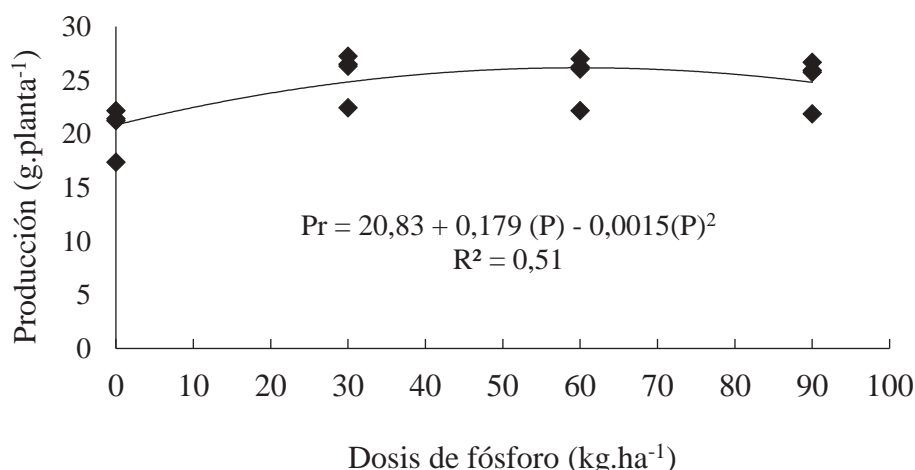


Gráfico 38. Modelo polinómico de grado 2, de la producción en función del fósforo

**Decisión:**

La combinación de 40 kg.ha<sup>-1</sup> de N y 30 kg.ha<sup>-1</sup> de P<sub>2</sub>O<sub>5</sub> posibilita obtener la mejor respuesta productiva.

6.3.2 EXPERIMENTOS MULTIFACTORIALES CON EL MÉTODO TAGUCHI

Los métodos Taguchi son un conjunto de técnicas usadas para mejorar la calidad y reducir la variación en procesos y productos mediante el diseño experimental cuidadoso y la optimización de los factores críticos. Los arreglos ortogonales en el diseño de experimentos implican una ventaja operativa y reducción de costos.

Los diseños factoriales clásicos se tornan complejos en la planeación, operación y análisis por el alto número de unidades experimentales requeridas. Con el diseño ortogonal se reduce el número de tratamientos y se simplifica el análisis estadístico (Báez et al., 2010, p. 69; Hernández et al., 2015, p. 66). Los diseños ortogonales recomendados en la investigación agropecuaria y agroindustrial, corresponden a las series 2<sup>n</sup> y 3<sup>n</sup> como: L<sub>4</sub>(2)<sup>3</sup>, L<sub>8</sub>(2)<sup>7</sup> y L<sub>9</sub>(3)<sup>4</sup> (Gráfico 39), aunque hay diseños multifactoriales más complejos, aplicados en la industria, entre ellos: L<sub>12</sub>(2)<sup>11</sup>, L<sub>16</sub>(2)<sup>15</sup>, L<sub>16</sub>(4)<sup>5</sup>, L<sub>25</sub>(5)<sup>6</sup> y L<sub>27</sub>(3)<sup>13</sup>.

Corridas Experimentales	A	B	C
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

L<sub>4</sub>(2)<sup>3</sup>

Corridas Experimentales	A	B	C	D	E	F	G
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

L<sub>8</sub>(2)<sup>7</sup>

Corridas Experimentales	A	B	C	D
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

L<sub>9</sub>(3)<sup>4</sup>

Gráfico 39. Diseños ortogonales sugeridos para la investigación agropecuaria y agroindustrial



La separación de los efectos por factor en estudio es un desglose de la combinación factorial. El proceso de análisis se conoce como “análisis regular” que comprende 4 pasos : (1) elaboración de la tabla de respuesta, (2) la identificación de la combinación óptima de factores y niveles, (3) la elaboración de gráficas de los efectos factoriales y (4) la predicción de la respuesta óptima esperada. Para la planeación y análisis de diseños ortogonales puede usarse los programas estadísticos: CGG-SOFT, XLSTAT, Qualitek-4 y Minitab.

**Construcción de la tabla de respuesta.-** Los datos de las “corridas experimentales” (tratamientos) y repeticiones, se desglosan por cada factor y nivel de estudio, organizando en una tabla de doble entrada, donde se promedian los efectos factoriales.

**Identificación de la combinación óptima de factores y niveles.-** En la tabla de respuesta se seleccionan los niveles apropiados por cada factor, teniendo en cuenta si la variable es del tipo “mayor es mejor” (p.e.: rendimiento) o del tipo “menor es mejor” (p.e.: defectos físicos).

**Graficación de las respuestas.-** Los datos de la tabla de respuesta se representan en barras, colocando los factores y niveles en el eje X y la variable experimental en el eje Y. Algunos autores usan gráficos con líneas para resaltar la intensidad de las pendientes.

**Predicción de la respuesta óptima.-** La respuesta óptima se puede predecir con el modelo indicado por Báez et al. (2010, p. 72):

$$P_{RO} = \bar{Y} + [(A_M - \bar{Y}) + (B_M - \bar{Y}) + (C_M - \bar{Y}) + \dots]$$

Dónde:

$P_{RO}$  = Predicción de la respuesta óptima

$\bar{Y}$  = Media general del experimento

$A_M, B_M, C_M$  = Factores A, B, C,... con los mejores niveles.

### 6.3.2.1. DISEÑO DE TRES FACTORES CON DOS NIVELES

Un diseño experimental de tres factores en dos niveles  $(2)^3$ , equivale a 8 tratamientos (A x B x C) que se desglosan de la siguiente manera:  $2 \times 2 \times 2 = 8$  (Cuadro 127). Por lo tanto, si se planea probar en dos repeticiones se necesitan 16 unidades experimentales y con tres repeticiones se requieren de 24 parcelas.

Cuadro 127. Diseño factorial  $2^3$  clásico

Tratamientos	Factores en estudio			Código
	A	B	C	
1	1	1	1	A1B1C1
2	1	1	2	A1B1C2
3	1	2	1	A1B2C1
4	1	2	2	A1B2C2
5	2	1	1	A2B1C1
6	2	1	2	A2B1C2
7	2	2	1	A2B2C1
8	2	2	2	A2B2C2

Usando los métodos Taguchi, un experimento de tres factores con dos niveles se ajusta al diseño ortogonal:  $L_4(2)^3$  (Cuadro 128). Esta nomenclatura indica que el experimento ortogonal requiere de cuatro tratamientos o corridas experimentales ( $L_4$ ) para probar tres factores (A, B, C) en dos niveles cada uno.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 128. Modelo de diseño ortogonal  $L_4(2)^3$*

Tratamientos	Factores en estudio			Código del arreglo ortogonal
	A	B	C	
1	1	1	1	A1B1C1
2	1	2	2	A1B2C2
3	2	1	2	A2B1C2
4	2	2	1	A2B2C1

**Ejercicio 125:**

Caracterización de un material biodegradable a partir de desechos cerveceros y micelio del hongo *Pleurotus ostreatus* (Bohórquez y Ramírez, 2020)

**Datos y análisis:**

El objetivo fue determinar un medio biodegradable de menor densidad para el cultivo de hongos comestibles. La variable experimental medida es la “densidad del sustrato biodegradable” para cultivar hongos comestibles:  $\text{kg.m}^3$ , siendo el tipo de variable “menor es mejor”.

Los factores y niveles en estudio para caracterizar el material biodegradable en el cultivo de hongos comestibles se indica en el Cuadro 129.

*Cuadro 129. Factores y niveles, según diseño  $L_4(2)^3$*

Factores en estudio	Niveles	
	1	2
A. Sustrato principal	Afrecho de malta	Polvillo de cebada
B. Proporción de sustrato principal (%)	100	70
C. Tiempo de secado (h)	12	24

El diseño de tratamientos, en los métodos Taguchi, es fijo y se basa en las combinaciones de factores y niveles en un arreglo ortogonal (Cuadro 130). La distribución de los factores y niveles en una matriz predefinida por Taguchi, no son modificables y se considera una ventaja del método (Naranjo et al., 2020, p. 2). Los datos de densidad del sustrato ( $\text{kg.m}^3$ ), registrados se exponen en el Cuadro 131.

*Cuadro 130. Diseño ortogonal  $L_4(2)^3$*

Tratamientos	Factores			Código	Descripción de los factores y niveles		
	A	B	C		A Sustrato principal	B Proporción de sustrato principal (%)	C Tiempo de secado (horas)
1	1	1	1	A1B1C1	Afrecho de malta	100	12
2	1	2	2	A1B2C2	Afrecho de malta	70	24
3	2	1	2	A2B1C2	Polvillo de cebada	100	24
4	2	2	1	A2B2C1	Polvillo de cebada	70	12

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

Cuadro 131. Promedios de la variable de respuesta “densidad sustrato” (kg.m<sup>3</sup>) en función del diseño L<sub>4</sub>(2)<sup>3</sup>

Tratamientos	Código	A Sustrato principal	B Proporción de sustrato (%)	C Tiempo de secado (h)	Repeticiones			Densidad Media (kg.m <sup>3</sup> )
					I	II	III	
1	A1B1C1	1	1	1	409	448	380	412
2	A1B2C2	1	2	2	322	297	327	315
3	A2B1C2	2	1	2	364	403	358	375
4	A2B2C1	2	2	1	212	153	199	188
								$\bar{Y} = 323$

Fuente: Bohórquez y Ramírez (2020, p.15)

**Análisis regular:**

*Tabla de respuesta.*- Se calcula los efectos parciales de los niveles por cada factor. Las medias de las características de calidad expresan los efectos aditivos, en la variable de interés cuya variación resulta de la manipulación en el experimento. La variable en estudio es la “densidad del sustrato” que corresponde al tipo “menor es mejor”.

Análisis del factor A: se estima el efecto principal de A1 y A2, de la siguiente manera:

$$A1 = \frac{412 + 315}{2} = 364 \text{ kg.m}^3 \qquad A2 = \frac{375+188}{2} = 282 \text{ kg.m}^3$$

Análisis del factor B: se estima el efecto principal de B1 y B2, de la siguiente manera:

$$B1 = \frac{412 + 375}{2} = 394 \text{ kg.m}^3 \qquad B2 = \frac{315 + 188}{2} = 252 \text{ kg.m}^3$$

Análisis del factor C: se estima el efecto principal de C1 y C2, de la siguiente manera:

$$C1 = \frac{412 + 188}{2} = 300 \text{ kg.m}^3 \qquad C2 = \frac{315 + 375}{2} = 345 \text{ kg.m}^3$$

Con la información de los efectos por nivel de los factores en estudio, se organiza en la “tabla de respuesta (Cuadro 132). La mejor respuesta se asocia a la menor densidad del sustrato, “menor es mejor”, entonces: A2, B2 y C1 es la combinación óptima.

Cuadro 132. Respuesta de “densidad del sustrato” (kg.m<sup>3</sup>) en función del diseño L<sub>4</sub>(2)<sup>3</sup>

Niveles	Factores en estudio		
	A Sustrato principal	B Proporción de sustrato principal	C Tiempo de Secado
1	364	394	<b>300</b>
2	<b>282</b>	<b>252</b>	345

*Combinación óptima de factores y niveles.*- Como la variable densidad (kg.m<sup>3</sup>) es del tipo “menor es mejor”, la combinación óptima es la A2B2C1, verificada en la tabla de resultados y corresponde al tratamiento 4, que tuvo una media de 188 kg.m<sup>3</sup>, inferior a los promedios de los otros tratamientos. La combinación A2B2C1 es un arreglo ortogonal, donde, la predicción, que es parte del análisis regular, lo verifica. Los datos de las repeticiones sirven para calcular la desviación estándar por tratamiento, donde aquellos que tienen valores más bajos tienen menor variación y por consiguiente son más estables y convenientes.

Representación gráfica de los efectos de cada factor.- Los datos de la tabla de respuesta se grafican usando barras o líneas, para destacar las diferencias entre niveles por factor, así como la intensidad de la pendiente. Como la variable es del tipo “menor es mejor”, en el factor A, el nivel A2 es mejor que A1. En el factor B, el nivel B2 es mejor que B1 y en el factor C, el nivel C1 es mejor que C2 (Gráfico 40).

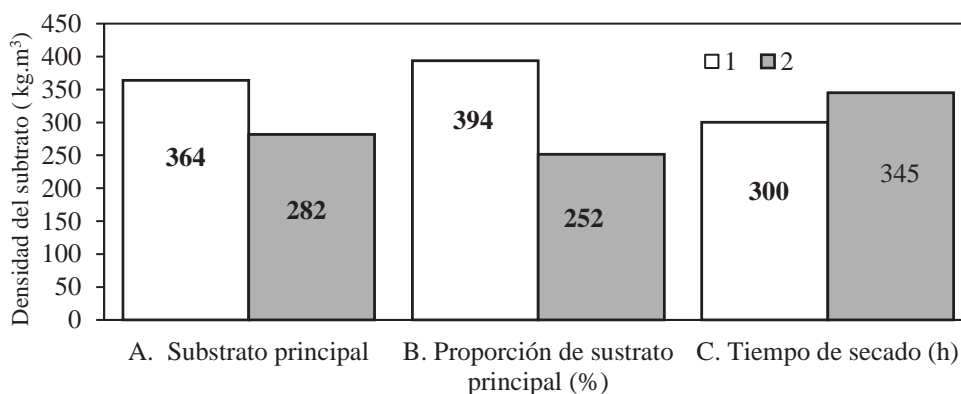


Gráfico 40. Densidad del sustrato para el cultivo de hongos comestibles en función de tres factores y dos niveles de estudio

Predicción de la respuesta óptima.- Para estimar la respuesta óptima, con base a los resultados de niveles/factor, se usa el modelo siguiente:

$$P_{RO} = \bar{Y} + [(A_M - \bar{Y}) + (B_M - \bar{Y}) + (C_M - \bar{Y})]$$

$P_{RO}$  = Predicción de la respuesta óptima

$\bar{Y}$  = Media general del experimento

$A_M, B_M, C_M$  = Factores A, B y C con los mejores niveles:

$$\bar{Y} = 322,68 \text{ kg.m}^3$$

$$A_2 = 281,57 \text{ kg.m}^3$$

$$B_2 = 251,61 \text{ kg.m}^3$$

$$C_1 = 300,20 \text{ kg.m}^3$$

$$P_{RO} = \bar{Y} + [(A_2 - \bar{Y}) + (B_2 - \bar{Y}) + (C_1 - \bar{Y})]$$

$$P_{RO} = 322,68 + [(281,57 - 322,68) + (251,61 - 322,68) + (300,20 - 322,68)]$$

$$P_{RO} = 322,68 + [(-41,11) + (-71,07) + (-22,48)]$$

$$P_{RO} = 188,01 \text{ kg.m}^3$$

### Decisión:

El sustrato biodegradable más adecuado para el cultivo de hongos comestibles resulta de la combinación: Polvillo de cebada (A2), en una proporción del 70 % (B2) y con un tiempo de secado de 12 horas (C1). La densidad del material lograda con la combinación óptima A2B2C1 es de 188,01 kg.m<sup>3</sup>.

## 6. EXPERIMENTO DE SIETE FACTORES EN DOS NIVELES

Un experimento usando los métodos Taguchi corresponde al diseño L<sub>8</sub> (2)<sup>7</sup>, que se refiere a un factorial de 7 factores (ABCDEFG) en dos niveles cada uno. Esto significa: 2 x 2 x 2 x 2 x 2 x 2 x 2 = 128 tratamientos. Si el experimento se probara con r = 3 repeticiones, se requerirán de 384 unidades experimentales.

Con el diseño ortogonal  $L_8(2)^7$  se reduce a 8 el número de tratamientos. En un ensayo con repeticiones ( $r = 3$ ), se requiere de 24 unidades experimentales. Esta reducción de los tratamientos significa una reducción de los costos del experimento, además de las ventajas operativas (Cuadro 133).

Cuadro 133. Diseño ortogonal  $L_8(2)^7$

Tratamientos	A	B	C	D	E	F	G	Código
1	1	1	1	1	1	1	1	A1B1C1D1E1F1G1
2	1	1	1	2	2	2	2	A1B1C1D2E2F2G2
3	1	2	2	1	1	2	2	A1B2C2D1E1F2G2
4	1	2	2	2	2	1	1	A1B2C2D2E2F1G1
5	2	1	2	1	2	1	2	A2B1C2D1E2F1G2
6	2	1	2	2	1	2	1	A2B1C2D2E1F2G1
7	2	2	1	1	2	2	1	A2B2C1D1E2F2G1
8	2	2	1	2	1	1	2	A2B2C1D2E1F1G2

### Ejercicio 126:

En un ensayo  $L_8(2)^7$  se prueban siete factores en dos niveles en el objetivo de mejorar la calidad del café soluble utilizando el método Taguchi (Zapata y Sarache, 2014)

#### Datos y análisis:

El ensayo  $L_8(2)^7$  equivale a probar siete factores (A, B, C, D, E, F y G) en dos niveles cada (Cuadro 134), que en un experimento factorial completo significa 128 tratamientos. Con el método Taguchi se reduce a 8 tratamientos (corridas experimentales) con un arreglo ortogonal (Cuadro 135), en el mismo que se exponen, además los resultados del experimento.

La variable de respuesta es la calidad de taza calificada sobre 100 puntos y corresponde al tipo “mayor es mejor”.

Cuadro 134. Factores y niveles en estudio en un diseño  $L_8(2)^7$

Factores en estudio	Niveles	
	1	2
A. Tipo de secado	Ambiente	Calentador
B. Tiempo de extracción de sólidos solubles	1 hora	2 horas
C. Temperatura de extracción de los sólidos solubles	40°C	50°C
D. Tiempo de concentración	16 horas	24 horas
E. Tiempo de secado del extracto de café concentrado	30 horas	60 horas
F. Tamaño del grano	Medianas	Grandes
G. Relación aroma/buqué	125 (ouE/m <sup>3</sup> )	126 (ouE/m <sup>3</sup> )

Nota: ouE/m<sup>3</sup> es la unidad de recuperación de aromas en el café soluble

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

Cuadro 135. Calidad sensorial del café solubles en función del diseño  $L_8(2)^7$

Tratamiento	Factores en estudio							Código	Repeticiones			Puntaje sensorial
	A	B	C	D	E	F	G		1	2	3	
1	1	1	1	1	1	1	1	A1B1C1D1E1F1G1	50,0	51,1	54,25	51,78
2	1	1	1	2	2	2	2	A1B1C1D2E2F2G2	45,50	48,2	49,50	48,83
3	1	2	2	1	1	2	2	A1B2C2D1E1F2G2	38,0	39,8	40,00	39,25
4	1	2	2	2	2	1	1	A1B2C2D2E2F1G1	37,0	38,3	40,66	38,66
5	2	1	2	1	2	1	2	A2B1C2D1E2F1G2	56,7	57,0	59,50	57,75
6	2	1	2	2	1	2	1	A2B1C2D2E1F2G1	58,5	59,0	60,00	59,15
7	2	2	1	1	2	2	1	A2B2C1D1E2F2G1	45,8	39,7	43,50	42,98
8	2	2	1	2	1	1	2	A2B2C1D2E1F1G2	52,5	51,8	51,00	51,75
								$\bar{Y}$	48,4	48,1	49,8	48,77

**Análisis regular:**

Tabla de respuesta de la calidad de taza.- En el Cuadro 136, se expone el resultado de la separación de los efectos aditivos por factores y niveles.

Cuadro 136. Efecto de los factores y niveles en estudio sobre la calidad de taza

Niveles	A Tipo de secado	B Tiempo de extracción de sólidos solubles	C Temperatura de extracción de sólidos solubles	D Tiempo de concentración	E Tiempo de secado del extracto concentrado	F Tamaño del grano	G Relación aroma/buqué
1	44,6	54,4	48,8	47,9	50,5	50,0	48,1
2	52,9	43,2	48,7	49,6	47,1	47,6	49,4
Diferencia	8,3	-11,2	-0,1	1,7	-3,4	-2,4	1,2
%	18,5	-20,6	-0,3	3,5	-6,8	-4,9	2,6

Fuente: Zapata y Sarache (2014, p.121)

Combinación óptima de factores y niveles

A2B1C1D2E1F1G2

Representación gráfica de los factores y niveles.- En el Gráfico 41, se expone la representación gráfica de los efectos de los factores y niveles en estudio.

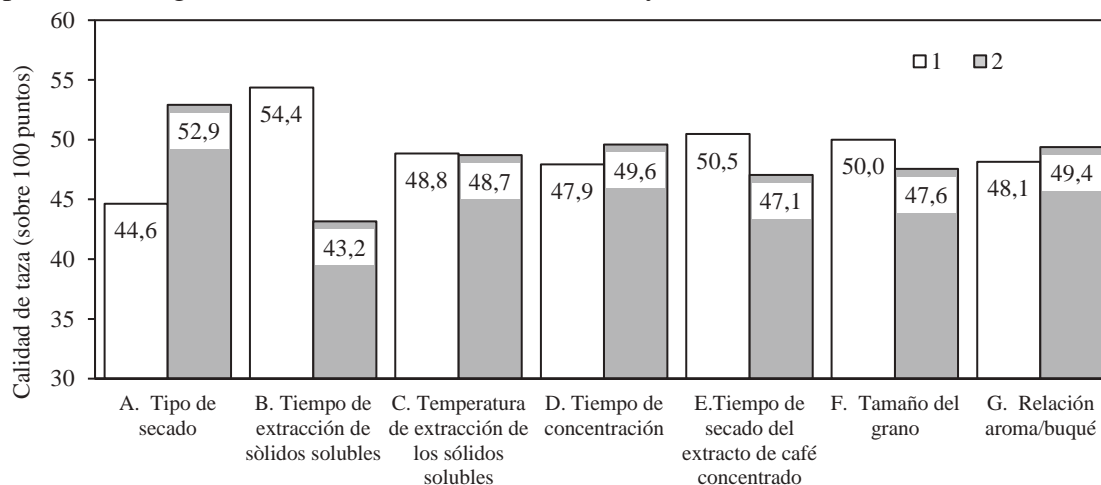


Gráfico 41. Calidad de taza del café soluble en función de siete factores con dos niveles en estudio

Predicción de la respuesta óptima

$$\bar{Y} = 48,77 \text{ puntos}$$

Combinación óptima de factores y niveles: A2B1C1D2E1F1G2

$$A2 = 52,9 \text{ puntos}$$

$$B1 = 54,4 \text{ puntos}$$

$$C1 = 48,8 \text{ puntos}$$

$$D2 = 49,6 \text{ puntos}$$

$$E1 = 50,5 \text{ puntos}$$

$$F1 = 50,0 \text{ puntos}$$

$$G2 = 49,4 \text{ puntos}$$

$$P_{RO} = \bar{Y} + [(A_2 - \bar{Y}) + (B_1 - \bar{Y}) + (C_1 - \bar{Y})(D_2 - \bar{Y}) + (E_1 - \bar{Y}) + (F_1 - \bar{Y}) + (G_2 - \bar{Y})]$$

$$P_{RO} = 48,77 + [(52,9 - 48,77) + (54,4 - 48,77) + (48,8 - 48,77)$$

$$+ [(49,6 - 48,77) + (50,5 - 48,77) + (50,0 - 48,77) + (49,4 - 48,77)]$$

$$P_{RO} = 62,0 \text{ puntos}$$

**Decisión:**

La mejora de la calidad del café soluble se logra secando el grano en el calentador (A2), con un tiempo de extracción de sólidos solubles de 1 hora (B1) a 40°C (C1), un tiempo de concentración de 24 horas (D2), un tiempo de secado de extracto concentrado de 30 horas (E1), un tamaño mediano del grano (F1) y una relación aroma/bouqué de 126 (ouE/m<sup>3</sup>) (G2).

**Ejercicio 127:**

Determinar el efecto de la enzima pectintranseliminasa sobre la fermentación de las almendras de cacao de las variedades CCN-51 y nacional (Lucero, 2021).

**Datos y análisis:**

Los factores y niveles en estudio se indican en el Cuadro 137. Los análisis de la variable “almendras fermentadas”, expresadas en porcentaje, se exponen en el Cuadro 138. El objetivo del experimento fue determinar el efecto de las alternativas de poscosecha sobre la fermentación de las almendras de cacao.

Cuadro 137. Factores y niveles en estudio de un experimento  $L_8(2)^7$

Factores en estudio	Niveles	
	1	2
A. Variedad de cacao	CCN-51	Nacional
B. Momento del beneficio	Inmediatamente después de cosecha	Después de 72 horas de haber cosechado
C. Adición de pectinasa	Dosis 1 mL/5 kg cacao en baba	Dosis 2 mL/5 kg cacao en baba
D. Tiempo de fermentación enzimática	30 minutos	60 minutos
E. Escurrido del mucílago	Escurrido y sin lavado	Escurrido y lavado completo
F. Inicio del secado	Secado sin reposo en poslavado	Reposo 72 horas antes de iniciar secado
G. Método de secado	Directo en tendal de cemento	En marquesina

Fuente: Lucero (2021, p. 19 - 41)

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

Cuadro 138. Almendras bien fermentadas de cacao (%) como respuesta al experimento  $L_8(2)^7$

Tratamientos	Factores en estudio							Código	Repeticiones			Media (%)
	A	B	C	D	E	F	G		1	2	3	
1	1	1	1	1	1	1	1	A1B1C1D1E1F1G1	72,0	70,6	73,4	72,0
2	1	1	1	2	2	2	2	A1B1C1D2E2F2G2	75,5	75,3	73,6	74,8
3	1	2	2	1	1	2	2	A1B2C2D1E1F2G2	76,5	77,5	76,5	76,8
4	1	2	2	2	2	1	1	A1B2C2D2E2F1G1	71,6	72,3	71,6	71,8
5	2	1	2	1	2	1	2	A2B1C2D1E2F1G2	78,9	79,6	79,8	79,4
6	2	1	2	2	1	2	1	A2B1C2D2E1F2G1	77,5	78,7	79,0	78,4
7	2	2	1	1	2	2	1	A2B2C1D1E2F2G1	81,3	82,8	83,2	82,4
8	2	2	1	2	1	1	2	A2B2C1D2E1F1G2	80,6	81,5	80,7	80,9
											$\bar{Y} =$	77,1

**Análisis regular:**

*Tabla de respuesta.*- En el Cuadro 139, se indican los porcentajes de almendras fermentadas en función de los factores y niveles.

Cuadro 139. Almendras de cacao fermentadas (%) en función de los factores/niveles en estudio

Niveles	A	B	C	D	E	F	G
1	73,9	76,2	77,5	77,7	77,0	76,0	76,2
2	80,3	78,0	76,6	76,5	77,1	78,1	78,0
Diferencia	6,4	1,8	-0,9	-1,2	0,1	2,1	1,8
%	8,7	2,4	-1,2	-1,5	0,1	2,7	2,4

*Combinación óptima de factores y niveles*

La combinación óptima fue: A2B2C1D1E2F2G2

*Representación gráfica por factores y niveles.*- En el Gráfico 42, se visualiza el efecto de los factores y niveles en estudio sobre el porcentaje de almendras bien fermentadas.

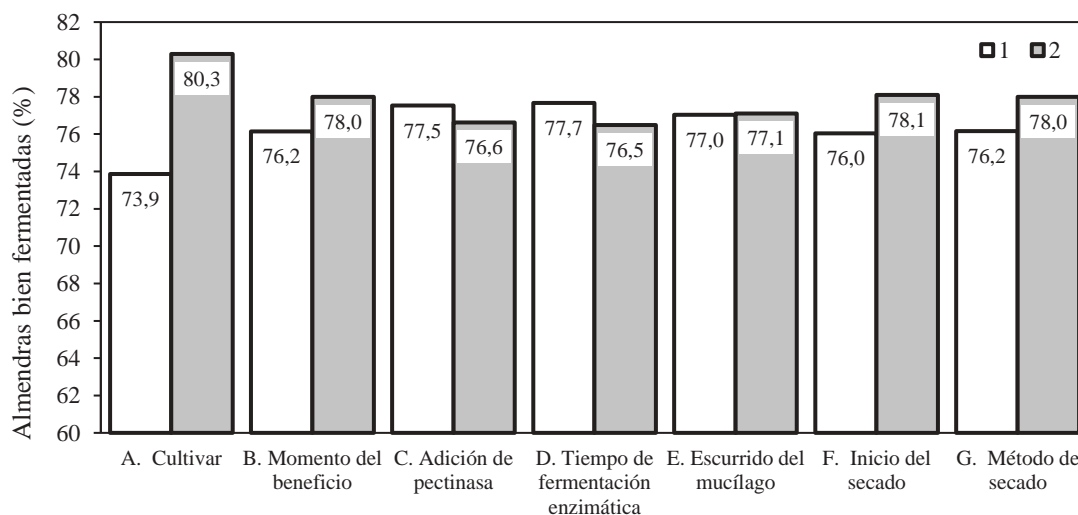


Gráfico 42. Porcentaje de almendras bien fermentadas en función de los factores de la poscosecha



Predicción de respuesta óptima almendras fermentadas (%)

$$\bar{Y} = 77,1 \%$$

Combinación óptima: A2B2C1D1E2F2G2

A2 = 80,3; B2 = 78,0; C1 = 77,5; D1 = 77,7; E2 = 77,1; F2 = 78,1 y G2 = 78,0

$$P_{RO} = \bar{Y} + [(A_2 - \bar{Y}) + (B_2 - \bar{Y}) + (C_1 - \bar{Y}) + (D_1 - \bar{Y}) + (E_2 - \bar{Y}) + (F_2 - \bar{Y}) + (G_2 - \bar{Y})]$$

$$P_{RO} = 77,1 + [(80,3 - 77,1) + (78,0 - 77,1) + (77,5 - 77,1) + (77,7 - 77,1) + (77,1 - 77,1) + (78,1 - 77,1) + (78,0 - 77,1)]$$

$$P_{RO} = 77,1 + (3,2 + 0,9 + 0,4 + 0,6 + 0 + 1 + 0,9)$$

P<sub>RO</sub> = 84,1% almendras fermentadas

**Decisión:**

El tratamiento A2B2C1D1E2F2G2 permitió predecir una respuesta óptima de 84,1% de almendras fermentadas. Este tratamiento corresponde a Cacao Nacional (A2), beneficiado después de 72 horas de la cosecha (B2), aplicando la enzima pectintranseliminasa en dosis 1 mL.5 kg<sup>-1</sup> de cacao en baba (C1), dejando en fermentación enzimática por 30 minutos (D1), con un escurrido y lavado completo del cacao fermentado (E2), manteniendo en reposo por 72 horas antes de iniciar el secado (F2), que se realiza sobre una marquesina (G2).

6.3.2.3 EXPERIMENTO DE CUATRO FACTORES EN TRES NIVELES

En un diseño clásico, un experimento factorial: A x B x C x D, cada uno en tres niveles, equivale a 3 x 3 x 3 x 3 = 81 tratamientos. Si este ensayo estuviera solo en dos repeticiones significa que se requieren de 164 unidades experimentales. Estos estudios a nivel de campo resultan técnicamente no viables.

Los métodos Taguchi fueron desarrollados para reducir el número de tratamientos en un experimento, en base a un arreglo ortogonal, manteniendo la validez del estudio. En el ensayo de 81 tratamientos, con el diseño ortogonal L<sub>9</sub>(3)<sup>4</sup>, se reduce a solo 9 tratamientos combinados (Cuadro 140). La ventaja de este diseño de cuatro factores en tres niveles es la posibilidad de optimizar el uso de insumos (niveles/factor) cuando se evidencia un comportamiento no lineal, tipo polinómico de grados 2 o 3.

Cuadro 140. Diseño ortogonal L<sub>9</sub>(3)<sup>4</sup> de cuatro factores en tres niveles

Tratamientos	Factores y niveles				Código
	A	B	C	D.	
1	1	1	1	1	A1B1C1D1
2	1	2	2	2	A1B2C2D2
3	1	3	3	3	A1B3C3D3
4	2	1	2	3	A2B1C2D3
5	2	2	3	1	A2B2C3D1
6	2	3	1	2	A2B3C1D2
7	3	1	3	2	A3B1C3D2
8	3	2	1	3	A3B2C1D3
9	3	3	2	1	A3B3C2D1

**Ejercicio 128:**

Obtención de pulpa de celulosa a partir de residuos de agaváceas en el propósito de identificar el potencial para elaborar papel tipo artesanal (Jiménez, 2017).

**Datos y análisis:**

La variable experimental es el rendimiento de pulpa (%). La variable es del tipo “mayor es mejor”.

El objetivo es obtener pulpa de celulosa de Agave por el método organosolv (solventes orgánicos). Los factores y niveles en estudio, ajustados al diseño  $L_9(3)^4$ , se exponen en Cuadro 141. En el Cuadro 142, se expone los resultados del experimento, en rendimiento de pulpa de celulosa para papel.

Cuadro 141. Factores y niveles en estudio del diseño  $L_9(3)^4$  para obtener pulpa de celulosa a partir de Agave

Factores en estudio	Niveles		
	1	2	3
A. Agua m/v (%)	60	40	20
B. Tiempo (horas)	2	4	6
C. Ácido acético m/v (%)	40	60	80
D. Ácido clorhídrico m/v (%)	0,5	1,0	1,5

Cuadro 142. Rendimiento de pulpa de celulosa para papel en función de los factores agua (A), tiempo (B) y concentraciones de ácidos acético (C) y clorhídrico (D) en tres niveles

Tratamientos	A	B	C	D	Código	I	II	III	Media de pulpa (%)
1	1	1	1	1	A1B1C1D1	36,67	39,27	33,27	36,40
2	1	2	2	2	A1B2C2D2	34,11	33,1	25,55	30,92
3	1	3	3	3	A1B3C3D3	32,71	36,3	38,26	35,76
4	2	1	2	3	A2B1C2D3	28,58	27,31	26,94	27,61
5	2	2	3	1	A2B2C3D1	35,06	33,57	38,21	35,61
6	2	3	1	2	A2B3C1D2	34,13	36,32	29,69	33,38
7	3	1	3	2	A3B1C3D2	31,48	33,23	32,94	32,55
8	3	2	1	3	A3B2C1D3	34,65	32,71	32,15	33,17
9	3	3	2	1	A3B3C2D1	36,03	29,85	33,86	33,25
$\bar{Y} =$									<b>33,18</b>

**Análisis regular:**

Tabla de respuesta.- En el Cuadro 143, se exponen las respuestas de los factores y niveles sobre la extracción de pulpa de agave.

Cuadro 143. Extracción de pulpa de celulosa en función de cuatro factores en tres niveles de estudio

Niveles	A Agua m/v (%)	B Tiempo (horas)	C Ácido acético m/v (%)	D Ácido clorhídrico m/v (%)	Media (%)
1	34,36	32,19	34,32	35,09	33,99
2	32,20	33,23	30,59	32,28	32,08
3	32,99	34,13	34,64	32,18	33,48
Media (%)	33,18	33,18	33,18	33,18	33,18

Nota: Tipo de variable “mayor es mejor”. En cursiva se indican los mejores niveles por factor

Combinación óptima de factores y niveles

A1B3C3D1

Representación gráfica de efectos por nivel/factor.- Los efectos de factores/niveles, para extraer pulpa de Agave, se expone en el Gráfico 43.

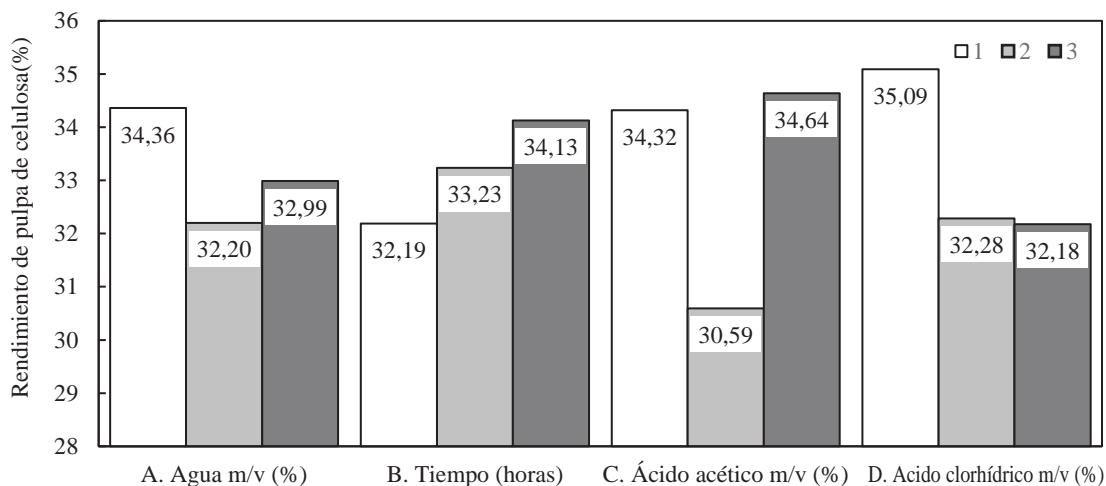


Gráfico 43. Rendimiento de pulpa de celulosa en función de cuatro factores con tres niveles en estudio

Predicción de la respuesta óptima

$$\bar{Y} = 33,18 \%$$

Combinación óptima de factores y niveles: A1B3C3D1, en porcentaje.

$$A1 = 34,36 ; B3 = 34,13; C3 = 34,64 \text{ y } D1 = 35,09$$

$$P_{RO} = \bar{Y} + [(A_1 - \bar{Y}) + (B_3 - \bar{Y}) + (C_3 - \bar{Y}) + (D_1 - \bar{Y})]$$

$$P_{RO} = 33,18 + [(34,36 - 33,18) + (34,13 - 33,18) + (34,64 - 33,18) + (35,09 - 33,18)]$$

$$P_{RO} = 38,67 \%$$

**Decisión:**

El rendimiento potencial para la obtención de pulpa de agave es 38,67 % que se puede lograr con 60 % de agua (A1), en un tiempo de 6 horas (B3), usando ácido acético al 80 % (C3) y ácido clorhídrico al 0,5 % (D1).

6.3.3 ANÁLISIS DE LA VARIANZA EN LOS DISEÑOS TAGUCHI

El método Taguchi integra por un conjunto de técnicas estadísticas y de diseño de experimentos con el objetivo de mejorar la calidad de los productos. Se orientan a tratar experimentos factoriales en diseños  $A^B$ , donde la letra A representa el número de factores en estudio y la letra B el número de niveles por factor (Pardo et al., 2007, p.344).

En los diseños Taguchi, la variabilidad de un producto o proceso se mide a través de la señal/ruido, en el propósito de minimizar los efectos de los factores que no se pueden controlar, llamados factores de ruido<sup>31</sup>. Este procedimiento fundamentalmente se basa en la relación de las medias y de las desviaciones estándar.

<sup>31</sup> El tema del análisis señal/ruido no se aborda en este documento.

**Ejercicio 129:**

Realizar el análisis de la varianza del ensayo sobre “Obtención de pulpa de celulosa a partir de residuos de agaváceas, en procura de identificar el potencial para la elaboración de papel tipo artesanal” (Jiménez, 2017).

**Datos y análisis:**

La variable experimental corresponde a “extracción de pulpa de celulosa (%)” que es del tipo “mayor es mejor”.

Los factores y niveles en estudio fueron:

Factor A: Agua m/v (%) → A1 = 60, A2 = 40, A3 = 20

Factor B: Tiempo (horas) → B1 = 2 horas, B2 = 4 horas, B3 = 6 horas

Factor C: Ácido acético (%) → C1 = 40, C2 = 60, C3 = 80

Factor D: Ácido clorhídrico (%) → D1: 0,5; D2: 1,0; D3: 1,5

Los datos del experimento  $L_9(4)^3$ , se indican en el Cuadro 144. Se destaca que el ensayo correspondería a una combinación factorial:  $3 \times 3 \times 3 \times 3 = 81$  tratamientos.

Cuadro 144. Extracción de pulpa de celulosa de Agave en función de los factores agua (A), tiempo (B), concentración de ácido acético (C) y concentración de ácido clorhídrico (D)

Repeticiones	Tratamientos	Código	A	B	C	D	Extracción de pulpa (%)
1	1	A1B1C1D1	1	1	1	1	36,67
1	2	A1B2C2D2	1	2	2	2	34,11
1	3	A1B3C3D3	1	3	3	3	32,71
1	4	A2B1C2D3	2	1	2	3	28,58
1	5	A2B2C3D1	2	2	3	1	35,06
1	6	A2B3C1D2	2	3	1	2	34,13
	7	A3B1C3D2	3	1	3	2	31,48
1	8	A3B2C1D3	3	2	1	3	34,65
1	9	A3B3C2D1	3	3	2	1	36,03
2	1	A1B1C1D1	1	1	1	1	39,27
2	2	A1B2C2D2	1	2	2	2	33,10
2	3	A1B3C3D3	1	3	3	3	36,30
2	4	A2B1C2D3	2	1	2	3	27,31
2	5	A2B2C3D1	2	2	3	1	33,57
2	6	A2B3C1D2	2	3	1	2	36,32
2	7	A3B1C3D2	3	1	3	2	33,23
2	8	A3B2C1D3	3	2	1	3	32,71
2	9	A3B3C2D1	3	3	2	1	29,85
3	1	A1B1C1D1	1	1	1	1	33,27
3	2	A1B2C2D2	1	2	2	2	25,55
3	3	A1B3C3D3	1	3	3	3	38,26
3	4	A2B1C2D3	2	1	2	3	26,94
3	5	A2B2C3D1	2	2	3	1	38,21
3	6	A2B3C1D2	2	3	1	2	29,69
3	7	A3B1C3D2	3	1	3	2	32,94
3	8	A3B2C1D3	3	2	1	3	32,15
3	9	A3B3C2D1	3	3	2	1	33,86

Fuente: Jiménez (2017, p. 99)

**Análisis estadístico:**

*Estadística descriptiva.*- Los estadígrafos del experimento se indican en el Cuadro 145. Se destaca que la media, mediana y moda tienden a coincidir alrededor de 33 %, la asimetría y la curtosis tienen valores absolutos menores que 0,50 y la proporción de error es 2 %. Por lo tanto, la serie de datos total del experimento para extraer pulpa de celulosa a partir de *Agave*, tiende a mostrar una distribución normal. El experimento, en general, tiene una precisión del 98 %.

Cuadro 145. Estadígrafos de la variable extracción de pulpa de celulosa a partir de *Agave*

Estadísticos	Extracción de pulpa de celulosa	Unidades	Interpretación
Media	33,2	%	Distribución normal
Mediana	33,3	%	
Moda	32,7	%	
Varianza	12,2	(%) <sup>2</sup>	Hay variación aprovechable
Desviación estándar	3,5	%	Reducida dispersión de los datos
Error típico	0,7	%	Reducido error de la media
Coefficiente de variación	10,5	%	Reducida desviación estándar
Variación relativa	2,0	%	Adecuada proporción de error
Precisión	98,0	%	Alta precisión en el experimento
Curtosis	-0,1	índice	Tiende a ser mesocúrtica
Coefficiente de asimetría	-0,4	índice	Tiende a ser simétrica
Mínimo	25,6	%	Valor obtenido más bajo
Máximo	39,3	%	Valor obtenido más alto
Rango	13,7	%	Diferencia entre valores extremos
Intervalo de confianza (95%)	1,3	%	Amplitud de la media poblacional $\mu$
Límite inferior de $\mu$	31,9	%	El parámetro $\mu$ varía desde este punto
Límite superior de $\mu$	34,5	%	El parámetro $\mu$ varía hasta este punto
n	27	datos	Unidades experimentales

La media fue  $\bar{X} = 33,2$  %, la desviación típica  $S = 3,5$  % y el error típico  $EE = 0,7$  %. La variación relativa (VR) fue 2 % que equivale a una precisión del 98 %. La curtosis  $K = -0,1$  y la asimetría  $A = -0,40$  indican, en su orden, que la distribución tiende a ser mesocúrtica y simétrica. Los valores mínimo, máximo y rango explican los puntos extremos. El Intervalo de confianza  $IC_{(1-\frac{\alpha}{2})}$  indica que la media poblacional ( $\mu$ ) varía de 31,9 % a 34,5 %.

*Análisis de varianza.*- El ADEVA se desagrega en comparaciones de los niveles por factor en estudio y la significación estadística se prueba con las comparaciones ortogonales, para desagregar los efectos de niveles por factor<sup>32</sup>. En este ensayo se determinó que las concentraciones de ácido acético 80 % y ácido clorhídrico 0,5 % tienen un efecto directo sobre la variable de extracción de pulpa de celulosa (Cuadro 146).

<sup>32</sup> Los contrastes ortogonales se pueden realizar en una hoja de cálculo de Excel o aplicando los contrastes del programa INFOSTAT

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 146. Análisis de la varianza de la característica “Extracción de pulpa de celulosa” en Agave mediante el método organosolv*

Fuente de variación	GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p de F	Variación %
A1 = A2	(1)	20,97	20,97	2,733 ns	4,414	8,285	0,116	6,12
A2 = A3	(1)	3,59	3,59	0,468 ns	4,414	8,285	0,503	1,05
B1 = B2	(1)	4,93	4,93	0,643 ns	4,414	8,285	0,433	1,44
B2 = B3	(1)	3,59	3,59	0,468 ns	4,414	8,285	0,503	1,05
C1 > C2	(1)	62,46	62,46	8,141 *	4,414	8,285	0,011	18,22
C2 < C3	(1)	73,73	73,73	9,610 **	4,414	8,285	0,006	21,51
D1 > D2	(1)	35,39	35,39	4,613 *	4,414	8,285	0,046	10,32
D2 = D3	(1)	0,05	0,05	0,007 ns	4,414	8,285	0,936	0,01
Error experimental	18	138,10	7,67					40,28
Total	26	342,81						100,00

Nota: ns = no hay diferencia estadística significativa, \* = hay diferencia estadística significativa (p < 0,05) y \*\* = hay diferencia estadística altamente significativa (p < 0,01)

En el Cuadro 147, se resume el análisis, desglosando nivel por factor, donde se evidencia que la mejor combinación corresponde al tratamiento (A1B3C3D1): Agua 60 %, tiempo 6 horas, ácido acético 80 % y ácido clorhídrico 0,5%.

*Cuadro 147. Extracción de pulpa de celulosa en Agave, en función de cuatro factores en tres niveles*

Niveles	A (agua)		B (tiempo)		C (ácido acético)		D (ácido clorhídrico)	
	%	Extracción de pulpa (%)	Horas	Extracción de pulpa (%)	%	Extracción de pulpa (%)	%	Extracción de pulpa (%)
1	60	34,4	2	32,2	40	34,3	0,5	35,1
2	40	32,2	4	33,2	60	30,6	1,0	32,3
3	20	33,0	6	34,1	80	34,6	1,5	32,2
Significación	A1 = A2 (ns)		B1 = B2 (ns)		C1 > C2 (*)		D1 > D2 (*)	
	A2 = A3 (ns)		B2 = B3 (ns)		C2 < C3 (**)		D2 = D3 (ns)	

Nota: ns = no hay diferencia estadística, \* = hay diferencia significativa (p < 0,05) y \*\* = hay diferencia altamente significativa (p < 0,01).

**Decisión:**

Los factores agua y los tiempos no tuvieron efecto significativo. Las concentraciones de ácido acético al 80 % y de ácido clorhídrico al 0,5 % favorecieron la extracción de pulpa.

## 6.4 ANÁLISIS DE COVARIANZA

El análisis de covarianza (ANCOVA) es un procedimiento estadístico que permite eliminar la heterogeneidad causada por una o más variables cuantitativas independientes sobre la variable de respuesta o dependiente (Badii et al., 2008, p. 26). El ANCOVA es un análisis de varianza al que se ha eliminado el efecto predicho por una o más covariables, por regresión lineal múltiple. La inclusión de covariables puede aumentar la potencia estadística porque a menudo reduce la variabilidad. En la experimentación agrosocial, agroforestal, agroindustrial, ambiental y otros tipos que implican estudios de fenómenos naturales, organismos vivos o procesos, frecuentemente ocurren inconvenientes con el campo experimental, unidades experimentales y registro de datos. Una de las formas de realizar los ajustes de datos es el ANCOVA (Gómez & Gómez, 1983, p. 430-433).

El término covarianza significa variación simultánea de dos variables correlacionadas. El análisis permite separar las causas de variación de las variables y de la variación conjunta. El ANCOVA aumenta la potencia de la prueba F, al eliminar el ruido que puede ocasionar una covariable sobre la variable dependiente, permitiendo ajustar las medias. La covariación conjunta de un par de variables aleatorias permite saber si hay entre ellas dependencia; así como la recta de regresión o el coeficiente de correlación lineal. Si la covarianza es positiva significa que los valores altos de una variable tienen correspondencia con los valores altos de la otra variable, ocurriendo la misma tendencia con los valores mínimos. La covarianza resulta negativa cuando los valores más altos de una variable se asocian a los más bajos de la otra variable.

Para reducir el error experimental hay técnicas se usan antes de la aplicación de los tratamientos y de la recolección de datos (“a priori”) y las que se emplean luego de obtener los datos y en el proceso de análisis (“a posteriori”). El ANCOVA es una técnica usada cuando se analiza los datos de la investigación. Concretamente, el ANCOVA consiste en un procedimiento estadístico por el cual se consigue eliminar la heterogeneidad que aparece en la variable que estamos estudiando (siendo ésta una variable dependiente; por ejemplo los niveles de ansiedad), debida a la influencia de una (o más) variables independientes, que son cuantitativas, y a las que llamaremos covariables (por ejemplo terapia en diferentes grados de intensidad).

El ANCOVA puede aplicarse a las muestras y a los experimentos. En los dos casos, se emplean las fórmulas:

$$COV_{XY} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{n-1} \quad \rightarrow \quad COV_{XY} = \frac{\sum (X-\bar{X})^2 (Y-\bar{Y})^2}{n-1}$$

El modelo estadístico del análisis de covarianza (ANCOVA) es el siguiente:

$$Y_{ij} = \mu + \beta(X_{ij} - x_{..}) + \epsilon_{ij}$$

i-ésimo tratamiento  
j-ésimo observación

Las fórmulas de las varianzas de X e Y se indican a continuación:

$$S_x^2 = \frac{1}{n} \sum (X - \bar{X})^2 \quad S_y^2 = \frac{1}{n} \sum (Y - \bar{Y})^2$$

El coeficiente de correlación *r se relaciona* con otros estadísticos de la regresión:

$$r = \frac{COV_{xy}}{\sqrt{S_x^2(S_y^2)}} \quad b = r \left[ \frac{\sqrt{S_y^2}}{\sqrt{S_x^2}} \right] \quad \rightarrow \quad b = r \left( \frac{S_y}{S_x} \right) \quad a = \bar{Y} - b(\bar{X})$$

$$S_{error}^2 = S_y^2(1 - R_{ajustado}^2)$$

$$\hat{Y} = \bar{Y} + b_{yx} (X - \bar{X})$$

Los objetivos del ANCOVA se orientan a eliminar cualquier error sistemático fuera del control del investigador que puede sesgar los resultados y valorar las diferencias en las respuestas inherentes al estudio (Badii et al., 2008, p. 26). Por lo tanto, sirve para controlar el error, aumentar la precisión y ajustar las medias de los tratamientos en la variable dependiente con base a la variable independiente, que deben estar correlacionados significativamente.



### 6.4.1 COVARIANZA EN LA INVESTIGACIÓN PECUARIA

En la investigación pecuaria es frecuente la experimentación con dietas y mezclas de ingredientes. La producción animal depende de la genética de la población, de las condiciones ambientales (limpieza, protección, comodidad) y manejo (sanidad, alimentación y cuidados). Las dietas, suplementos alimenticios y otras prácticas tienen sus efectos sobre la ganancia de peso de los animales, esto significa que el peso final ( $Y_f$ ) tendría como covariable el peso inicial ( $Y_o$ ), que corresponde a la medición antes de aplicar los tratamientos.

Un primer momento del análisis es la valoración de la asociación entre las variables:  $Y_f = f(Y_o)$ . ¿El peso final depende del peso inicial? Una primera aproximación a responder la pregunta es realizar el análisis de correlación. Si el valor  $r$  resulta significativo, entonces se puede continuar con el análisis. En un segundo momento, al suponer que el peso final ( $Y_f$ ) está influenciado por el peso inicial ( $Y_o$ ), se convierte en covariable. Le conjetura que el peso inicial ocasiona un ruido en la plena expresión del peso final ( $Y_f$ ). La separación de esa proporción de efectos de  $Y_o$  sobre  $Y_f$  posibilita “limpiar” la variable de respuesta de los errores originados por la covariable.

El ANCOVA estima los valores de  $Y_f$  corregidos ( $\hat{Y}$ ). Con los valores de  $Y_f$  se realiza el análisis de varianza. Luego del ADEVA con los datos corregidos, de haber significación estadística para los tratamientos, se procederá a la separación de las medias (pruebas *post hoc*), usando la DMS, Tukey, Duncan o comparaciones ortogonales.

#### Ejercicio 130:

En cuatro camadas de cerdos se seleccionan aleatoriamente cinco animales y se determinó el peso inicial ( $P_i$ ), luego se los alimento con distintas dietas y al cabo de un período se registraron los pesos de los animales, individualmente ( $P_f$ ), en  $\text{kg}\cdot\text{animal}^{-1}$ , con el propósito de determinar si el peso inicial es una covariable del peso final.

#### Datos y análisis:

En el ensayo de cerdos se mide el peso por animal en distintas camadas, al inicio y al final del ensayo, así como la ganancia de peso en función de cinco dietas (Cuadro 148).

Cuadro 148. Pesos inicial y final de cerdos de distintas camadas alimentadas con cinco dietas

Camadas	Dietas	Peso inicial en kg por animal ( $P_i$ )	Peso final en kg por animal ( $P_f$ )	Ganancia de peso ( $P_f - P_i$ )
1	1	4	10	6
1	2	2	5	3
1	3	4	8	4
1	4	6	15	9
1	5	4	12	8
2	1	1	5	4
2	2	3	9	6
2	3	2	5	3
2	4	1	8	7
2	5	3	7	4
3	1	2	5	3
3	2	6	12	6
3	3	4	10	6
3	4	6	10	4
3	5	2	8	6
4	1	3	5	2
4	2	1	7	6
4	3	4	6	2
4	4	2	9	7
4	5	1	6	5
	Media	3,05	8,10	5,05
	Suma	61	162	101



**Análisis de la relación entre las variables X e Y:**

El promedio de peso inicial (Pi) por animal fue  $\bar{X} = 3,05$  kg y el promedio de peso final (Pf) fue  $\bar{Y} = 8,10$  kg. Esto equivale a una ganancia de peso de  $5,05 \text{ kg}\cdot\text{animal}^{-1}$ , con coeficientes  $r = 0,73^{**}$  y  $R^2 = 0,53$ .

El modelo matemático se sistematiza en:

$$Pf = 4,36 + 1,23(Pi).$$

La función:  $Pf = f(Pi)$  se expone en el Gráfico 44. El peso inicial (Pi) y las dietas determinan la ganancia de peso, que se valora en la medición del peso final (Pf), cuyo coeficiente de determinación es 53,1 %.

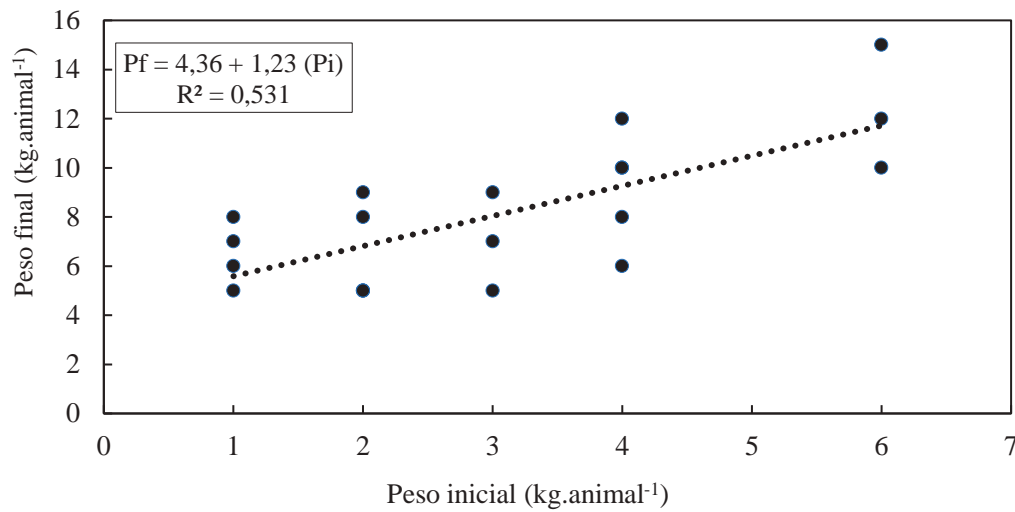


Gráfico 44. Regresión entre pesos inicial y final de cuatro camadas de cerdos

El intercepto,  $a = 4,36$ , no tiene ningún valor biológico, pues no puede haber animal que no tenga peso al nacimiento. Este concepto tiene que ser cuidadosamente tratado cuando se realizan los análisis de regresión.

El coeficiente de la regresión,  $b = 1,23$ , indica que por cada kilo de peso inicial, aumenta el valor en el peso final a razón de 1,23 kg que equivale a 1230 gramos por animal.

En el Gráfico 45, se puede observar el comportamiento del ensayo, destacando que la dieta 4, en las camadas 1 y 4, así como la dieta 2 en las camadas 2 y 3 sobresalieron en peso final con clara influencia de los pesos iniciales.

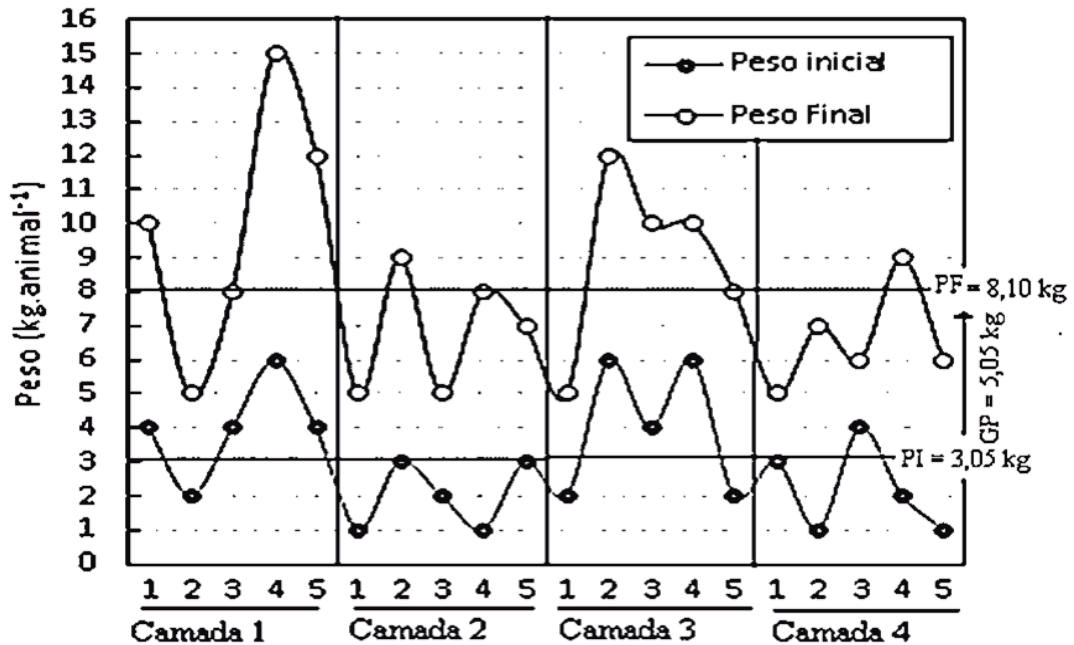


Gráfico 45. Cambios observados entre los pesos inicial y final de cuatro camadas de cerdos tratados con cinco dietas

**Cálculos de las varianzas:**

Varianzas de los pesos inicial (X) y final (Y)

$$S_x^2 = \frac{1}{n} \sum (X - \bar{X})^2 \rightarrow S_x^2 = \frac{1}{20} (53) = 2,65$$

$$S_y^2 = \frac{1}{n} \sum (Y - \bar{Y})^2 \rightarrow S_y^2 = \frac{1}{20} (150) = 7,50$$

**Cálculo de la covarianza:**

$$COV_{xy} = r(S_x \cdot S_y) \rightarrow COV_{xy} = 0,729[(2,737)(1,627)] = 3,25$$

**Cálculo del Coeficiente de correlación (r):**

$$r = \frac{COV_{xy}}{\sqrt{S_x^2(S_y^2)}} \quad r = \frac{3,25}{\sqrt{2,65(7,50)}} = \frac{3,25}{4,458} \quad r = 0,729$$

**Cálculo de Coeficiente de determinación (R<sup>2</sup>):**

$$R^2 = (0,729)^2 = 0,531$$

**Cálculo del Coeficiente de determinación ajustado (R<sup>2</sup>):**

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} [1 - R^2]$$

Dónde:

$\bar{R}^2$  = Coeficiente de determinación ajustado

n = Tamaño de la muestra

k = Número de variables explicativas.

$$\bar{R}^2 = 1 - \frac{20-1}{20-1-1} [1 - 0,531] \rightarrow \bar{R}^2 = 1 - \frac{19}{18} [0,469] \rightarrow \bar{R}^2 = 1 - 1,0556[0,469]$$

$$\bar{R}^2 = 0,505 \sim 50,5 \%$$

**Cálculo de la pendiente de la regresión (b) y del intercepto (a):**

$$b = r \left[ \frac{\sqrt{S_Y^2}}{\sqrt{S_X^2}} \right] \quad b = 0,729 \left[ \frac{\sqrt{7,50}}{\sqrt{2,65}} \right] = 0,729 \left[ \frac{2,737}{1,627} \right] = 1,226$$

$$a = \bar{Y} - b(\bar{X}) \quad \rightarrow \quad a = 8,10 - 1,226(3,05) = 4,36$$

**Predicción  $Y_p$ : peso final corregido (Pfc)**

Se basa en la regresión del peso final observado (Pf) y el peso inicial (Pi = X).

$$Y_p = a + b_{yx}(X) \quad \rightarrow \quad Y_p = 4,36 + 1,23(X_1): \quad \rightarrow \text{Modelo lineal (verificado).}$$

Los resultados de la predicción de  $\hat{Y}$ , usando la fórmula alternativa, son equivalentes:

$$\hat{Y} = \bar{Y} + b_{yx}(\bar{X} - \bar{X}) \quad [109]$$

En el Cuadro 149 se resume los estadísticos asociados al análisis de covarianza.

*Cuadro 149. Estadísticos básicos de la variable peso por animal, al inicio y al final del ensayo*

Camadas	Dietas	Peso inicial (Pi)	Peso final observado (Pf)	Ganancia de peso (Pf - Pi)	Peso final corregido Pfc	Residuo (Pf - Pfc)
1	1	4	10	6	9,26	0,74
1	2	2	5	3	6,81	-1,81
1	3	4	8	4	9,26	-1,26
1	4	6	15	9	11,72	3,28
1	5	4	12	8	9,26	2,74
2	1	1	5	4	5,59	-0,59
2	2	3	9	6	8,04	0,96
2	3	2	5	3	6,81	-1,81
2	4	1	8	7	5,59	2,41
2	5	3	7	4	8,04	-1,04
3	1	2	5	3	6,81	-1,81
3	2	6	12	6	11,72	0,28
3	3	4	10	6	9,26	0,74
3	4	6	10	4	11,72	-1,72
3	5	2	8	6	6,81	1,19
4	1	3	5	2	8,04	-3,04
4	2	1	7	6	5,59	1,41
4	3	4	6	2	9,26	-3,26
4	4	2	9	7	6,81	2,19
4	5	1	6	5	5,59	0,41
	Media	3,05	8,10	5,05	8,10	0,00
	$\Sigma$	61	162	101	162	0

En el Gráfico 46, se valoran las diferencias entre los pesos finales observados (Pf) y las predicciones de peso final (Pfc).

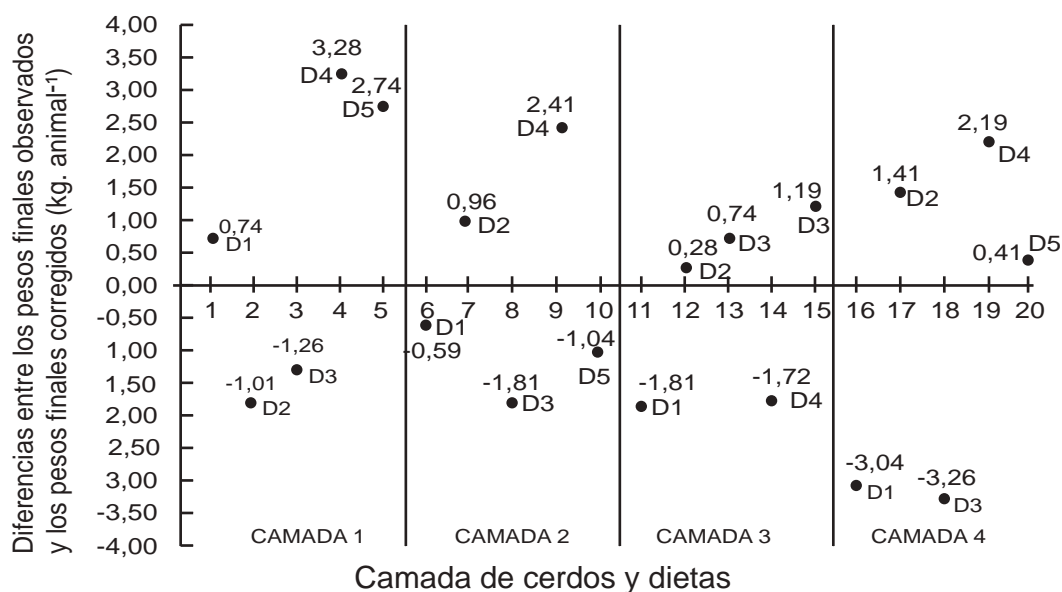


Gráfico 46. Diferencias entre pesos finales observados y pesos finales corregidos por covarianza

**Análisis de varianza con los datos corregidos:**

El ADEVA del peso inicial indica que no hay diferencias estadísticas entre camadas ni entre animales. El ADEVA del peso final indica que hay diferencias estadísticas entre camadas pero no entre dietas. Se constató un efecto significativo de la covariable “peso inicial”, justificándose la corrección por covarianza. El  $CME_{\text{corregido}} = 3,71$  (Cuadro 150).

Cuadro 150. Análisis de varianza de los pesos inicial y final de los animales con la corrección por covarianza

<u>Peso inicial de los animales</u>						
FV	GL	SC	CM	F calculada	p	
Camadas	3	18,15	6,05	2,45 ns	0,1136	
Dietas	4	5,20	1,30	0,53 ns	0,7182	
Error experimental	12	29,60	2,47			
Total	19	52,95				
CV(%)			51,49			
<u>Peso final de los animales</u>						
FV	GL	SC	CM	F calculada	p	
Camadas	3	41,80	13,93	3,95 *	0,0389	
Dietas	4	39,80	9,95	2,82 ns	0,0789	
Covariable peso inicial	1	29,40	29,40	8,34 *	0,0148	
Error experimental	11	38,80	3,53			
Total	19	149,80				
CV(%)			23,19			
<u>Peso final de los animales, corregido por covarianza</u>						
FV	GL	SC	CM	F calculada	p	
Camadas	3	27,22	9,07	2,45 ns	0,1143	
Dietas	4	7,82	1,95	0,53 ns	0,7185	
Error experimental	12	44,53	3,71			
Total	19	79,57				
CV(%)			23,78			

Nota: ns = no hay diferencia estadística, \* = hay diferencia estadística significativa.

La varianza del error se puede estimar con la fórmula:

$$S_{\text{error}}^2 = S_y^2 (1 - R_{\text{ajustado}}^2) \quad [110]$$

$$S_{\text{error}}^2 = 7,50(1 - 0,505) = 7,50(0,495) = 3,71$$

El valor calculado de  $S_{\text{error}}^2$  se aproxima al CME (cuadrado medio del error) calculado para los residuos, en el análisis de la varianza de la regresión. Interpretando el análisis de la varianza de la regresión lineal se determina que, efectivamente, Y depende de X con una probabilidad de equivocarse de 3 casos por cada 10 mil eventos ( $p = 0,0003$ ). La variación, cuyo valor es 53,1 %, corresponde al Coeficiente de determinación  $R^2$  sin ajustar.

### **Decisión:**

El peso final de los animales ( $Y_2$ ) depende del peso inicial (X), por lo tanto, el ajuste por covarianza es pertinente. El análisis de varianza con los datos corregidos evidencia que no hay diferencias estadísticas significativas entre tratamientos.

## 6.4.2 COVARIANZA EN LA INVESTIGACIÓN AGRONÓMICA

En la investigación agronómica se prueban fertilizantes, plaguicidas, variedades, densidades, alternativas de riego, sistemas de cultivos, entre otros. Los objetivos permiten identificar el nivel investigativo, la variable de respuesta, el factor en estudio, el objeto de estudio (cultivo o sistema) y la ubicación. Algunos ejemplos de objetivos: “Comparar el comportamiento productivo de las variedades...en tres densidades poblacionales...”, “Valorar la eficacia de bioinsumos...”

En cada caso surgen cuestiones como: ¿El comportamiento productivo de las variedades puede ser afectado por las densidades?. Se presupone que sí. El ANCOVA “limpia” ese efecto de las densidades sobre la producción. Al elaborar el dispersograma y añadir la línea de tendencia<sup>33</sup> se tiene dos escenarios probables: ¿Cuándo hay cero plantas.parcela<sup>-1</sup>, habrá producción?. No, obvio, por lo tanto no puede haber intercepto. En el otro caso, como los estudios de fertilización edáfica, si no lo aplica si habrá producción, porque siempre hay nutrientes remanentes, por lo tanto, el modelo debe tener intercepto.

En otro momento del análisis, al suponer que la producción (Y) está influenciado por una covariable surge la necesidad de “limpiar la data de errores” y corregir la variable dependiente Y, que corregida se representa como  $\hat{Y}$ . Los análisis estadísticos, incluido el ADEVA se deben realizar con los datos corregidos por covarianza.

### **Ejercicio 131:**

En un ensayo de fertilización de siete tratamientos y seis repeticiones en maíz, se presupone que la producción de grano al 12 % de humedad ( $Y = \text{kg. parcela}^{-1}$ ) tiene como covariable a la densidad poblacional ( $X = \text{plantas. parcela}^{-1}$ ). Verificar si la densidad es una covariable de la producción y corregir los datos de producción por covarianza.

### **Datos y análisis:**

Los resultados del experimento de fertilización en maíz y de su covariable densidad poblacional, donde se midió la producción de grano, se expone en el Cuadro 151.

<sup>33</sup> En el dispersograma que relaciona densidad poblacional (X) y producción observada (Y) se agrega la línea de tendencia, modelo matemático y coeficiente de determinación.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 151. Producción de maíz ( $Y = \text{kg.parcela}^{-1}$ ) en función de la fertilización y densidad poblacional*

Trata- miento	I		II		III		IV		V		VI		$\Sigma X$	$\bar{X}$	$\Sigma Y$	$\bar{Y}$
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y				
1	41	4,08	24	2,78	31	2,79	46	4,24	32	4,17	38	2,62	212	29	20,68	3,32
2	40	4,26	36	4,23	44	5,60	48	6,36	47	4,33	47	4,03	262	38	28,81	4,91
3	37	4,72	32	4,92	38	4,50	41	5,62	42	5,15	40	4,32	230	33	29,23	4,90
4	32	4,25	38	4,53	40	4,83	40	4,30	51	5,43	46	4,52	247	37	27,86	4,72
5	37	4,00	34	5,05	47	5,54	41	6,46	50	6,65	39	5,70	248	36	33,40	5,88
6	42	6,16	38	5,00	34	4,61	40	5,41	42	5,13	49	5,86	245	35	32,17	5,20
7	35	4,59	22	3,63	44	6,20	35	5,47	40	5,16	40	6,07	216	31	31,12	5,31
Sumas	264	32,06	224	30,14	278	34,07	291	37,86	304	36,02	299	33,12	1660	238	203,27	34,24
Medias	38	4,58	32	4,31	40	4,87	42	5,41	43	5,15	43	4,73	237	34	29,04	4,89

**Análisis de la línea de tendencia:**

En el Gráfico 47, se exponen los modelos lineales “sin intercepto” y “con intercepto”. En base del análisis de los valores del Coeficiente de regresión (b) y de los Coeficientes de determinación se selecciona el modelo pertinente.

Aunque las líneas de tendencia son similares, los Coeficientes son distintos. La pertinencia del modelo se valora a través del Coeficiente de determinación  $R^2$ : mientras más cercano a la unidad, el modelo es más adecuado.

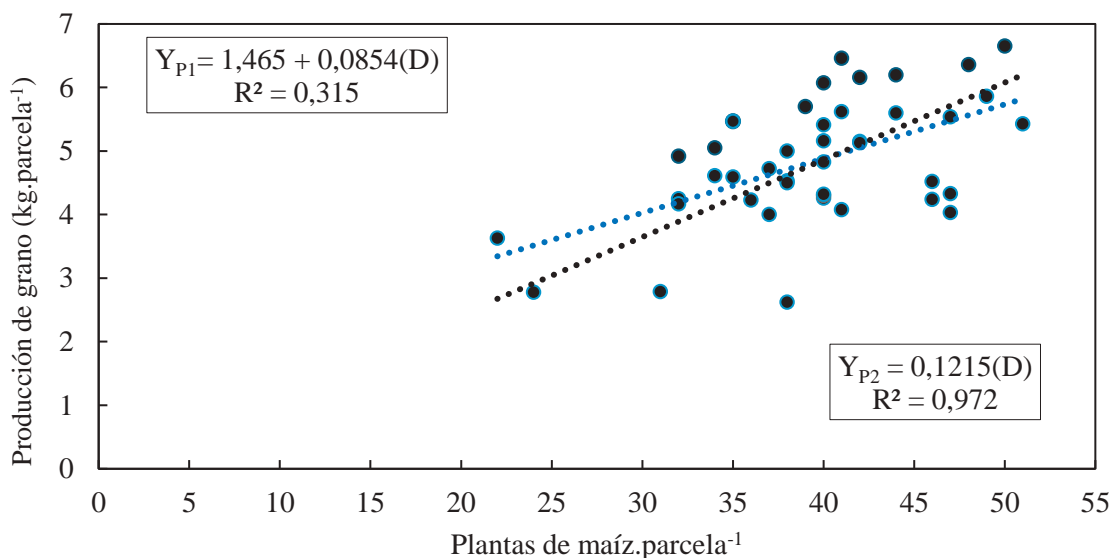
El modelo matemático en el modelo “con intercepto”:  $Y_{P1} = 1,465 + 0,0854 (D)$  con un  $R^2 = 0,315$ , definitivamente no es aplicable<sup>34</sup>. Se interpretaría como: si la densidad poblacional  $X = 0$ , la producción  $Y = 1,465 \text{ kg.parcela}^{-1}$ , lo cual es un absurdo.

El modelo  $Y_{P2} = 0,1215 (D)$ , se interpreta como: si  $X = 0$ ,  $Y = 0$ . No hay plantas no hay producción, lo cual es razonable. El valor de la pendientes  $b = 0,121$  indica que cada planta contribuye al aumento de la producción en  $0,121 \text{ kg.parcela}^{-1}$ . Considerando que la media es  $4,89 \text{ kg.parcela}^{-1}$  y la producción por plantas es  $0,121 \text{ kg}$ , se deduce que la parcela está conformada por 40 plantas:  $(\frac{4,89}{0,121} = 40)$ .

El Coeficiente de determinación  $R^2 = 0,972 \sim 97,23 \%$ , indica que el modelo lineal válido es:  $Y_{P2} = 0,1215 (D)$ . Este resultado se interpreta como: la producción de maíz si depende de la densidad.

El coeficiente de regresión (b) del modelo y el coeficiente de determinación  $R^2$  condicionan la realización del análisis de covarianza. Si en el modelo,  $R^2 > 60 \%$ , se justifica continuar con el análisis de covarianza. En el Cuadro 152, se expone el análisis de regresión lineal entre plantas.parcela<sup>-1</sup> y producción de grano ( $\text{kg.parcela}^{-1}$ ).

<sup>34</sup> Al usar programas estadísticos hay que hacer los ajustes pertinentes al modelo, especificando el intercepto cero, con el razonamiento: en densidad igual a cero la producción será cero. Por tanto, no debe haber intercepto.



Nota: La comparación de los modelos  $Y_{P1}$  y  $Y_{P2}$  se realiza con fines didácticos. El  $Y_{P1}$  no aplica

Gráfico 47. Modelos lineales de la variable producción de grano en función de las plantas.parcela<sup>-1</sup>

Cuadro 152. Análisis de regresión entre densidad y producción de grano por parcela en maíz

Fuente de variación	GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p
Regresión	1	12,02	12,02	18,42 **	4,085	7,314	0,00011
Residuos	40	26,09	0,65				
Total	41	38,11					

Nota: El análisis de regresión lineal indica que hay significación estadística con el 95% de confianza

Un resumen de los estadísticos del ANCOVA, usando las fórmulas expuestas, posibilita la interpretación rápida de la regresión (Cuadro 153). El Coeficiente de determinación  $R^2_{ajustado}$  es prácticamente igual a  $R^2 = 0,972$  porque en la regresión solo intervino una variable independiente. El Coeficiente de regresión  $b = 0,1215$  indica que el aumento de 1 planta por parcela posibilita un incremento de la producción en 121 gramos. De este parte del análisis se deduce, además, que la producción media es 121 g.planta<sup>-1</sup>.

Cuadro 153. Estadísticos de la covarianza entre producción (kg.parcela<sup>-1</sup>) y número de plantas.parcela<sup>-1</sup>

Modelo	n	R <sup>2</sup>	R <sup>2</sup> <sub>ajustado</sub>	b	$\bar{X}$	$\bar{Y}$	S <sub>X</sub> <sup>2</sup>	S <sub>Y</sub> <sup>2</sup>	S <sub>X</sub>	S <sub>Y</sub>	COV <sub>Xy</sub>	S <sub>E</sub> <sup>2</sup>
Y <sub>P2</sub>	42	0,972	0,9716	0,1215	40	4,8	40	0,93	6,34	0,96	3,43	0,0264

**Predicción:**

En el Cuadro 154, se indican las plantas.parcela<sup>-1</sup> (X), la producción observada (Y) y la predicción de la producción (modelo Y<sub>P2</sub>). El promedio  $\bar{X} = 40$  plantas.parcela<sup>-1</sup> y producción de 4,8 kg.parcela<sup>-1</sup> proyectan a los tratamientos 2, 5 y 6 como promisorios.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 154. Plantas.parcela<sup>-1</sup>, medias de producción observada y predichas y residuos en el modelo lineal*

Tratamientos	Plantas.parcela <sup>-1</sup> (X)	Producción observada kg.parcela <sup>-1</sup> (Y <sub>o</sub> )	Producción predicha kg.parcela <sup>-1</sup> (Y <sub>p2</sub> )	Residuos Y <sub>p2</sub>
1	35	3,45	4,29	-0,84
2	44	4,80	5,31	-0,51
3	38	4,87	4,66	0,21
4	41	4,64	5,00	-0,36
5	41	5,57	5,02	0,55
6	41	5,36	4,96	0,40
7	36	5,19	4,37	0,82
Medias	40	4,8	4,8	0,0
Suma	277	34	34	0

**Análisis de varianza para Y<sub>o</sub> y Y<sub>p2</sub>:**

En el Cuadro 155, se indica el ADEVA usando los datos de producción corregidos por la covarianza (covariable densidad).

*Cuadro 155. Análisis de varianza de la producción corregida por la covarianza densidad*

Fuente de variación	GL	SC	CM	F calculada	P
Repeticiones	5	9,25	1,85	5,49 *	0,0110
Tratamientos	6	4,97	0,83	2,46 *	0,0470
T1 < T2	(1)	(3,06)	3,06	9,08 **	0,0052
T2 > T7	(1)	(2,60)	2,60	7,73 **	0,0093
T2 = T3	(1)	(1,25)	1,25	3,72 ns	0,0632
T2 = T4	(1)	(0,27)	0,27	0,81 ns	0,3752
T2 = T5	(1)	(0,24)	0,24	0,71 ns	0,4073
T2 = T6	(1)	(0,36)	0,36	1,06 ns	0,3115
Error experimental	30	10,11	0,34		
Total	41	24,33			

Nota: ns = no hay diferencia significativa, \* = diferencia significativa (p < 0,05) y \*\* = diferencia altamente significativa (p < 0,01).

**Decisión:**

Usando los datos corregidos por covarianza, el tratamiento 2 resultó estadísticamente mayor que T1 y T7 (p < 0,01). Con los otros tratamientos T2 resultó estadísticamente igual.

## 6.5 COMPARACIÓN DE MEDIAS

En el análisis de varianza, al tomar la decisión de aceptar la hipótesis nula (H<sub>0</sub>) se concluye que no hay diferencia estadística significativa. Al aceptar H<sub>1</sub> se está rechazando H<sub>0</sub>, por lo tanto, se concluye que al menos hay un tratamiento estadísticamente diferente. La decisión estadística de aceptar H<sub>1</sub>, se expresa en términos probabilísticos, asociados al valor p. Supóngase, que en un estudio de cuatro genotipos, mediante el ADEVA, se acepta H<sub>1</sub>, “al menos uno de los genotipos es estadísticamente diferente”. El paso siguiente es identificar cuál es la variedad estadísticamente diferente, en la variable de estudio. Esto se conoce como comparación de medias, separación de medias o pruebas *post hoc*.

En la estadística paramétrica y no paramétrica se han desarrollado diversas técnicas *post hoc*, que comparan medias de los tratamientos, como el análisis de los intervalos de confianza de μ, la diferencia mínima significativa (DMS), las pruebas de Tukey, de Duncan y de Newman-Keuls; las comparaciones ortogonales (extensión del ADEVA) y los comparadores estadísticos de Friedman y de Kruskal-Wallis.



6.5.1 USO DE LA PRUEBA “t DE STUDENT” DE UNA MUESTRA

La prueba t de Student se direcciona hacia una cola. La fórmula a usar es la siguiente:

$$t_{\text{calculada}} = \frac{\bar{Y} - A}{\frac{S_Y}{\sqrt{n}}} \quad [111]$$

Dónde:

$t_{\text{calculada}}$  = Estadístico

$\bar{Y}$  = Media de la muestra

A = Valor de referencia o patrón

$S_{\bar{Y}} = \frac{S_Y}{\sqrt{n}}$  = Error típico de la variable Y.

**Ejercicio 132:**

Determinar si amerita la aplicación de insecticida, con el 95% de confianza, cuando en una muestra  $n = 10$ , la infestación media  $\bar{Y} = 17\%$  y  $EE = 1,31\%$ . Se conoce que el umbral de daño económico es  $A = 10\%$ , arriba del cual hay que aplicar insecticida.

**Datos y análisis:**

$\bar{Y} = 17\%$  ,  $EE = 1,31\%$  ,  $A = 10\%$  y  $n = 10$

**Hipótesis:**

$H_0: \bar{Y} \leq A$ . La media muestral es menor o igual que el umbral de daño económico

$H_1: \bar{Y} > A$ . La media muestral es mayor que el umbral de daño económico.

**Regla de decisión:**

Si  $t_{\text{calculada}} < t_{\text{crítico}}$ : Aceptar  $H_0$

Si  $t_{\text{calculada}} \geq t_{\text{crítico}}$ : Aceptar  $H_1$ . Aplicar insecticida

**Cálculo de t:**

$$t_{\text{calculada}} = \frac{\bar{Y} - A}{\frac{S_Y}{\sqrt{n}}} \rightarrow t_{\text{calculada}} = \frac{17 - 10}{1,31} = \frac{7}{1,31} = 5,35$$

Grados de libertad (GL):  $n - 1 = 10 - 1 = 9$

**Contraste de t calculada con los valores críticos:**

En Excel, los valores críticos de t se obtienen con la función:

=DISTR.T.INV(probabilidad; grados de libertad).

**Valor crítico:**

Para  $\alpha = 0,05$  y  $GL = 9$  el valor crítico  $t_{0,05} = 2,685$

**Interpretación:**

$t_{\text{calculada}} = 5,35 > t_{0,05} = 2,685$ : Acepta  $H_1: \bar{Y} > A$ .

**Decisión:**

La media muestral es mayor que el umbral de daño ( $p < 0,05$ ). Aplíquese insecticida.

### 6.5.2 USO DE t PARA SEPARAR MEDIAS INDEPENDIENTES

La hipótesis nula ( $H_0$ ) supone igualdad de dos parámetros poblacionales. Las diferencias numéricas observadas se deben a los efectos insignificantes de los errores aleatorios. Para aplicar la prueba t y comparar dos medias independientes se debe verificar que se cumplan las condiciones paramétricas: independencia entre los grupos, distribución normal y homocedasticidad.

La fórmula de t a usarse en esta prueba es la siguiente:

$$t_{\text{calculada}} = \frac{[\bar{X}_1 - \bar{X}_2]}{\sqrt{\frac{S_{X1}^2}{n_1} + \frac{S_{X2}^2}{n_2}}} \quad [112]$$

Dónde:

$t_{\text{calculada}}$  = Estadístico

$\bar{X}_1$  = Media del tratamiento 1

$\bar{X}_2$  = Media del tratamiento 2

$S_{X1}^2$  = Varianza de grupo 1

$S_{X2}^2$  = Varianza de grupo 2

$n_1$  = Tamaño de muestra en el tratamiento 1

$n_2$  = Tamaño de muestra en el tratamiento 2

Los grados de libertad (GL) se calculan mediante la fórmula:  $GL = n_1 + n_2 - 2$

Las hipótesis estadísticas se formulan como:

$H_0: \bar{X}_1 = \bar{X}_2$

$H_1: \bar{X}_1 \neq \bar{X}_2$

La regla de decisión es la siguiente:

Si  $t_{\text{calculada}} < t_{\text{crítico}}$ : Aceptar  $H_0$

Si  $t_{\text{calculada}} \geq t_{\text{crítico}}$ : Aceptar  $H_1$

#### Ejercicio 133:

En un ensayo de banano, en 12 unidades productivas, se probó una alternativa de fertilización comparado con un testigo local, registrándose el peso de los racimos (libras.racimo<sup>-1</sup>). Comparar las medias de los dos tratamientos, usando la prueba t.

#### Datos y análisis:

Los resultados del ensayo de banano se indican en el Cuadro 156.

Cuadro 156. Promedios de peso de los racimos de banano en función de dos niveles de fertilización compleja

Racimos de banano →	1	2	3	4	5	6	7	8	9	10	11	12	n	Peso promedio (libras.racimo <sup>-1</sup> )	Varianza
FA: Fertilización alternativa	85	78	70	100	75	75	78	85	83	94	80	80	12	81,9	70
FL: Testigo	70	73	45	36	76	70	70	74	75	60	80	32	12	63,4	272

$$t_{\text{calculada}} = \frac{[\bar{X}_1 - \bar{X}_2]}{\sqrt{\frac{S_{X1}^2}{n_1} + \frac{S_{X2}^2}{n_2}}} \quad \rightarrow \quad t_{\text{calculada}} = \frac{[81,9 - 63,4]}{\sqrt{\frac{70}{12} + \frac{272}{12}}} = 3,466$$

**Hipótesis estadísticas:**

H<sub>0</sub>:  $\bar{X}_1 = \bar{X}_2$ . H<sub>0</sub>: 81,9 libras/racimo (FA) = 63,4 libras/racimo (FL)

H<sub>1</sub>:  $\bar{X}_1 \neq \bar{X}_2$ . H<sub>1</sub>: 81,9 libras/racimo (FA)  $\neq$  63,4 libras/racimo (FL)

**Regla de decisión:**

Si  $t_{calculada} < t_{crítico} =$  Aceptar H<sub>0</sub>

Si  $t_{calculada} \geq t_{crítico} =$  Aceptar H<sub>1</sub>

**Valores críticos:**

$$t_{0,05} = 2,074$$

$$GL = n_1 + n_2 - 2 = 12 + 12 - 2 = 22$$

**Interpretación:**

$$t_{calculada} = 3,466 > t_{0,05} = 2,074$$

Aceptar H<sub>1</sub> con el 95 % de confianza.

**Decisión:**

La media de peso del racimo con fertilización alternativa (FA) es significativamente mayor que la del testigo (FL).

**Ejercicio 134:**

En un rodal de una especie arbórea, dividido en lotes A y B, sembrado hace 10 años, se toman n = 56 árboles al azar, donde: n<sub>A</sub> = 21 y n<sub>B</sub> = 35. Se midió el tamaño del fuste (m). Comparar las medias de altura de fuste de los árboles, en los lotes A y B y tomar decisiones con el 99 % de confianza.

**Datos y análisis:**

Los resultados obtenidos en submuestras de los árboles en los rodales A y B, donde se midió el tamaño de fuste, se presenta en el Cuadro 157.

Cuadro 157. Estadísticos básicos del tamaño del fuste (m) en los rodales A y B

Submuestras	N	$\bar{X}$	S <sup>2</sup>	S	EE
A	21	21,78	17,85	4,23	0,92
B	35	20,85	14,43	3,79	0,64

$$t_{calculada} = \frac{[\bar{X}_1 - \bar{X}_2]}{\sqrt{\frac{S_{X1}^2}{n_1} + \frac{S_{X2}^2}{n_2}}} \quad \rightarrow \quad t_{calculada} = \frac{[21,78 - 20,85]}{\sqrt{\frac{17,85}{21} + \frac{14,43}{35}}}$$

$$T_{calculada} = 0,823$$

**Hipótesis estadísticas:**

H<sub>0</sub>:  $\bar{X}_1 = \bar{X}_2$ . H<sub>0</sub>: 21,78 = 20,85

H<sub>1</sub>:  $\bar{X}_1 \neq \bar{X}_2$ . H<sub>1</sub>: 21,78  $\neq$  20,85

**Regla de decisión:**

Si  $t_{calculada} < t_{crítico}$ : Aceptar H<sub>0</sub>

Si  $t_{calculada} \geq t_{crítico}$ : Aceptar H<sub>1</sub>

**Valores críticos:**

$$t_{0,05} = 2,074$$

$$t_{0,01} = 2,819$$

$$GL = n_1 + n_2 - 2 = 21 + 35 - 2 = 54$$

El valor  $p$  se puede obtener directamente en Excel con la función:

=DISTR.T(x; grados de libertad; colas)

Valor  $p = 0,414$

**Interpretación:**

$$t_{\text{calculada}} = 0,823 < t_{0,01} = 2,819: \text{Aceptar } H_0$$

**Decisión:**

Las alturas de fuste de los árboles en los lotes A y B son estadísticamente iguales

6.5.3 COMPARACIÓN DE DOS MEDIAS RELACIONADAS

La “prueba de  $t$ ” para muestras relacionadas, también identificadas como dependientes o pareadas permite comparar las medias de una variable experimental, proveniente de dos series de mediciones relacionadas (Di Rienzo et al., 2009, p. 174). El método compara pares de medias y puede adaptarse para realizar una comparación de varios tratamientos. El procedimiento tiene el propósito de determinar si las diferencias entre las medias de las dos series de datos relacionados variables difieren de 0.

Algunos ejemplos de estudios con muestras relacionadas son: Comparar la tasa fotosintética de 30 plantas en dos momentos distintos: en la mañana y en la tarde. La pregunta de investigación ¿La tasa fotosintética varía según el momento de la medición? Comparar el número de insectos plaga vivos en un cultivo, antes y después de la aplicación de un bioinsecticida. La pregunta de investigación ¿La mortalidad de los insectos plaga depende de la aplicación de bioinsecticida? Comparar la altura de planta (cm) de un híbrido de un cultivo específico, tratado con distintos bioestimulantes del crecimiento, en distintos momentos. Cuantificar el pH del suelo en mediciones antes y después del encalado.

La prueba es aplicable en experimentos longitudinales orientados a cuantificar el crecimiento morfológico, las curvas epidemiológicas, el comportamiento de las arvenses, arbustos o árboles; la evolución de las poblaciones de organismos del suelo y la dinámica de nutrientes en el suelo. En estudios transversales puede aplicarse cuando se relacionan con uso de variedades o aplicación de buenas prácticas agrícolas, en dos o más localidades.

La fórmula usada para este tipo de pruebas es la siguiente:

$$t_{\text{calculada}} = \frac{\bar{d}}{S_{\bar{d}}} \quad [113]$$

Dónde:

$t_{\text{calculada}}$  = Estadístico

$\bar{d}$  = Diferencia de las medias de los tratamientos

$S_{\bar{d}}$  Error estándar de las diferencias

La varianza de las diferencias se calcula con la fórmula

$$S_{\bar{d}}^2 = \frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n - 1}$$

La  $S_d^2$  es la raíz cuadrada de la varianza de las diferencias.

$$S_d = \sqrt{S_d^2}$$

El error estándar de las diferencias se calcula con la fórmula:

$$S_{\bar{d}} = \frac{S_d}{\sqrt{n}}$$

Los grados de libertad  $GL = n - 1$ , donde  $n =$  número de pares de datos.

La hipótesis se formula como:

$$H_0: \bar{X}_2 - \bar{X}_1 = 0$$

$$H_1: \bar{X}_1 - \bar{X}_2 \neq 0$$

La regla de decisión es la siguiente:

Si  $t_{calculada} < t_{crítico}$ : Aceptar  $H_0$

Si  $t_{calculada} \geq t_{crítico}$ : Aceptar  $H_1$

### Ejercicio 135:

En un experimento se compararon las producciones de grano (kg.parcela<sup>-1</sup>) de las variedades A y B de sorgo en 10 fincas. Determinar si la diferencia en las medias de producción de las dos variedades es estadísticamente igual con el 95% de confianza (Reyes, 1980, p. 89).

#### Datos y análisis:

Los datos del ensayo de dos variedades de sorgo en 10 fincas, donde se mide la variable (kg.parcela<sup>-1</sup>), se indican en el Cuadro 158.

Cuadro 158. Producción de grano dos variedades de sorgo en 10 fincas

Fincas	Producción de grano (kg.parcela <sup>-1</sup> )		Diferencias	
	Variedad A	Variedad B	D	d <sup>2</sup>
1	9	8	1	1
2	17	15	2	4
3	14	11	3	9
4	13	11	2	4
5	15	9	6	36
6	10	12	-2	4
7	11	11	0	0
8	13	10	3	9
9	13	9	4	16
10	15	14	1	1
<i>n</i>	<i>10</i>	<i>10</i>	<i>10</i>	
<i>Suma</i>	<i>130</i>	<i>110</i>	<i>20</i>	<i>84</i>
<i>Media</i>	<i>13,0</i>	<i>11,0</i>	<b><i>2,0</i></b>	
<i>Varianza</i>	<i>6,00</i>	<i>4,89</i>	<i>4,89</i>	
<i>Desviación estándar</i>	<i>2,45</i>	<i>2,21</i>	<i>2,21</i>	
<i>Error estándar</i>	<i>0,77</i>	<i>0,70</i>	<i>0,70</i>	

$$t_{calculada} = \frac{\bar{d}}{S_{\bar{d}}} \rightarrow t_{calculada} = \frac{2}{0,70} = 2,86$$

#### Hipótesis estadísticas:

$$H_0: \bar{X}_A - \bar{X}_B = 0 \rightarrow H_0: 13 - 11 \text{ tiende a } 0$$

$H_1: \bar{X}_A - \bar{X}_B \neq 0 \rightarrow H_1: 13 > 11$  son estadísticamente diferentes

**Regla de decisión:**

Si  $t_{\text{calculada}} < t_{\text{crítico}} = \text{Aceptar } H_0$

Si  $t_{\text{calculada}} \geq t_{\text{crítico}} = \text{Aceptar } H_1$

**Valores críticos:**

$t_{0,05} = 2,685$  con  $GL = n - 1 = 9$ .

**Interpretación:**

$t_{\text{calculada}} = 2,86 > t_{0,05} = 2,285$ : Aceptar  $H_1$

El valor  $p = 0,0188$  (obtenido en Excel)

La confianza en la decisión es  $(1 - p) = 0,9812 \sim 98,12 \%$ .

**Decisión:**

La variedad de sorgo VA tiene mayor rendimiento que la variedad VB con el 98 % de confianza.

6.5.4 DIFERENCIA MÍNIMA SIGNIFICATIVA

La diferencia mínima significativa (DMS) también es una prueba *t post hoc* utilizada luego de que en el análisis de la varianza (ADEVA) se determine que hay diferencias estadísticas entre tratamientos (Gómez & Gómez, 1983, p. 188).

Las fórmulas para calcular la DMS en ensayos balanceados y desbalanceados, son:

Para los diseños balanceados:

$$DMS = t_{\frac{\alpha}{2}, GL \text{ del CME}} \sqrt{\frac{2 \text{ (CME)}}{r}} \quad [114]$$

Para los diseños no balanceados:

$$DMS = t_{\frac{\alpha}{2}, GLE} \sqrt{\text{CME} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \quad [115]$$

La formulación de la hipótesis es la siguiente:

$H_0: \bar{X}_1 - \bar{X}_2 = 0$  No hay diferencia estadística significativa

$H_1: \bar{X}_1 - \bar{X}_2 \neq 0$  Hay diferencia estadística significativa

La regla de decisión se formula como:

Si la  $|\bar{X}_1 - \bar{X}_2| < DMS >$  Aceptar  $H_0$ . No hay diferencia significativa

Si la  $|\bar{X}_1 - \bar{X}_2| \geq DMS >$  Aceptar  $H_1$ . Hay diferencia estadística significativa

Cuando hay más de dos medias a compararse se recomienda ordenar de mayor y comparar en orden, para facilitar el procedimiento.

**Ejercicio 136:**

Realizar la prueba DMS para el experimento de cuatro dietas para alimentación de la broca, en un DCA, analizando la variable número de larvas de broca del café/recipiente.

**Datos y análisis:**

$\bar{D}_1 = 46$                        $n_1 = 4$   
 $\bar{D}_2 = 34$                        $n_2 = 2$   
 $\bar{D}_3 = 34$                        $n_3 = 5$   
 $\bar{D}_4 = 42$                        $n_4 = 9$   
 $t = 4$                                $n = 20$   
 $CME = 2,0$   
 $\alpha = 0,05$

**Hipótesis estadísticas:**

$H_0: \bar{X}_1 - \bar{X}_2 = 0$ . Las medias son estadísticamente iguales  
 $H_1: \bar{X}_1 - \bar{X}_2 \neq 0$ . Las medias son estadísticamente diferentes

**Regla de decisión:**

Si la  $|\bar{X}_1 - \bar{X}_2| < DMS$  > Aceptar  $H_0$ . No hay diferencia significativa  
 Si la  $|\bar{X}_1 - \bar{X}_2| \geq DMS$  > Aceptar  $H_1$ . Hay diferencia estadística significativa

Valor  $t_{\alpha/2} = 2,473$

**Cálculo de la DMS:**

$$DMS = T_{\frac{\alpha}{2}, GLE} \sqrt{CME \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$DMS_{1,4} = 2,473 \sqrt{2 \left( \frac{1}{4} + \frac{1}{9} \right)} = 2,102$$

$$DMS_{2,4} = 2,473 \sqrt{2 \left( \frac{1}{2} + \frac{1}{9} \right)} = 2,734$$

$$DMS_{2,3} = 2,473 \sqrt{2 \left( \frac{1}{2} + \frac{1}{5} \right)} = 2,916$$

La separación de medias, usando la prueba DMS, se expone en el Cuadro 159.

Cuadro 159. Comparaciones entre pares de tratamientos basado en la diferencia mínima significativa

Tratamientos	Observaciones	Medias de tratamientos	Diferencias	Comparaciones	DMS <sub>0,05</sub>
D1	4	46	-	-	-
D4	9	42	4 (*)	D1 > D4	2,102
D2	2	34	8 (*)	D4 > D2	2,734
D3	5	34	0 (ns)	D2 = D3	2,926

Nota: \* = hay diferencia estadística significativa, ns = no hay diferencia estadística

**Interpretación:**

La DMS = 2,1 se resta del promedio D1 = 46 y se obtiene un punto teórico de 43,9 hasta donde las medias serían iguales. Como no hay ninguna dieta en este intervalo, en este rango A solo estaría D1 = 46. A la media D4 = 42 se resta 3,03 y se define un segundo rango B que se extiende hasta 38,97. La media de D2 y de D3 es 34 (rango C), lejos de D4. Por lo tanto: D1 > D4 > D2 = D3 (Gráfico 48).

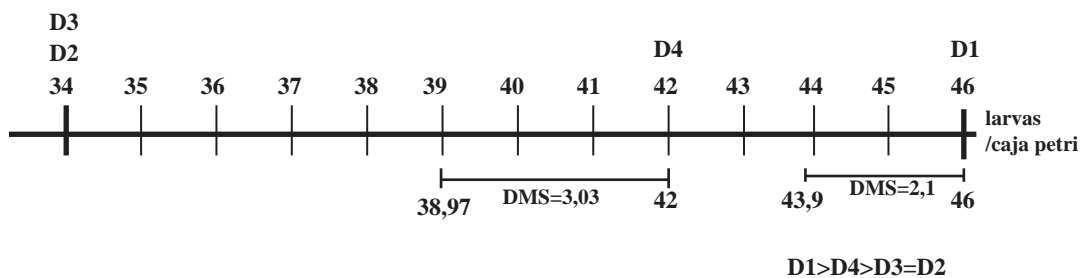


Gráfico 48. Diagrama de la separación de medias usando la DMS en experimentos desbalanceados

**Decisión:**

La dieta D1 es estadísticamente superior a las demás con el 95 % de confianza.

**Ejercicio 137:**

Usando la DMS, separar las medias de producción de forraje (kg.parcela<sup>-1</sup>) de cuatro variedades de sorgo, cinco repeticiones, CME = 2,2 y GLE = 12

**Datos y análisis:**

Los datos de las sumas y de los promedios de las 4 variedades de sorgo forrajero, se exponen en el Cuadro 160.

Cuadro 160. Sumas y promedios de producción de grano de cuatro genotipos de sorgo forrajero

Variedades	Suma	Medias (kg.parcela <sup>-1</sup> )
Honey	172,10	34,42
Milk Maker	173,90	34,78
Silo King	168,50	33,70
Beef Builder	141,90	28,38

En el Cuadro 161, se expone las comparaciones de medias usando la DMS.

**Hipótesis estadísticas:**

H<sub>0</sub>:  $\bar{X}_1 - \bar{X}_2 = 0$

H<sub>1</sub>:  $\bar{X}_1 - \bar{X}_2 \neq 0$

**Regla de decisión:**

Si la  $|\bar{X}_1 - \bar{X}_2| < DMS >$  Aceptar H<sub>0</sub>. No hay diferencia significativa

Si la  $|\bar{X}_1 - \bar{X}_2| \geq DMS >$  Aceptar H<sub>1</sub>. Hay diferencia estadística significativa



$$DMS = T_{\frac{\alpha}{2}, GL \text{ del CME}} \sqrt{\frac{2(CME)}{r}} \rightarrow DMS = 2,56 \sqrt{\frac{2(2,2)}{5}} = 2,4$$

Cuadro 161. Separación de medias de cuatro genotipos de sorgo forrajero usando la DMS

Variedades	Medias	Diferencias	Comparaciones	DMS <sub>0,05</sub>
Milk Maker	34,78	-	-	-
Honey	34,42	0,36 (ns)	Milk Maker = Honey	2,40
Silo King	33,70	1,08 (ns)	Milk Maker = Silo King	2,40
Beef Builder	28,38	6,40 (*)	Milk Maker > Beef Builder	2,40

Nota: ns = no hay diferencia significativa (p > 0,05), \* = si hay diferencia estadística significativa (p < 0,05)

**Interpretación:**

La comparación de medias de los genotipos con la DMS, se expone en el Gráfico 49.

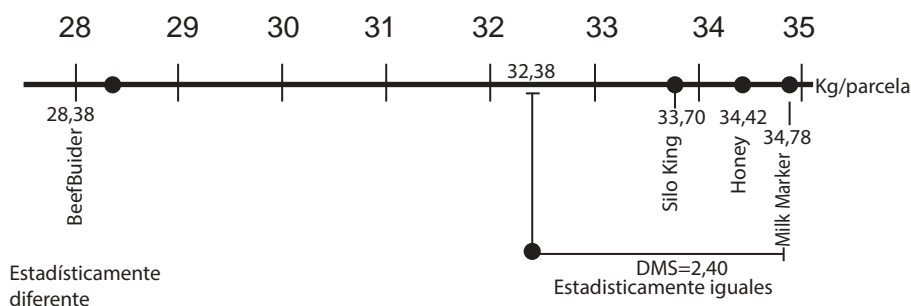


Gráfico 49. Diagrama de la separación de medias de genotipos de sorgo con la DMS

**Decisión:**

En producción, la variedad Milk Maker resultó estadísticamente igual que Honey y Silo King (p > 0,05) pero mayor que la variedad Beef Builder (p < 0,05).

**Ejercicio 138:**

Realizar la separación de medias de las 25 variedades de maíz en un ensayo en diseño LT, usando la DMS, identificando las de mayor rendimiento con un 95 % de confianza.

**Datos y análisis:**

Los promedios de rendimiento de 25 variedades de maíz, probadas en diseño LT se exponen en el Cuadro 162. Además: CME = 41.267, GLE = 16, α = 0,05 y r = 2.

Cuadro 162. Promedios de rendimiento de 25 variedades de maíz ordenados de mayor a menor

Variedades	V10	V13	V5	V9	V8	V11	V6	V7	V19	V3	V1	V18	V4	V17	V25
Media kg.ha <sup>-1</sup>	912	908	900	895	885	828	719	686	624	521	511	488	454	437	434

**Cálculo de la DMS:**

$$DMS = T_{\frac{\alpha}{2}, GL \text{ del CME}} \sqrt{\frac{2(CME)}{r}} \rightarrow DMS = T_{0,025, 16} \sqrt{\frac{2(41267)}{2}} = 2,47(203) = 502$$

**Interpretación:**

El promedio más alto corresponde a la variedad V10 = 912 kg.ha<sup>-1</sup>. Si a este valor se resta la DMS = 502 kg.ha<sup>-1</sup>, da un valor teórico de 410 kg.ha<sup>-1</sup>. Por lo tanto, como todas las variedades tienen rendimientos que varían de 410 a 912 kg.ha<sup>-1</sup>, resultan estadísticamente iguales. La contribución del factor genético a la variación total fue calculada en 64 %.

Caben las preguntas: ¿La media V25 = 434 es igual a V10 = 912 kg.ha<sup>-1</sup>? ¿Habría error tipo I?

**Decisión:**

Las variedades resultaron estadísticamente iguales en producción. Se sugiere reprocesar el análisis usando el análisis de varianza por rangos de Kruskal-Wallis.

6.5.5 PRUEBA DE DUNCAN

La prueba del rango múltiple de Duncan se usa para comparar medias de los tratamientos en experimentos balanceados, después de haber rechazado H<sub>0</sub> en el análisis de la varianza. Esta prueba de medias, típicamente de dos colas, se basa en la distancia que existe entre cada par de tratamientos.

Las fórmulas usadas son las siguientes:

$$CED = T_{\alpha(\text{Medias,GLE})} \sqrt{\frac{CME}{r}} \quad [116]$$

Dónde:

CED = Comparador estadístico de Duncan

T<sub>α(Medias,GLE)</sub> = Valor tabular de rango estudentizado de Duncan

El error estándar se calcula a partir de la varianza del error (CME) con r = repeticiones y un nivel de significación α, usando la fórmula:

$$S_{\bar{x}} = \sqrt{\frac{CME}{r}} \quad [117]$$

Las hipótesis se formulan como:

H<sub>0</sub>:  $\bar{X}_1 - \bar{X}_2 = 0$ . Las medias son estadísticamente iguales

H<sub>1</sub>:  $\bar{X}_1 - \bar{X}_2 \neq 0$ . Las medias son estadísticamente diferentes

La regla de decisión es la siguiente:

Si  $|\bar{X}_1 - \bar{X}_2| < CED$ : Aceptar H<sub>0</sub>. No hay diferencia estadística significativa

Si  $|\bar{X}_1 - \bar{X}_2| \geq CED$ : Aceptar H<sub>1</sub>. Hay diferencia estadística significativa

**Ejercicio 139:**

Separar las medias de un experimento en BA de cuatro genotipos de sorgo con cinco repeticiones, usando la prueba de Duncan (α = 5 %), donde se midió el rendimiento de heno (forraje seco) en kg.parcela<sup>-1</sup> (Reyes, 1980, p. 135).

**Datos y análisis:**

Genotipos = 4, repeticiones = 5, CME = 2,2 y GLE = 75.

**Cálculo de CED:**

$$CED = t_{\alpha(\text{Medias,GLE})} \sqrt{\frac{CME}{r}}$$

$$S_{\bar{X}} = \sqrt{\frac{CME}{r}} \quad \rightarrow S_{\bar{X}} = \sqrt{\frac{2,2}{5}} \quad \rightarrow S_{\bar{X}} = 0,663$$

**Valores críticos estudentizados de Duncan:**

	$T_{\alpha(\text{Medias,GLE})}$	$S_{\bar{X}}$	CED
$t_{0,05}$ (4 medias, 75 GLE)	3,08	0,663	2,04
$t_{0,05}$ (3 medias, 75 GLE)	2,98	0,663	1,98
$t_{0,05}$ (2 medias, 75 GLE)	2,83	0,663	1,88

**Hipótesis estadísticas:**

$H_0: \bar{X}_1 - \bar{X}_2 = 0$ . Las medias son estadísticamente iguales

$H_1: \bar{X}_1 - \bar{X}_2 \neq 0$ . Las medias son estadísticamente diferentes

**Regla de decisión:**

Si  $|\bar{X}_1 - \bar{X}_2| < CED$ : Aceptar  $H_0$

Si  $|\bar{X}_1 - \bar{X}_2| \geq CED$ : Aceptar  $H_1$

En el Cuadro 163, se indican las diferencias entre medias de los genotipos, los valores del comparador estadístico de Duncan (CED), las diferencias entre tratamientos y los rangos de Duncan ( $\alpha = 0,05$ ). Si a la media de  $F1 = 20,2 \text{ t.ha}^{-1}$  se le resta el  $CED = 2,04$ , se tiene el valor 18,16 que indica el punto (teórico) hasta el cual compartirían el mismo rango.

La diferencia entre 18,2 t (F2) y 20,2 t (F2) tienen el rango A, por tanto, son estadísticamente iguales. Si al genotipo  $F2 = 18,2$  se resta el valor  $CED = 1,98$  da el valor 16,2, por lo tanto, comparte el rango B con el progenitor 2 = 17,2  $\text{t.ha}^{-1}$ . Si a F2 se le resta  $CED = 1,88$  se obtiene el valor 15,32. Los progenitores 1 y 2 comparten el rango C.

Cuadro 163. Separación de medias de producción de forraje de genotipos de sorgo con la prueba de Duncan

Genotipos de sorgo	Producción de forraje seco	Diferencias entre genotipos ( $\text{t.ha}^{-1}$ )				CED	Diferencia Media - CED	Rangos
		F1	F2	Progenitor 2	Progenitor 1			
	Media	20,2	18,2	17,2	16,0			
F1	20,2	0,0	2,0	3,0	4,2	2,04	18,16	A
F2	18,2		0,0	1,0	2,2	1,98	16,22	B
Progenitor 2	17,2			0,0	1,2	1,88	15,32	B C
Progenitor 1	16,0				0,0			C

Nota: Letras iguales en los rangos de Duncan significa que no hay diferencia estadística

**Interpretación:**

El genotipo  $F1 > F2$  ( $p > 0,05$ ). Los genotipos  $F2 = P2$  y  $P1 = P2$  ( $p > 0,05$ ).

**Decisión:**

El genotipo F1 produce más forraje que los otros ( $p > 0,05$ ).

### 6.5.6 PRUEBA DE TUKEY

La prueba de Tukey<sup>35</sup> se usa para separar pares de medias de un conjunto de tratamientos usando el comparador estadístico (CTk), para discriminar las diferencias.

Las fórmulas usadas son las siguientes:

$$CTk = q_{\alpha(\text{tratamientos,GLE})} \sqrt{\frac{CME}{r}} \quad [118]$$

Dónde:

CTk = Comparador estadístico de Tukey

$q_{\alpha(\text{tratamientos,GLE})}$  = Valor tabular de amplitud estudentizada de Tukey

$S_{\bar{X}} = \sqrt{\frac{CME}{r}}$  Error estándar de la media calculada a partir de la varianza del error (CME)

r = número de repeticiones

$\alpha = 0,05$

Para facilitar el análisis, se recomienda ordenar las medias de los tratamientos de mayor a menor. El procedimiento e interpretación es similar a la prueba de Duncan.

#### Ejercicio 140:

Separar las medias de cuatro genotipos de sorgo usando la prueba de Tukey ( $\alpha = 5\%$ ), en un experimento en BA de 5 repeticiones, donde se midió el rendimiento de forraje seco en kg.parcela<sup>-1</sup> (Reyes, 1980, p. 135).

#### Datos y análisis:

Genotipos = 4, r = 5, CME = 2,2, GLE = 75 y  $\alpha = 0,05$

#### Cálculo de CTk:

$$CTk = q_{\alpha(\text{tratamientos, GLE})} \sqrt{\frac{CME}{r}}$$

$$S_{\bar{X}} = \sqrt{\frac{CME}{r}} \rightarrow S_{\bar{X}} = \sqrt{\frac{2,2}{5}} = 0,663$$

El valor  $q_{\alpha}$  para 4 tratamientos, 75 GLE y  $\alpha = 0,05$ , obtenida en la tabla de amplitud estudentizada de Tukey es:  $q_{\alpha} = 3,73$ .

Por lo tanto:  $CTk = 3,73 (0,663) = 2,47$

#### Hipótesis:

$H_0: \bar{X}_1 - \bar{X}_2 = 0$ . Las medias son estadísticamente iguales

$H_1: \bar{X}_1 - \bar{X}_2 \neq 0$ . Las medias son estadísticamente diferentes

#### Regla de decisión:

Si  $|\bar{X}_1 - \bar{X}_2| < CTk$ : Aceptar  $H_0$

Si  $|\bar{X}_1 - \bar{X}_2| \geq CTk$ : Aceptar  $H_1$

<sup>35</sup> En el Anexo 7 se indica el procedimiento de lectura en la tabla de amplitud estudentizada para obtener los valores q de Tukey

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

Los valores CED y los rangos de Tukey ( $\alpha = 0,05$ ) se indican el Cuadro 164. Si a la media de “F1” = 20,2 t.ha<sup>-1</sup> se resta el CTK = 2,48, se obtiene un punto = 17,72. Entonces; las medias que se encuentren entre 17,72 y 20,2 compartirán el rango A de Tukey y se declaran que son estadísticamente iguales. Continuando, si al genotipo F2 = 18,2, se resta el CTK = 2,48, se tendría el punto 15,72, En este caso, el “progenitor 1”, el “progenitor 2” y la F2, están compartiendo el rango B, por lo tanto, son estadísticamente iguales.

*Cuadro 164. Separación de medias de producción de forraje seco de cuatro genotipos usando la prueba Tukey*

Genotipos de sorgo	Producción de forraje seco Medias →	Diferencias entre tratamientos (t.ha <sup>-1</sup> )				CTk	Diferencia t – CTK	Rangos
		F1	F2	P2	P1			
		20,2	18,2	17,2	16,0			
F1	20,2	0,0	2,0	3,0	4,2	2,47	17,72	A
F2	18,2		0,0	1,0	2,2	2,47	15,72	A B
Progenitor P2	17,2			0,0	1,2	2,47		B
Progenitor P1	16,0				0,0			B

Nota: Letras iguales en los rangos de Tukey significa que no hay diferencia estadística

**Decisión:**

Las medias de los genotipos F1 y F2 resultaron estadísticamente iguales, con mayor rendimiento de forraje que los otros genotipos ( $p < 0,05$ ).

**Ejercicio 141:**

Realizar la separación de medias, usando la prueba de Tukey, para el experimento AB + 1 + 1, referido a 3 variedades x 3 alternativas de fertilización y dos testigos.

**Datos y análisis:**

La prueba de Tukey se realizó con el programa INFOSTAT (Cuadro 165), donde se aprecia que V3F3, V3F2, V2F2, V3F1 y V2F1 comparten el rango A (son estadísticamente iguales). El traslape de V3F2 = V2F3 y V2F2 = V1F2 enmascara los efectos.

*Cuadro 165. Separación de medias del experimento bifactorial, variedades x fertilización, con la prueba Tukey*

Tratamientos	Código	Medias (kg.parcela <sup>-1</sup> )	Rangos de Tukey ( $\alpha = 0,05$ )
9	V3F3	17,60	A
8	V3F2	17,20	AB
5	V2F2	15,40	ABC
7	V3F1	15,40	ABC
4	V2F1	15,00	ABC
6	V2F3	14,20	BC
3	V1F3	12,80	CD
11	Testigo comercial	12,80	CD
2	V1F2	12,60	CD
1	V1F1	10,60	DE
10	Testigo absoluto	9,00	E
	Media	13,90	

Nota: Medias con una letra común no son estadísticamente diferentes

**Decisión:**

Las variedades de mayor rendimiento fueron V3 y V2. Los niveles de fertilización F3 y F2 fueron mejores. El traslape entre rangos genera incertidumbre.

### 6.5.7 PRUEBA DE NEWMAN-KEULS

La prueba Newman-Keuls es un procedimiento de comparaciones múltiples escalonadas utilizada para identificar medias significativamente diferentes. Se debe aplicar luego de haber rechazado la hipótesis de nulidad ( $H_0$ ) en el ADEVA. Esta prueba de dos colas es similar a la prueba de Tukey y se diferencia por usar otros valores críticos.

Para calcular el comparador estadístico de Newman-Keuls (CNK) en ensayos balanceados se usa la fórmula:

$$CNK = q_{(\alpha, \text{tratamientos}, GLE)} \sqrt{\frac{CME}{n}} \quad [119]$$

Para calcular CNK en experimentos no balanceados se usa la fórmula:

$$CNK = q_{(\alpha, \text{tratamientos}, GLE)} \sqrt{\frac{CME}{2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \quad [120]$$

Dónde:

CNK = Comparador estadístico de Newman-Keuls

CME = Cuadrado medio del error

$$S_{\bar{x}} = \sqrt{\frac{CME}{n}} = \text{Error típico}$$

$q_{\alpha(\text{tratamientos}, GLE)}$  = valor tabular de amplitud estudentizada

$n$  = número de repeticiones

$n_1$  y  $n_2$  = observaciones de cada una de las medias comparadas

$\alpha = 0,05$

Al usar la prueba de Newman-Keuls (CNK) se recomienda ordenar las medias de los tratamientos de mayor a menor para agilizar el procedimiento.

Las hipótesis estadísticas se formulan como:

$H_0: \bar{X}_1 - \bar{X}_2 = 0$ . Las medias son estadísticamente iguales

$H_1: \bar{X}_1 - \bar{X}_2 \neq 0$ . Las medias son estadísticamente diferentes

La regla de decisión se plantea de la siguiente manera:

Si  $|\bar{X}_1 - \bar{X}_2| < CNK$ : Aceptar  $H_0$

Si  $|\bar{X}_1 - \bar{X}_2| \geq CNK$ : Aceptar  $H_1$

#### Ejercicio 142:

Separar las medias de los rendimientos de forraje seco ( $\text{kg} \cdot \text{ha}^{-1}$ ) de 4 genotipos de sorgo usando la prueba de Newman-Keuls con el 95 % de confianza, de un ensayo en BA con 5 repeticiones (Reyes, 1980, p. 135).

#### Datos y análisis:

Genotipos = 4,  $r = 5$ ,  $CME = 2,2$ ;  $GLE = 75$  y  $\alpha = 0,05$

#### Cálculo de CNK:

$$CNK = q_{\alpha}(\text{tratamientos}; GLE) \sqrt{\frac{CME}{r}}$$

$$S_{\bar{x}} = \sqrt{\frac{CME}{n}} \rightarrow S_{\bar{x}} = \sqrt{\frac{2,2}{5}} = 0,663$$

**Amplitud estudentizada para medias comparadas:**

	Valor tabular q	$S_{\bar{x}}$	CNK
$q_{0,05 (4; 75)}$	3,73	0,663	2,47
$q_{0,05 (3; 75)}$	3,39	0,663	2,25
$q_{0,05 (2; 75)}$	2,82	0,663	1,87

**Hipótesis:**

H<sub>0</sub>: No hay diferencia entre las medias comparadas

H<sub>1</sub>: Si hay diferencia estadística significativa entre las medias comparadas

**Regla de decisión:**

Si la “Diferencia de medias comparadas” < CNK: Aceptar H<sub>0</sub>

Si la “Diferencia de medias comparadas” ≥ CNK: Aceptar H<sub>1</sub>

Si a la media F1 = 20,2 se resta CNK = 2,47 y se obtiene el valor 17,73 que es el indicativo de que las medias de 17,73 a 20,2 son estadísticamente iguales y comparten el rango con la letra A.

Cuando a media del genotipo F2 = 18,2 se resta el valor CNK = 2,25 da el valor teórico 15,95 indicativo de la amplitud del rango B, que abarca a los progenitores 1 y 2, además de F2 (Cuadro 166).

Cuadro 166. Separación de medias de cuatro genotipos de sorgo usando la prueba Newman-Keuls

Genotipos de sorgo	Producción de forraje Medias →	Diferencias entre tratamientos (t.ha <sup>-1</sup> )				CNK	Diferencia t - CNK	Rangos
		F1	F2	Progenitor 2	Progenitor 1			
		20,2	18,2	17,2	16,0			
F1	20,2	0,0	2,0	<b>3,0</b>	<b>4,2</b>	2,47	17,73	A
F2	18,2		0,0	1,0	<b>2,2</b>	2,25	15,95	AB
Progenitor 2	17,2			0,0	1,2	1,87		B
Progenitor 1	16,0				0,0			B

Nota: Letras iguales significa que no hay diferencia estadística con la prueba de rangos NK

**Interpretación:**

Las medias de los genotipos F1 y F2 resultaron estadísticamente iguales y con mayor rendimiento de forraje que los otros genotipos (p < 0,05).

**Decisión:**

Los genotipos F1 y F2 resultaron estadísticamente iguales, con mayor rendimiento de forraje que los otros.

### 6.5.8 COMPARACIONES ORTOGONALES

Las comparaciones ortogonales es una técnica de análisis estadístico que permite comparar de manera simultánea grupos de tratamientos o pares de tratamientos de interés preferencial. Es la continuación de un ADEVA diseñado para extraer información específica sobre las comparaciones de interés.

Cabe destacar que las comparaciones ortogonales dependen del diseño del experimento y de la elección de los contrastes lineales. Por lo tanto, es fundamental planificar cuidadosamente el diseño del estudio y definir los contrastes de manera adecuada para obtener resultados válidos e interpretables<sup>36</sup>.

Para el análisis de comparaciones ortogonales se requiere como datos; las sumatorias de los tratamientos (no las medias), el cuadrado medio del error (CME), los grados de libertad del error (GLE), el número de tratamientos (t), el número de repeticiones (r) y el nivel de significación ( $\alpha$ ).

En este análisis, para obtener los valores críticos de F, al nivel de significación concreto (como  $\alpha = 0,01$  o  $0,05$ ), se considera un grado de libertad para el numerador y los grados de libertad del error para el denominador (tomado del ADEVA). Hay que tener presente que dos comparaciones son ortogonales cuando la suma de los productos de sus coeficientes es igual a cero.

Las fórmulas usadas para el análisis de comparaciones ortogonales son:

$C_i$ : Comparación de dos tratamientos o grupos de T con r repeticiones.

Sumatoria de los coeficientes al cuadrado:

$$k = \sum C_i^2 \quad [121]$$

Sumatoria producto “tratamientos x coeficientes”

$$Q = \sum (T_i C_i) \quad [122]$$

Suma de cuadrados con 1 grado de libertad = Cuadrado medio.

$$CM = \frac{Q^2}{r \cdot k} \quad [123]$$

$$\text{Fórmula para obtener } F_{\text{calculada}} = \frac{CM}{CME}$$

Los valores críticos se obtienen de la tabla F con la probabilidad  $\alpha$ , GL del numerador = 1 y GLE para el denominador (tomado del ADEVA).

En Excel, los valores críticos de F se obtienen con la siguiente función:

=DISTR.F.INV(probabilidad;gradosdelibertad numerador;GLE)

En las hipótesis estadísticas se explicita la comparación de las medias:

$H_0: \mu_1 = \mu_2$ . Las medias son estadísticamente iguales

$H_1: \mu_1 \neq \mu_2$ . Las medias son estadísticamente diferentes

<sup>36</sup> En el programa INFOSTAT, los contrastes entre pares de tratamientos o pares de grupos de interés se realiza definiendo con antelación los coeficientes comparativos



La regla de decisión es la siguiente:

Si  $F_{\text{calculada}} < F_{\text{crítico}}$ : Aceptar  $H_0$

Si  $F_{\text{calculada}} \geq F_{\text{crítico}}$ : Aceptar  $H_1$

Las decisiones estadísticas, muchos investigadores las toman en base al valor  $p$ . Cuando  $p > 0,05$  se acepta  $H_0$ , si  $p < 0,05$  se acepta  $H_1$  con el 95 % de confianza y si  $p < 0,01$  se acepta  $H_1$  con el 99 % de confianza.

### Ejercicio 143:

Realizar las comparaciones ortogonales del ensayo “Efecto de las frecuencias del riego sobre el peso de los racimos de plátano” e identificar la mejor alternativa.

### Datos y análisis:

Los resultados del experimento de tres frecuencias de riego y las medias de peso de los racimos, se indican en el Cuadro 167. Los datos del ensayo son:  $t = 3$ ,  $r = 20$ ,  $GLE = 57$  y  $CME = 4,72$ . Se añade el valor escalar al máximo para interpretar los resultados.

Cuadro 167. Pesos promedios de racimos de plátano en función de las frecuencias de riego

Tratamientos	Frecuencias de riego	Sumatoria	R	Media de peso (kg.racimo <sup>-1</sup> )	Escalar al máximo (EAM)
t1	Cada 1 día	772	20	38,60	100
t2	Cada 2 días	749	20	37,45	97
t3	Cada 3 días	534	20	26,70	69

### Hipótesis estadísticas:

#### Comparación 1:

$H_0$ : La frecuencia de riego diaria = frecuencia de riego cada dos días

$H_1$ : La frecuencia de riego diaria  $\neq$  frecuencia de riego cada dos días

#### Comparación 2:

$H_0$ : Las frecuencias de riego diaria y cada dos días = frecuencia de riego cada tres días

$H_1$ : Las frecuencias de riego diaria y cada dos días  $\neq$  frecuencia de riego cada tres días

### Regla de decisión:

Si  $F_{\text{calculada}} < F_{\text{crítica}}$ : Aceptar  $H_0$

Si  $F_{\text{calculada}} \geq F_{\text{crítica}}$ : Aceptar  $H_1$

### Análisis comparativo:

Comparación de los tratamientos: T1 vs T2

$$K = \sum(C_1)^2 \rightarrow K = (1)^2 + (-1)^2 = 2$$

Comparación de los grupos: T1,T2 vs T3

$$K = \sum(C_2)^2 \rightarrow K = [(1)^2 + (1)^2] + (-2)^2 = 1+1+4 = 6$$

Suma producto de tratamientos x coeficientes:

$$Q_{c1} = \sum(T_1 \cdot C_1) \rightarrow Q_{c1} = [(772)(1)] + (749)(-1) = 23$$

$$Q_{c2} = \sum(T_1; T_2 \cdot C_2) \rightarrow Q_{c2} = [(772)(1) + (749)(1)] + (534)(-2) = 1521 - 1068 = 453$$

Suma de cuadrados:

$$CM = \frac{Q^2}{rk} \rightarrow CM_{C1} = \frac{23^2}{40} = \frac{529}{40} = 13$$

$$CM_{C2} = \frac{453^2}{120} = \frac{205209}{120} = 1710$$

$$F_{\text{calculada } C1} = \frac{CM_{C1}}{CME} = \frac{13}{4,72} = 2,80$$

$$F_{\text{calculada } C2} = \frac{CM_{C2}}{CME} = \frac{1710}{4,72} = 362$$

El análisis de contrastes o comparaciones ortogonales se expone en el Cuadro 168. En el Cuadro 169 se indica el ADEVA con la inclusión de las comparaciones ortogonales.

Cuadro 168. Contrastes ortogonales del peso del racimo de plátano en función a las frecuencias de riego

Comparación	Σ T1	Σ T2	Σ T3	R	k	r(k)	Q	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p
	772	749	534									
C1	1	-1	0	20	2	40	23	13	2,75 ns	4,01	7,10	0,0996
C2	1	1	-2	20	6	120	453	1710	362 **	4,01	7,10	0,000001

Nota: ns = no hay diferencia estadística, \*\* = hay diferencia altamente significativa (p < 0,01)

Cuadro 169. ADEVA del peso de racimo de plátano con inclusión de los contrastes de frecuencias de riego

Fuente de variación	GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p de F
Frecuencia de riego	2	2208	1103,98	233,7 **	3,16	5,00	0,000001
Riego diario vs riego cada dos días	(1)	13	13	2,8 ns	4,01	7,10	0,09963
Riego diario y cada 2 días vs riego cada 3 días	(1)	1710	1710	362 **	4,01	7,10	0,000001
Error experimental	57	269	4,72				
Total	59	2477	41,99				

Nota: ns = no hay diferencia estadística significativa, \*\* = hay diferencia estadística altamente significativa

### Interpretación:

Los pesos de los racimos de plátano con riegos diarios y cada dos días resultaron estadísticamente mayores que con riego cada tres días. El valor escalar al máximo (EAM) destaca que los riegos diarios (T1) dan los mayores pesos (100%), con los riegos cada dos días (T2) el peso reduce en 3 % y con los riegos cada tres días los pesos se reducen en 31 %.

### Decisión:

Los riegos diarios y cada dos días posibilitan obtener pesos de racimos más altos en comparación del peso con riegos cada tres días, donde se pierde el 31 % de peso.

### Ejercicio 144:

Determinar la mejor dieta en el ensayo “Efecto de cuatro dietas en gallinas sobre el peso de los huevos”, medidos en decagramos (Dg), usando el análisis de comparaciones ortogonales.

### Datos y análisis:

En ensayo tiene como características: t = 4, GLE = 6, CME = 5,1 y α = 0,05. Se planea comparar: testigo vs dieta C, testigo vs dietas B y C, dieta B vs dieta A (Cuadro 170).

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

Cuadro 170. Separación de medias mediante comparaciones ortogonales del peso de huevos (Dg) en función de cuatro dietas

Comparación	Tratamientos				r	K	r(k)	Q	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p
	ΣB	ΣA	Σ Testigo	ΣC									
	237	220	201	179									
C1	0	0	1	-1	4	2	8	22	61	11,86 *	5,99	13,75	0,014
C2	1	1	-2	0	4	6	24	55	126	25,0 **	5,99	13,75	0,003
C3	1	-1	0	0	4	2	8	17	36	7,08 *	5,99	13,75	0,037

Nota: \* = diferencia estadística significativa, \*\* = diferencia estadística altamente significativa

**Interpretación:**

Las dietas A y B > Testigo, la dieta B > dieta A y el testigo > dieta C.

**Decisión:**

La dieta B en gallinas ponedoras resultó superior a las dietas B, C y testigo (p < 0,05).

**Ejercicio 145:**

Realizar las comparaciones ortogonales para determinar el contenido de magnesio en las hojas de fréjol como efecto de la aplicación de tres alternativas de fertilización edáfica.

**Datos y análisis:**

Los resultados del ensayo de contenido de Mg en las hojas como efecto de la fertilización edáfica, se exponen en el Cuadro 171. Otra información importante para el análisis es la siguiente: T = 3, r = 8, GLE = 21, CME = 0,17 y α = 0,05.

Cuadro 171. Sumas, promedios y valores escalares al máximo del contenido de magnesio en las hojas de fréjol en función de tres tratamientos de fertilización edáfica

Fertilización edáfica	ΣY: Magnesio en las hojas de fréjol	r	Magnesio (ppm) $\bar{Y}$	Escalar Al Máximo (EAM)
T1	29,4	8	3,68	48
T2	35,7	8	4,46	58
T3	61,2	8	7,65	100

En el Cuadro 172, se indica el análisis de comparaciones ortogonales. La variable de transformada al valor EAM indica que con T3 se logra el mejor resultado (100 %), con T1 se logró el 48 % de Mg y aplicando T2 se alcanzó el 58 %.

Cuadro 172. Separación de medias mediante comparaciones ortogonales del contenido de Mg en hojas de fréjol en función de tres tratamientos de fertilización edáfica

Comparación	ΣT3	ΣT2	ΣT1	r	k	r(k)	Q	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p
	61,2	35,7	29,4									
C1	1	-1	0	8	2	16	25,5	41	239 **	4,32	8,02	0,00001
C2	0	1	-1	8	2	16	6,3	2	15 **	4,32	8,02	0,00100

Nota: \*\* = Diferencia estadística altamente significativa

**Interpretación:**

T3 > T2 > T1

**Decisión:**

El contenido de Mg en las hojas de fréjol, más alto fue T3 con el 99 % de confianza.

**Ejercicio 146:**

Realizar las comparaciones ortogonales de la producción en un ensayo factorial 3 x 3 +1, de tres variedades de maíz en tres niveles de fertilización más un testigo.

**Datos análisis:**

Los datos son: r = 5, t = 10, CME = 2,11, GLE = 36 y  $\alpha = 0,05$ . Los cálculos se indican en el Cuadro 173.

Cuadro 173. Comparaciones ortogonales para la separación de medias en un experimento AB + 1

	$\Sigma T1$	$\Sigma T2$	$\Sigma T3$	$\Sigma T4$	$\Sigma T5$	$\Sigma T6$	$\Sigma T7$	$\Sigma T8$	$\Sigma T9$	$\Sigma T10$										
Comp.	53	63	64	75	77	71	77	86	88	44	r	k	r(k)	Q	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p	
C1	1	1	1	1	1	1	1	1	1	-9	5	90	450	258	147,9	69,5**	4,11	7,40	0,0001	
C2	1	0	0	0	0	0	0	0	0	-1	5	2	10	9	8,1	3,80ns	4,11	7,40	0,059	
C3	0	0	0	0	0	0	0	0	1	-1	5	2	10	44	193,6	90,9**	4,11	7,40	0,0001	
C4	0	0	0	0	0	0	0	1	0	-1	5	2	10	42	176,4	82,8**	4,11	7,40	0,0001	
C5	0	0	0	0	0	0	1	0	0	-1	5	2	10	33	108,9	51,1**	4,11	7,40	0,0001	
C6	0	0	0	0	0	1	0	0	0	-1	5	2	10	27	72,9	34,2**	4,11	7,40	0,0001	
C7	0	0	0	0	0	0	0	1	-1	0	5	2	10	-2	0,4	0,19 ns	4,11	7,40	0,667	
C8	0	0	0	0	0	0	1	0	-1	0	5	2	10	-11	12,1	5,68*	4,11	7,40	0,022	
C9	0	0	0	0	0	1	0	0	-1	0	5	2	10	-17	28,9	13,6**	4,11	7,40	0,007	

Nota: ns = no hay diferencia estadística, \* = hay diferencia significativa y \*\* = hay diferencia altamente significativa

**Interpretación:**

T1, T2, T3, T4, T5, T6, T7, T8, T9 > T10 (testigo) \*\*. Testigo = T1 ns, T9 > T10 \*\*, T8 > T10 \*\*, T7 > T10 \*\*, T6 > T10 \*\*, T8 = T9 ns, T7 < T9 \* y T9 > T6 \*\*.

**Decisión:**

Los tratamientos T9 y T8 resultaron estadísticamente iguales y superiores a los demás con el 99 % de confianza.

**Ejercicio 147:**

Realizar el análisis de comparaciones ortogonales para el experimento AB + 1 + 1 referido a variedades (V) x fertilización (F) + testigo absoluto + testigo comercial, en maíz

**Datos y análisis:**

Las comparaciones ortogonales se exponen en el Cuadro 174.

Cuadro 174. Comparaciones ortogonales para la separación de medias en un experimento AB + 1 + 1

	$\Sigma T1$	$\Sigma T2$	$\Sigma T3$	$\Sigma T4$	$\Sigma T5$	$\Sigma T6$	$\Sigma T7$	$\Sigma T8$	$\Sigma T9$	$\Sigma T10$	$\Sigma T11$										
Comp.	53	63	64	75	77	71	77	86	88	44	64	r	k	r(k)	Q	CM	F <sub>calc.</sub>	F <sub>0,05</sub>	F <sub>0,01</sub>	p	
C1	0	0	0	0	0	0	0	0	0	1	-1	5	2	10	-20	40	18,78	4,11	7,40	0,001	
C2	0	0	0	0	0	0	0	1	-1	0	0	5	2	10	-2	0,4	0,19	4,11	7,40	0,67	
C3	0	0	0	0	0	0	-1	1	0	0	0	5	2	10	9	8,1	3,80	4,11	7,40	0,06	
C4	0	0	0	0	0	-1	1	0	0	0	0	5	2	10	6	3,6	1,69	4,11	7,40	0,20	
C5	0	0	0	-1	1	0	0	0	0	0	0	5	2	10	2	0,4	0,19	4,11	7,40	0,67	
C6	0	0	1	0	0	0	0	0	0	0	-1	5	2	10	0	0	0,00	4,11	7,40	1,00	
C7	0	-1	1	0	0	0	0	0	0	0	0	5	2	10	1	0,1	0,05	4,11	7,40	0,83	
C8	-1	1	0	0	0	0	0	0	0	0	0	5	2	10	10	10	4,69	4,11	7,40	0,04	
C9	-1	0	0	0	0	0	0	0	0	0	1	5	2	10	11	12	5,68	4,11	7,40	0,02	
C10	1	0	0	0	0	0	0	0	0	-1	0	5	2	10	9	8,1	3,80	4,11	7,40	0,06	

Nota: En este caso se usa el valor p para interpretar las comparaciones entre tratamientos.

**Interpretación:**

T10 < T11 (p < 0,01), T9 = T8 (p = 0,67), T8 = T7 (p = 0,06), T6 = T7 (p = 0,20), T5 = T4 (p = 0,67), T6, T3 = T11 (p = 1,0), T2 = T3 (p = 0,83), T1 < T2 (p = 0,04), T11 > T1 (p = 0,02) y T1 = T10 (p = 0,06).

**Decisión:**

Los mejores tratamientos fueron: T9 (V3F3) = T8 (V3F2) = T7 (V3F1). Estos corresponden a la variedad V3 con los tres niveles de fertilización (F1, F2 y F3).

## 6.6 ANÁLISIS DE VARIANZA NO PARAMÉTRICO

El análisis de varianza no paramétrico es utilizado para comparar medias de muestras independientes o relacionadas en ensayos aun cuando no se cumplan los supuestos de normalidad, homocedasticidad y adecuado número de grados de libertad del error, analizando datos cualitativos o cuantitativos.

Las pruebas de Kruskal-Wallis y de Friedman son equivalentes a los análisis de varianza en diseños completamente al azar (DCA) y en bloques al azar (BA), respectivamente. Al aceptar  $H_0$  se afirma que no hubo efecto de los tratamientos, a un nivel de significación concreto. Al aceptar  $H_1$ , que equivale a rechazar  $H_0$ , se afirma que al menos un tratamiento es estadísticamente diferente. En una segunda parte del análisis estadístico, hay que realizar la comparación de las medias de los tratamientos.

En el Gráfico 50, se observa los histogramas que conforman las series de datos de dos tratamientos: T1 y T2. ¿Las medias de estos tratamientos son estadísticamente iguales?. El arte de la estadística es tomar decisiones, gestionando los riesgos y errores.

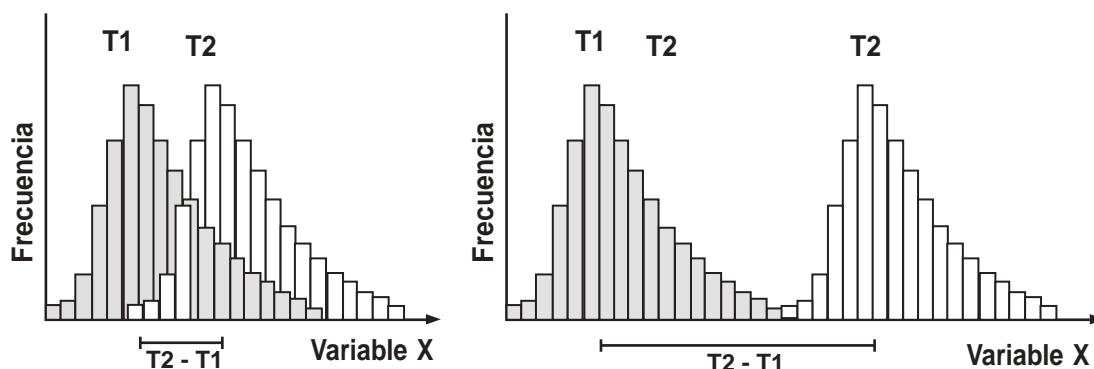


Gráfico 50. Distribuciones de las series de datos en dos tratamientos comparados

### 6.6.1 ANÁLISIS DE VARIANZA POR RANGOS DE KRUSKAL-WALLIS

El análisis de varianza por rangos de Kruskal y Wallis (KW) permiten decidir si  $k$  muestras independientes provienen de poblaciones diferentes (Badii et al., 2012, p. 154). Los datos pueden provenir de mediciones en cualquier escala, no exige que la distribución sea normal, ni que las varianzas sean homocedásticas o que el ensayo tenga un reducido número de grados de libertad del error. La prueba KW se inicia con la transformación a rangos, de la matriz de tratamientos, sin considerar los bloques. El valor más bajo tendrá el rango 1, el segundo valor tendrá el rango 2 y así sucesivamente. En caso de haber empates, se promedian las posiciones y continua el ordenamiento. El estadístico KW se identifica con la letra H, valor que se contrasta con  $\chi^2$  a un nivel de significación concreto (como  $\alpha = 0,05$ ) y grados de libertad,  $GL = t - 1$  (Martínez y Martínez, 2010; Lind et al., 2012, p. 699).

La fórmula para calcular el estadístico H (Siegel, 1990, p. 216), es la siguiente:

$$H = \frac{12}{n(n+1)} \sum \frac{R^2}{n_i} - 3(n+1) \quad [124]$$

Cuando hay ligas, el estadístico H, se corrige usando la fórmula:

$$H = \frac{\frac{12}{n(n+1)} \sum \frac{R^2}{n_i} - 3(n+1)}{1 - \frac{L^3 - L}{n^3 - n}} \quad [125]$$

Dónde:

H = Estadístico de Kruskal y Wallis

n = Número de observaciones total

$n_i$  = Número de observaciones por cada tratamiento

$\sum \frac{R^2}{n_i}$  = Promedio de los rangos de cada tratamiento

L = Ligas o datos repetidos en las series de datos.

Las hipótesis estadísticas se formulan como:

H<sub>0</sub>: Los tratamientos son estadísticamente iguales

H<sub>1</sub>: Al menos un tratamiento es estadísticamente diferente

La regla de decisión se plantea de la siguiente manera:

Si  $H < \chi^2_{0,05}$ : Aceptar H<sub>0</sub>

Si  $H > \chi^2_{0,05}$ : Aceptar H<sub>1</sub>

### Ejercicio 148:

Determinar si el peso (libras.animal<sup>-1</sup>) de los cerdos al nacer es afectado por el tamaño de la camada en base a datos de ocho camadas.

### Datos y análisis:

En el Cuadro 175, se exponen los datos de peso de las ocho camadas de cerdos al nacimiento (Y inicial), medido en libras.

Cuadro 175. Peso (libras.animal<sup>-1</sup>) de los cerdos (A<sub>i</sub>) al nacimiento en camadas de distinto tamaño

Camadas	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Medias (libras.cerdo <sup>-1</sup> )
Cam1	2,00	2,80	3,30	3,20	4,40	3,60	1,90	3,30	2,80	1,10	2,84
Cam2	3,50	2,80	3,20	3,50	2,30	2,40	2,00	1,60			2,66
Cam3	3,30	3,60	2,60	3,10	3,20	3,30	2,90	3,40	3,20	3,20	3,18
Cam4	3,20	3,30	3,20	2,90	3,30	2,50	2,60	2,80			2,98
Cam5	2,60	2,60	2,90	2,00	2,00	2,10					2,37
Cam6	3,10	2,90	3,10	2,50							2,90
Cam7	2,60	2,20	2,20	2,50	1,20	1,20					1,98
Cam8	2,50	2,40	3,00	1,50							2,35

Nota: Cam = camada, A = número de animal por camada

Fuente: Siegel (1990, p. 221)

### Hipótesis estadísticas:

H<sub>0</sub>: Los tratamientos son estadísticamente iguales

H<sub>1</sub>: Al menos un tratamiento es estadísticamente diferente

### Regla de decisión:

$$\left| \frac{R_1}{n_1} - \frac{R_2}{n_2} \right| < Z_{\alpha} \sqrt{\frac{n(n+1)}{12} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}: \text{Aceptar } H_0$$

$$\left| \frac{R_1}{n_1} - \frac{R_2}{n_2} \right| \geq Z_{\alpha} \sqrt{\frac{n(n+1)}{12} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}: \text{Aceptar } H_1$$

**Cálculos de las sumas y promedios usando los rangos:**

Los datos transformados a rangos se indica en el Cuadro 176.

Cuadro 176. Rangos (R) de peso de los cerdos al nacimiento en ocho camadas de distinto tamaño

Camadas	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	R	n <sub>i</sub>	$\frac{R^2}{n_i}$	
Cam1	8	27,5	47,5	41	56	54,5	6	47,5	27,5	1	316,5	10	10.017	
Cam2	52,5	27,5	41	52,5	14	15	8	5			215,5	8	5805	
Cam3	47,5	54,5	23	36	41	47,5	31,5	51	41	41	414	10	17.140	
Cam4	41	47,5	41	31,5	47,5	8,5	23	27,5			277,5	8	9626	
Cam5	23	23	31,5	8	10	11					106,5	6	1890	
Cam6	36	31,5	36	18,5							122	4	3721	
Cam7	23	12,5	12,5	18,5	2,5	2,5					71,5	6	852	
Cam8	18,5	16	34	4							72,5	4	1314	
												$\frac{\sum R^2}{n_i}$		50.365

Nota: n = 56, cam1 = 10, cam2 = 8, cam3 = 10, cam4 = 8, cam5 = 6, cam6 = 4, cam7 = 6, cam8 = 4. Cada camada es equivalente a n<sub>i</sub>. Se constató 13 ligas

**Cálculo del estadístico H:**

Sin corrección por empates o ligas:

$$H = \frac{12}{n(n+1)} \sum \frac{R^2}{n_i} - 3(n+1) \quad \rightarrow H = \frac{12}{56(56+1)} (50365) - 3(56+1) = 18,34$$

Con corrección por empates o ligas:

$$H = \frac{\frac{12}{n(n+1)} \left[ \sum \frac{R^2}{n_i} \right]^2 - 3(n+1)}{1 - \frac{L^3 - L}{n^3 - n}} \quad \rightarrow H = \frac{\frac{12}{56(57)} (6363) - 3(57)}{1 - \frac{13^3 - 13}{56^3 - 56}} = 18,57$$

Los estadísticos H calculados “sin la corrección por ligas” y “con la corrección por ligas” difieren en 23 centésimas. Se recomienda usar solo la fórmula de H sin corrección.

**Uso de herramientas de INFOSTAT:**

Con el programa INFOSTAT, el estadístico H = 18,46 corresponde a la media de los valores H con corrección y sin corrección por ligas.

**Obtención del valor crítico  $\chi^2$ :**

Grados de libertad (GL) = t - 1 = 7.  $\alpha = 0,05 \sim 5 \%$ .  $\chi^2_{(0,05;GL=7)} = 14,07$ .

El estadístico H se contrasta con  $\chi^2$  obtenido en tabla o usando la función de Excel: =PRUEBA.CHI.INV(probabilidad; grados de libertad) .

El valor p de Chi-cuadrado = 0,0096 (con la función de Excel)

En el contraste de H con  $\chi^2$  en Excel, usar la función: =DISTR.CHI(x; grados de libertad), que operativamente equivale a =DISTR.CHI(valor H; 1 grado de libertad)

**Interpretación:**

H = 18,34 >  $\chi^2_{0,05} = 14,07$ : Aceptar H<sub>1</sub>.

**Decisión:**

Al menos un tratamiento es estadísticamente diferente con el 95 % de confianza (p < 0,05). Se tiene que continuar el análisis para comparar las medias.

**Ejercicio 149:**

Realizar el análisis de varianza por rango de Kruskal-Wallis, en un estudio de cuatro dietas artificiales para la alimentación de broca del fruto del café, a nivel de laboratorio: la variable experimental es el número de larvas vivas por recipiente.

**Datos y análisis:**

La data corresponde a larvas de broca del café vivas, observadas a nivel de laboratorio, en función de cuatro dietas artificiales. La transformación de datos originales a rangos, toma en consideración la matriz de todos los tratamientos, que se ordenan de menor a mayor. En  $n = 20$  observaciones, los rangos varían de 1 a 20 y en el caso de haber empates, entre dos o más datos, se calcula el promedio de las posiciones.

El número de larvas más bajo es 33 que se repite en II-D2 y en V-D3, promediándose las posiciones:  $\frac{1+2}{2} = 1,5$ . El siguiente valor es 34 larvas en I-D3. II-D3 y IV-D3, por lo tanto, se promedian las posiciones:  $\frac{3+4+5}{3} = 4$ . El proceso continúa hasta cubrir los  $n = 20$  datos de la matriz (Cuadro 177).

Cuadro 177. Número de larvas vivas de broca del café por recipiente en función de cuatro dietas artificiales y transformación a rangos

Dietas	Datos originales (larvas vivas/recipiente)									Media
	I	II	III	IV	V	VI	VII	VIII	IX	
D1	45	46	49	44						46
D2	35	33								34
D3	34	34	35	34	33					34
D4	41	41	44	43	41	42	44	41	41	42
Dietas	Datos transformados a rangos (R)									ΣR
	I	II	III	IV	V	VI	VII	VIII	IX	
D1	18	19	20	16						73
D2	6,5	1,5								8
D3	4	4	6,5	4	1,5					20
D4	10	10	16	14	10	13	16	10	10	109

En Excel, los rangos en las series de datos se obtienen con la función:

=JERARQUÍA.MEDIA(CELDA DEL NÚMERO\_MATRIZ DE DATOS\_ORDEN).

**Resumen de rangos:**

En el Cuadro 178 se expone el resumen de los datos requeridos para el cálculo del estadístico H de la prueba Kruskal-Wallis.

Cuadro 178. Procedimiento de cálculo de los rangos para el análisis de varianza KW

Dietas artificiales para broca del café	R.	$n_i$	$\bar{R}$	$R^2$	$\frac{R^2}{n_i}$
D1	73	4	18	5329	1332,3
D2	8	2	4	64	32,0
D3	20	5	4	400	80,0
D4	109	9	12	11881	1320,1
Σ		20			2764,4



**Análisis de H:**

$$H = \frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1) \quad H = \frac{12}{20(20+1)} (2764,4) - 3(20+1) = 15,98$$

**Valor crítico  $\chi^2$ :**

GL = t - 1 = 4 - 1 = 3 con  $\alpha = 0,05 \sim 5 \%$

$$\chi^2_{(0,05; GL=3)} = 7,815$$

El valor crítico de  $\chi^2$  se obtiene en la tabla específica o con la función de Excel:

=PRUEBA.CHI.INV(probabilidad; grados de libertad).

El valor  $p$  se puede obtener en Excel con la función:

=DISTR.CHI(x; grados de libertad).

El valor  $x$  corresponde al estadístico H de la prueba Kruskal-Wallis.

$$p = 0,0011$$

**Interpretación:**

$H = 15,98 > \chi^2_{0,05} = 7,815$ : Aceptar  $H_1$ .

**Decisión:**

Al menos un tratamiento es estadísticamente diferente ( $p < 0,05$ ).

COMPARACIÓN *POST HOC* DE LA PRUEBA KRUSKAL-WALLIS

El análisis de varianza por rangos de Kruskal-Wallis es una prueba no paramétrica que da como resultado una decisión de aceptar o rechazar una hipótesis de nulidad (Martínez et al., 2020). El comparador estadístico de Kruskal-Wallis se abrevia como CKW.

La fórmula de cálculo es la siguiente:

$$CKW = Z_\alpha \sqrt{\frac{n(n+1)}{12} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \quad [126]$$

Dónde:

CKW = Comparador estadístico de Kruskal-Wallis

$Z_\alpha$  = Valor asociado a la probabilidad de equivocarse. Si  $\alpha = 0,05 \rightarrow Z = 1,96$

n = Suma de las observaciones de los dos tratamientos comparados

$n_1$  y  $n_2$  = Las observaciones para los dos tratamientos a compararse

Las hipótesis estadísticas se formulan como:

$H_0$ : No hay diferencia estadística significativa entre dos tratamientos comparados

$H_1$ : Si hay diferencia estadística significativa entre dos tratamientos comparados

La regla de decisión se plantea como:

$$\left| \frac{R_1}{n_1} - \frac{R_2}{n_2} \right| < CKW: \text{Aceptar } H_0$$

$$\left| \frac{R_1}{n_1} - \frac{R_2}{n_2} \right| > CKW: \text{Aceptar } H_1$$

$$\frac{R_i}{n_i} = \bar{R} \rightarrow \text{Promedio del rango por tratamiento}$$

**Ejercicio 150:**

Realizar la comparación de medias de peso al nacimiento de ocho camadas de cerdos, donde el análisis de varianza KW indicó que había diferencias estadísticamente significativas,

**Datos y análisis:**

Los datos corresponden a los pesos al nacimiento de 8 camadas de distinto tamaño.

$$n = n_1 + n_2$$

$n_1, n_2$ : números de observaciones por los tratamientos sujetos a comparación.

$$\frac{R_i}{n_i} = \bar{R} \text{ Promedio del rango por tratamiento}$$

**Hipótesis estadísticas para comparar tratamientos:**

$H_0$ : No hay diferencia estadística significativa entre los dos tratamientos comparados

$H_1$ : Si hay diferencia estadística significativa entre los dos tratamientos comparados

**Regla de decisión:**

$$\left| \frac{R_1}{n_1} - \frac{R_2}{n_2} \right| < Z_\alpha \sqrt{\frac{n(n+1)}{12} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}: \text{ Aceptar } H_0 \rightarrow (\text{ns})$$

$$\left| \frac{R_1}{n_1} - \frac{R_2}{n_2} \right| > Z_\alpha \sqrt{\frac{n(n+1)}{12} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}: \text{ Aceptar } H_1 \rightarrow (*)$$

**Análisis:**

$$|\bar{R}_3 - \bar{R}_4| = 41,4 - 34,7 = 6,71^* \quad 1,96 \sqrt{\frac{18(19)}{12} \left( \frac{1}{10} + \frac{1}{8} \right)} = 4,96$$

$$|\bar{R}_4 - \bar{R}_1| = 34,7 - 31,7 = 3,04\text{ns} \quad 1,96 \sqrt{\frac{18(19)}{12} \left( \frac{1}{8} + \frac{1}{10} \right)} = 4,96$$

$$|\bar{R}_1 - \bar{R}_6| = 31,7 - 30,5 = 1,15\text{ns} \quad 1,96 \sqrt{\frac{14(15)}{12} \left( \frac{1}{10} + \frac{1}{4} \right)} = 4,85$$

$$|\bar{R}_6 - \bar{R}_2| = 30,5 - 26,9 = 3,56\text{ns} \quad 1,96 \sqrt{\frac{12(13)}{12} \left( \frac{1}{4} + \frac{1}{8} \right)} = 4,33$$

$$|\bar{R}_2 - \bar{R}_8| = 26,9 - 18,1 = 8,81^* \quad 1,96 \sqrt{\frac{12(13)}{12} \left( \frac{1}{8} + \frac{1}{4} \right)} = 4,33$$

$$|\bar{R}_8 - \bar{R}_5| = 18,1 - 17,8 = 0,38\text{ns} \quad 1,96 \sqrt{\frac{10(11)}{12} \left( \frac{1}{4} + \frac{1}{6} \right)} = 3,83$$

$$|\bar{R}_5 - \bar{R}_7| = 17,8 - 11,9 = 5,83^* \quad 1,96 \sqrt{\frac{12(13)}{12} \left( \frac{1}{6} + \frac{1}{6} \right)} = 4,08$$

En el Cuadro 179, se indica el procedimiento de los cálculos del comparador estadístico KW

Cuadro 179. Cálculo del comparador estadístico de Kruskal-Wallis

Camada	$\bar{R}$	$n_i$	Comparación	Diferencia De $\bar{R}$	n	$\frac{n(n+1)}{12}$	$\frac{1}{n_1}$	$\frac{1}{n_2}$	$Z_{0,05}$	CKW
Cam3	41,4	10								
Cam4	34,7	8	Cam3 vs Cam4	6,71*	18	28,5	0,10	0,13	1,96	4,96
Cam1	31,7	10	Cam4 vs Cam1	3,04 ns	18	28,5	0,13	0,10	1,96	4,96
Cam6	30,5	4	Cam1 vs Cam6	1,15 ns	14	17,5	0,10	0,25	1,96	4,85
Cam2	26,9	8	Cam6 vs Cam2	3,56 ns	12	13,0	0,25	0,13	1,96	4,33
Cam8	18,1	4	Cam2 vs Cam8	8,81*	12	13,0	0,13	0,25	1,96	4,33
Cam5	17,8	6	Cam8 vs Cam5	0,38 ns	10	9,2	0,25	0,17	1,96	3,83
Cam7	11,9	6	Cam5 vs Cam7	5,83*	12	13,0	0,17	0,17	1,96	4,08

Nota: ns = no hay diferencia significativa, \* = hay diferencia estadística significativa

**Interpretación:**

Cam3 (10 animales) > Cam4 (8 animales) = Cam1 (10 animales) = Cam6 (4 animales) = Cam2 (8 animales) > Cam8 (4 animales) = Cam5 (6 animales) > Cam7 (6 animales). Los resultados se pueden interpretar apoyado en un modelo gráfico.

En la camada 1 hubo animales con pesos extremos (más bajo y más alto) que generó amplia dispersión. La camada 3 mostró menor dispersión y mediana más altas (Gráfico 51).

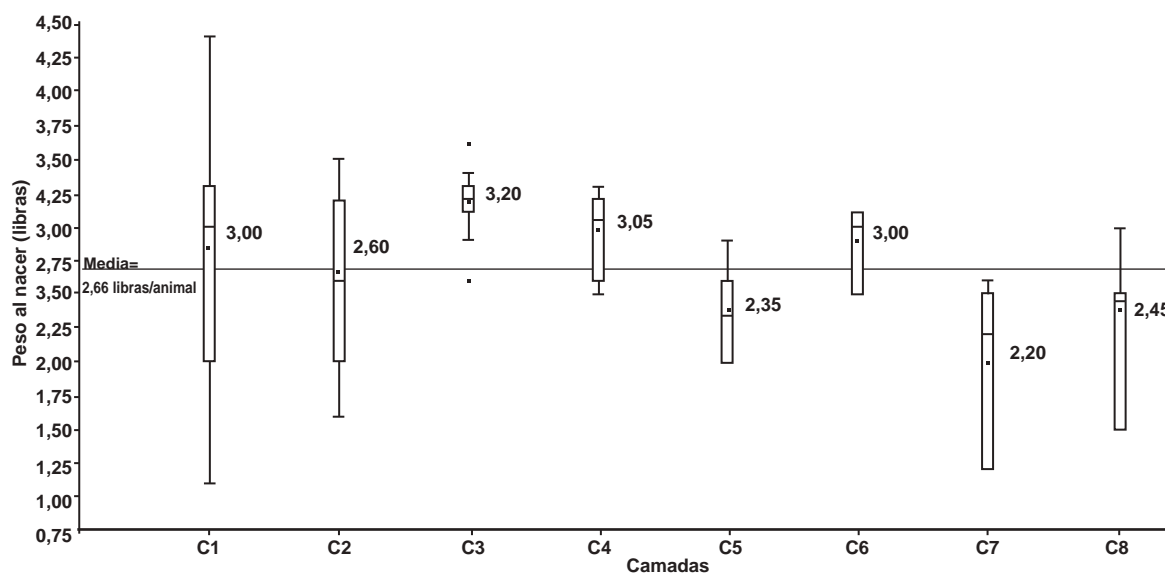


Gráfico 51. Dispersión de los datos y posición de las medianas de la variable peso al nacimiento de cerdos en camadas de distinto tamaño

**Decisión:**

Las camadas con 10 y 8 animales tienen pesos estadísticamente mayores que las camadas de 4 y 6 animales ( $p < 0,05$ ). La camada 3 es mejor.

**Ejercicio 151:**

Realizar la prueba *post hoc* CKW para separar las medias de las cuatro dietas artificiales en la alimentación de broca del fruto del café, siendo la variable de respuesta las larvas vivas por recipiente, continuando con el análisis KW.

Continuando con el análisis KW, se dispone de la información siguiente:

$n = n_1 + n_2$  corresponde a la suma de las observaciones de los tratamientos comparados.

$n_1 = 4, n_2 = 2, n_3 = 5, n_4 = 9$  son las observaciones por cada uno de los tratamientos.

**Hipótesis estadísticas para comparar tratamientos:**

H<sub>0</sub>: No hay diferencia estadística significativa entre los dos tratamientos comparados

H<sub>1</sub>: Si hay diferencia estadística significativa entre los dos tratamientos comparados

**Regla de decisión:**

$$\left| \frac{R_1}{n_1} - \frac{R_2}{n_2} \right| < C_{KW}: \text{Aceptar } H_0$$

$$\left| \frac{R_1}{n_1} - \frac{R_2}{n_2} \right| > C_{KW}: \text{Aceptar } H_1$$

**Análisis del comparador estadístico CKW:**

La separación de rangos de los tratamientos, referidos a las dietas artificiales de broca del café, se expone en el Cuadro 180.

$$\left| \frac{R_1}{n_1} - \frac{R_4}{n_4} \right| = 18 - 12 = 6^* > 1,96 \sqrt{\frac{13(13+1)}{12} \left( \frac{1}{4} + \frac{1}{9} \right)} = 4,59$$

$$\left| \frac{R_4}{n_4} - \frac{R_3}{n_3} \right| = 12 - 4 = 8^* > 1,96 \sqrt{\frac{14(14+1)}{12} \left( \frac{1}{9} + \frac{1}{5} \right)} = 4,57$$

$$\left| \frac{R_3}{n_3} - \frac{R_2}{n_2} \right| = 4 - 4 = 0 \text{ ns} < 1,96 \sqrt{\frac{7(7+1)}{12} \left( \frac{1}{5} + \frac{1}{2} \right)} = 3,54$$

Cuadro 180. Separación de medias con el estadístico CKW en ensayo de dietas artificiales para broca del café

Dietas	R	n <sub>i</sub>	$\bar{R}$	Comparación	N	$\frac{n(n+1)}{12}$	$\frac{1}{n_1} + \frac{1}{n_2}$	Z <sub>0,05</sub>	CKW	Diferencias entre $\bar{R}$ comparadas
D1	73	4	18	--	--	--	--	--	--	--
D4	109	9	12	D1 > D4	13	15,2	0,361	1,96	4,59	6 *
D3	20	5	4	D4 > D3	14	17,5	0,311	1,96	4,57	8 *
D2	8	2	4	D3 = D2	7	4,7	0,700	1,96	3,54	0 ns

Nota: ns = no hay diferencia significativa, \* Hay diferencia significativa

**Interpretación:**

D1 > D4 > D3 = D2.

**Decisión:**

La dieta D1 es estadísticamente mejor que las otras dietas artificiales usadas para la cría de broca del café en el laboratorio (p<0,05).

6.6.2 ANÁLISIS DE VARIANZA POR RANGOS DE FRIEDMAN

El análisis de varianza por rangos de Friedman es una técnica no paramétrica usada cuando los datos del ensayo en bloques al azar no cumplen con los supuestos de normalidad y homocedasticidad requeridos para el ADEVA paramétrico (Martínez et al., 2020).

La asignación de rangos se refiere en la ordenación por posiciones de las series de datos de cada tratamiento, siempre de menor a mayor. Al valor más bajo le corresponde el rango 1, el siguiente valor, en forma ascendente, el rango 2, y así sucesivamente. Cuando ocurren empates, se promedian las posiciones y se continua el ordenamiento.

El estadístico de Friedman  $\chi_r^2$  se contrasta con Chi cuadrado (Siegel, 1990, p. 197; Martínez y Martínez, 2010) en un nivel de significación prefijado como  $\alpha = 0,05$  y grados de libertad,  $GL = r - 1$ .

La fórmula para calcular el estadístico de Friedman es la siguiente:

$$\chi_r^2 = \frac{12}{r \cdot t(t+1)} \sum R_i^2 - 3r(t+1) \quad [127]$$

Dónde:

$\chi_r^2$  = Estadístico de Friedman

r = Número de repeticiones

t = Número de tratamientos

$\sum R_i^2$  = Sumatoria de los rangos de los tratamientos al cuadrado, por cada bloque.

La hipótesis estadística en la prueba de Friedman, se plantea como:

$H_0: \bar{X}_A - \bar{X}_B = 0$ . No hay diferencia estadística entre tratamientos

$H_1: \bar{X}_A - \bar{X}_B \neq 0$ . Si hay diferencia estadística significativa entre tratamientos

La regla de decisión es la siguiente:

Si  $\chi_r^2 < \chi_{0,05}^2$ : Aceptar  $H_0$ . No hay diferencia estadística

Si  $\chi_r^2 > \chi_{0,05}^2$ : Aceptar  $H_1$ . Si hay diferencia estadística significativa.

Luego de realizar el análisis de varianza por rangos de Friedman y aceptar la hipótesis alternativa ( $H_1$ : al menos un tratamiento es estadísticamente diferente), se tiene que realizar una prueba *post hoc* para comparar los tratamientos y determinar cuáles son estadísticamente diferentes o estadísticamente iguales.

El concepto de la separación de las medias de los tratamientos es similar al indicado en el análisis paramétrico. Se trata de identificar el mejor tratamiento. Cabe destacar que la decisión se relaciona con el tipo de variable: “mayor es mejor” o “menor es mejor”.

### Ejercicio 152:

Realizar el análisis de varianza por rangos de Friedman con su prueba *post hoc*, con los datos del conteo de insectos vivos antes de la aplicación (AA) y 14 días después de la aplicación de un bioplaguicida (14DDA) en un ensayo con diseño en Bloques al Azar.

#### Datos y análisis:

En este ensayo, la variable de respuesta es el “número de insectos vivos” que corresponde al tipo “menor es mejor”. Los datos del ensayo se exponen en el Cuadro 181.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 181. Datos del conteo de insectos vivos en plantas, antes (AA) y después de la aplicación (14DDA) de un bioplaguicida*

Repetición	Tratamientos	Dosis de bioinsecticida	0 AA	14 DDA	Diferencia (AA – 14 DDA)
1	T1	Dosis baja	71	45	26
1	T2	Dosis media	80	37	43
1	T3	Dosis alta	77	36	41
1	T4	Producto de referencia	79	53	26
1	T5	Testigo absoluto	77	181	-104
2	T1	Dosis baja	64	32	32
2	T2	Dosis media	81	32	49
2	T3	Dosis alta	74	25	49
2	T4	Producto de referencia	69	62	7
2	T5	Testigo absoluto	77	175	-98
3	T1	Dosis baja	73	62	11
3	T2	Dosis media	73	47	26
3	T3	Dosis alta	73	32	41
3	T4	Producto de referencia	65	69	-4
3	T5	Testigo absoluto	78	187	-109
4	T1	Dosis baja	79	52	27
4	T2	Dosis media	78	48	30
4	T3	Dosis alta	80	35	45
4	T4	Producto de referencia	70	92	-22
4	T5	Testigo absoluto	80	149	-69

Los datos de las diferencias AA- 14DDA, en valores absolutos, se reorganizan en una tabla de doble entrada, en base a los bloques. Estos datos se transforman a rangos por cada bloque, calculándose luego la suma de rangos ( $\Sigma R$ ) que se eleva al cuadrado ( $R^2$ ). La data para el análisis de la varianza de Friedman, se exponer en el Cuadro 182.

*Cuadro 182. Diferencias del número de insectos vivos, antes (AA) y 14 después de la aplicación de un bioplaguicida y transformaciones a rangos por bloques*

Tratamientos	Bloques				Diferencias medias
	I	II	III	IV	
T1	26	32	11	27	24
T2	43	49	26	30	37
T3	41	49	41	45	44
T4	26	7	4	22	15
T5	104	98	109	69	95

Tratamientos	Bloques				$\Sigma R$	$R^2$
	I	II	III	IV		
T1	1,5	2	2	2	7,5	56
T2	4	3,5	3	3	13,5	182
T3	3	3,5	4	4	14,5	210
T4	1,5	1	1	1	4,5	20
T5	5	5	5	5	20,0	400
					60	869

**Hipótesis estadísticas:**

H<sub>0</sub>: T1 = T2 = T3 = T4 = T5 No hay diferencia estadística significativa entre tratamientos

H<sub>1</sub>: Al menos un tratamiento es estadísticamente diferente

**Regla de decisión:**

Si  $\chi_r^2 < \chi_{0,05}^2$ : Aceptar H<sub>0</sub>

Si  $\chi_r^2 > \chi_{0,05}^2$ : Aceptar H<sub>1</sub>

**Cálculo del estadístico de Friedman:**

$$\chi_r^2 = \frac{12}{rt(t+1)} \sum R_i^2 - 3r(t+1) \rightarrow X_r^2 = \frac{12}{4(5)(5+1)} (869) - 3(4)(5+1) = 14,90$$

**Valores críticos:**

$\chi_{0,05}^2$  valor crítico con  $\alpha = 0,05$  y Grados de libertad (GL) = (r - 1) = 3

$\chi_{0,05}^2 = 7,81$  y  $\chi_{0,01}^2 = 11,34$

p de  $\chi^2 = 0,00089 \sim 0,089 \%$

En Excel se obtiene los valores críticos de  $\chi^2$  con la función:

=PRUEBA.CHI.INV(probabilidad; grados de libertad).

El valor p de Chi cuadrado se obtiene con la función:

= DISTR.CHI(X; grados de libertad).

La confianza en la decisión se calcula como: 1 - p.

1 - 0,00089 = 0,9991  $\sim$  99,91 % de confianza en la decisión.

**Interpretación:**

$\chi_r^2 = 14,9 > \chi_{0,05}^2 = 11,34$ : Aceptar H<sub>1</sub> con una confianza de 99 %

p = 0,00089

**Decisión:**

Al menos un tratamiento es estadísticamente diferente (p < 0,01).

**COMPARACIÓN POST-HOC DE LA PRUEBA DE FRIEDMAN**

Cuando se ha aceptado H<sub>1</sub> con la prueba de Friedman, se requiere separar las medias, de los tratamientos, usando una prueba post-hoc. Si hubiera dos tratamientos, por simple inspección se identifica la de mayor valor. Si hay más de dos medias se recomienda ordenar las medias o medianas de mayor a menor para agilizar el análisis.

La comparación entre pares de medias en esta prueba *post hoc* se puede realizar con la prueba de Mann-Whitney (Martínez et al., 2020), que equivale a la prueba t o mediante el uso del “comparador estadístico de Friedman” (CEF).

$$CEF = Z_{(1-\frac{\alpha}{2})} [\text{Error estándar}] \rightarrow CEF = Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{rt(t+1)}{6}}$$

El valor de Z se asocia al nivel de significación  $\alpha$ . En las pruebas de dos colas, el valor Z = 1,96 para  $\alpha = 0,05$  y Z = 2,58 para  $\alpha = 0,01$ .

El error estándar del estadístico de Friedman se calcula con la fórmula:

$$EE = \sqrt{\frac{rt(t+1)}{6}} \quad [128]$$

Las hipótesis estadísticas se formulan como:

H<sub>0</sub>: Los tratamientos son estadísticamente iguales

H<sub>1</sub>: Los tratamientos son estadísticamente diferentes

La regla de decisión se basa en:

$|R_1 - R_2| < CEF$ : Aceptar H<sub>0</sub>

$|R_1 - R_2| \geq CEF$ : Aceptar H<sub>1</sub>

### Ejercicio 153:

Realizar la prueba *post hoc* de Friedman (CEF) del ensayo sobre el efecto de 5 dosis del bioinsecticida donde se concluyó que al menos uno de los cinco tratamientos es estadísticamente diferente ( $p < 0,05$ ). La variable experimental fue el número de insectos vivos después de 14 días de la aplicación.

#### Análisis del comprador estadístico de Friedman:

$$CEF = Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{rt(t+1)}{6}} \rightarrow CEF = 1,96 \sqrt{\frac{4(5)(5+1)}{6}} \rightarrow CEF = 1,96 \sqrt{\frac{20(6)}{6}} = 1,96(4,47) = 8,8$$

#### Hipótesis estadísticas:

H<sub>0</sub>: Los dos tratamientos comparados son estadísticamente iguales

H<sub>1</sub>: Los dos tratamientos comparados son estadísticamente diferentes

#### Regla de decisión:

$|R_1 - R_2| < CEF$ : Aceptar H<sub>0</sub>

$|R_1 - R_2| \geq CEF$ : Aceptar H<sub>1</sub>

$$EE = \sqrt{\frac{rt(t+1)}{6}} \rightarrow EE = \sqrt{\frac{(4)(5)(6)}{6}} = \sqrt{\frac{120}{6}} = 4,47$$

#### Comparaciones de Friedman:

$|R_5 - R_4| = 15,5 > CEF = 8,8$ : T5 > T4

$|R_4 - R_3| = 10 > CEF = 8,8$ : T3 > T4

$|R_3 - R_1| = 14,5 > CEF = 8,8$ : T3 = T1

$|R_2 - R_1| = 6 > CEF = 8,8$ : T2 = T1

La comparación entre rangos, de los tratamientos, en la prueba de Friedman, se detalla en el Cuadro 183. En el cuadro 184, se exponen los resultados de la comparación de los rangos, de los tratamientos.

Cuadro 183. Comparación entre rangos de los tratamientos en la prueba de Friedman

Tratamientos	Comparación	R	Diferencias de R comparadas	Z <sub>0,05</sub>	Error estándar	CEF	Significación
T1	--	7,5		--	--	--	
T2	T2 vs T1	13,5	6	1,96	4,47	8,8	Aceptar H <sub>0</sub>
T3	T3 vs T1	14,5	7	1,96	4,47	8,8	Aceptar H <sub>0</sub>
T4	T3 vs T4	4,5	10	1,96	4,47	8,8	Aceptar H <sub>1</sub>
T5	T5 vs T4	20	15,5	1,96	4,47	8,8	Aceptar H <sub>1</sub>



**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 184. Resultados de la comparación entre tratamientos con la prueba de Friedman*

Tratamientos	Evaluadores					Rango	
	R	T5	T3	T2	T1		T4
		20	14,5	13,5	7,5	4,5	
T5	20	0					A
T3	14,5	5,5	0				A
T2	13,5	6,5	1,0	0			AB
T1	7,5	12,5	7,0	6,0	0		B
T4	4,5	15,5	10,0	9,0	3,0	0	B
	CEF	8,8					

Nota: Letras comunes en los rangos por tratamiento significa que son estadísticamente iguales

**Interpretación:**

El comparador estadístico de Friedman CEF = 8,8. Las diferencias entre rangos se contrastan con CEF. Para el tipo de variable “menor es mejor”, los mejores tratamientos son T4, T1 y T2.

**Decisión:**

Los mejores tratamientos fueron T4, T1 y T2. Los costos definirán la mejor opción.

**Ejercicio 154:**

En un estudio de la aceptabilidad de cuatro muestras de un arroz cocido, evaluados por cinco expertos, usando una escala ordinal de 1 a 10 (1 = muy baja y 10 = muy alta), se plantea determinar si hay diferencias estadísticas usando la prueba de Friedman.

**Datos y análisis:**

Los resultados de aceptabilidad dados por cinco evaluadores (r) para cuatro muestras de arroz, se indican en el Cuadro 185. En el Cuadro 186, se exponen los rangos.

*Cuadro 185. Calificaciones sensoriales asignadas a cuatro muestras de arroz por cinco evaluadores*

Muestras	Evaluadores					Mediana
	E1	E2	E3	E4	E5	
Muestra 1	9	6	9	9	8	9,0
Muestra 2	4	5	1	6	5	5,0
Muestra 3	1	2	2	3	5	2,0
Muestra 4	7	8	6	7	6	7,0

*Cuadro 186. Rangos asignados a cuatro muestras de arroz en evaluación sensorial*

Muestras	Evaluadores					Rangos	R <sup>2</sup>
	E1	E2	E3	E4	E5		
Muestra 1	4	3	4	4	4	19,0	361,0
Muestra 2	2	2	1	2	1,5	8,5	72,3
Muestra 3	1	1	2	1	1,5	6,5	42,3
Muestra 4	3	4	3	3	3,0	16,0	256,0
						Σ	732

**Hipótesis estadísticas:**

H<sub>0</sub>: No hay diferencia estadística significativa entre muestras. M1 = M2 = M3 = M4

H<sub>1</sub>: Al menos un tratamiento es estadísticamente diferente

**Regla de decisión:**

Si  $\chi_r^2 < \chi_{0,05}^2$  : Aceptar H<sub>0</sub>

Si  $\chi_r^2 > \chi_{0,05}^2$  : Aceptar H<sub>1</sub>

**Análisis del estadístico de Friedman:**

$$\chi_r^2 = \frac{12}{rt(t+1)} \sum R_i^2 - 3r(t+1) \rightarrow \chi_r^2 = \frac{12}{5(4)(4+1)} [732] - 3(5)(4+1)$$

$$\chi_r^2 = \frac{12}{20(5)} [732] - 15(5) = \frac{12}{100} (732) - 75 = 12,84$$

**Valores críticos:**

$$\chi_{0,05}^2 = 9,49 \text{ y } \chi_{0,01}^2 = 13,28$$

**Interpretación**

$$\chi_r^2 = 12,84 < \chi_{0,01}^2 = 13,28: \text{ Aceptar } H_0$$

$$\chi_r^2 = 12,84 > \chi_{0,05}^2 = 9,49: \text{ Aceptar } H_1$$

p de  $\chi^2 = 0,012 \sim 1,2 \%$

Confianza = 100 - 1,2% = 98,8 %

**Decisión:**

Al menos un tratamiento es estadísticamente diferente con el 98,8% de confianza.

**Ejercicio 155:**

Realizar la comparación de medias e identificar el tratamiento de mayor aceptabilidad, en ensayo de evaluación sensorial de cuatro muestras de cacao, a cargo de 5 analistas expertos, usando una escala ordinal de 1 a 10, donde 1 es muy poca aceptabilidad y 10 es alta aceptabilidad.

Este ensayo se caracteriza por: r = 5, t = 4

**Análisis del comprador estadístico de Friedman:**

$$CEF = Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{rt(t+1)}{6}} \rightarrow CEF = 1,96 \sqrt{\frac{5(4)(4+1)}{6}}$$

$$CEF = 1,96 \sqrt{\frac{20(5)}{6}} = 1,96(4,08) = 8,0$$

**Hipótesis estadísticas:**

H<sub>0</sub>: Las dos medias comparadas son estadísticamente iguales

H<sub>1</sub>: Las dos medias comparadas son estadísticamente diferentes

**Regla de decisión:**

$$|R_1 - R_2| < Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{rt(t+1)}{6}} : \text{ Aceptar } H_0$$

$$|R_1 - R_2| \geq Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{rt(t+1)}{6}} : \text{ Aceptar } H_1$$

**Comparaciones de Friedman:**

$$|R_1 - R_4| = 3,0 < CEF = 8,0: M1 = M4$$

$$|R_1 - R_2| = 10,5 > CEF = 8,0: M1 > M2$$

$$|R_1 - R_3| = 12,5 > CEF = 8,0: M1 > M3$$

El resumen de los análisis de rangos de Friedman, se expone en el Cuadro 187.

---

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

---

*Cuadro 187. Comparación de cuatro genotipos de arroz usando el método de Friedman*

Tratamientos	Medianas	Comparación	R	Diferencia de R	Z $\alpha/2$	Error estándar	CEF	
Muestra 1	9,0		19,0					
Muestra 4	7,0	M1 = M4	16,0	3,00	1,96	4,08	8	ns
Muestra 2	8,5	M1 > M2	8,5	10,50	1,96	4,08	8	*
Muestra 3	6,5	M1 > M3	6,5	12,50	1,96	4,08	8	*

Nota: M1 = M4, M1 > M2, M1 > M3. ns = no hay diferencia estadística significativa, \* = hay diferencia estadística con el 95 % de confianza.

***Decisión:***

La aceptabilidad es una variable del tipo "mayor es mejor". Las muestras M1 = M4 se identifican como más aceptadas con el 95 % de confianza.

## CAPÍTULO VII: INVESTIGACIÓN PREDICTIVA

La investigación predictiva tiene como propósito prever o anticipar situaciones futuras, requiere de la exploración, la descripción, la comparación, el análisis y la explicación de la causalidad de los fenómenos. El pronóstico tiene el propósito de la dirección futura de los eventos investigados, a partir de estudios exhaustivos de la evolución dinámica de los eventos, de su interrelación con el contexto, de las fuerzas de los actores y del estudio de las probabilidades.

Los términos pronóstico y predicción se usan indistintamente. El *pronóstico* es una aproximación a la ocurrencia de hechos futuros a partir del análisis de factores y correlaciones entre variables representados en modelos matemáticos y gráficos. La *predicción* se refiere a anunciar el comportamiento de hechos futuros. La probabilidad de que ocurra un hecho, siempre tendrá una medida de la incertidumbre ( $\alpha$ ) o de certidumbre ( $1 - \alpha$ ), expresada en términos de proporción entre 0 y 1 o de porcentaje de 0 a 100 % (Walpole et al., 2012, p. 279). La experimentación proporciona los datos e información en la que se basan los análisis predictivos.

En el contraste de hipótesis se compara las muestras a las cuales se aplicaron tratamientos. Las pruebas estadísticas más usadas son: F, t,  $\chi^2$  y Z. En la estimación de parámetros se proyecta el comportamiento de los parámetros de la población a partir de los estadísticos de las muestras.

### 7.1 INTERVALOS DE CONFIANZA DEL PARÁMETRO $\mu$

La estimación del parámetro  $\mu$  puede realizarse a partir del muestreo y cálculo de  $\bar{Y}$ , en series de datos univariados, el parámetro  $\sigma$  se puede calcular en las series de datos de las muestras y dependiendo del estudio, a partir del cuadrado medio del error del ADEVA.

$$\sigma = \sqrt{\text{CME}}$$

Los valores que toma la variable Y, en la predicción, depende de la distribución y del tamaño de muestra (n). Si las series de datos tienen distribución normal, entonces el parámetro  $\mu$  tiene límites, uno inferior y otro superior, que se explica con el modelo:

$$\mu = \bar{Y} \pm Z_{\alpha} \frac{S}{\sqrt{n}}$$

El valor Z no se asocia al tamaño de la muestra mientras que las distribuciones t y F son muy sensibles., en todo caso, conforme al teorema del límite central, mientras más grande sea el tamaño de la muestra más se aproxima hacia la normalidad.

En el análisis probabilístico, hay que calcular el error típico.

$$\text{Error típico: } \sigma_{\bar{Y}} = \frac{S}{\sqrt{n}}$$

La estimación del parámetro  $\mu$  es una aplicación de los intervalos de confianza (IC) a un nivel de significación concreto como  $\alpha = 0,05$ . La relación entre error típico e intervalo de confianza se indica en el Gráfico 52.

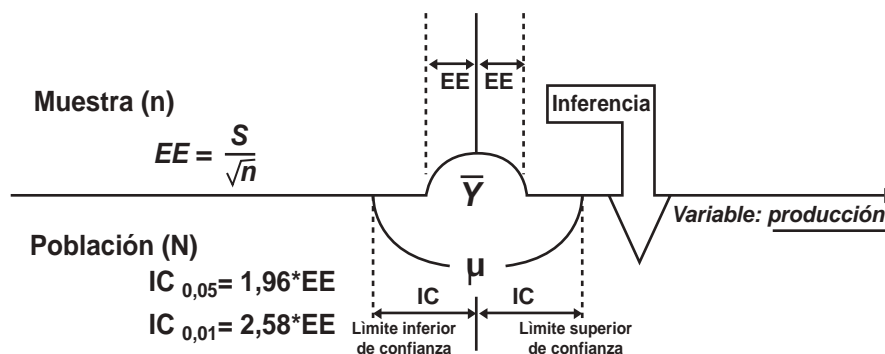


Gráfico 52. Relación entre media  $\pm$  error típico y los límites de confianza del parámetro  $\mu$

**Ejercicio 156:**

En una muestra aleatoria de 12 fincas dedicadas al cultivo de maíz ( $n = 12$ ) se aplicó una encuesta para conocer el rendimiento, en el último ciclo. Los estadísticos  $\bar{Y} \pm EE$  fueron:  $136 \pm 10$  quintales.hectárea<sup>-1</sup>. Se conoce también que en la comunidad hay 70 fincas que suman 245 hectáreas de maíz. ¿Cuál es la producción esperada en esa comunidad, con el 95 % de confianza?

**Datos y análisis:**

En una muestra aleatoria de productores de maíz ( $n = 12$  fincas), se estimó una producción media:  $\bar{Y} = 136 \pm 10$  quintales.ha<sup>-1</sup>.(Cuadro 188).

Cuadro 188. Datos de producción de maíz duro en una muestra aleatoria de 12 fincas

Fincas	Producción (quintales.ha <sup>-1</sup> ).	
1	85	
2	115	
3	160	
4	135	
5	150	
6	168	
7	210	
8	99	
9	105	
10	125	
11	135	
12	150	
N	12	fincas
Media	136	quintales
Mediana	135	quintales
Moda	135	quintales
Varianza	1178	(quintales) <sup>2</sup>
Desviación estándar	34	quintales
Error típico	10	quintales
Intervalo de confianza (95 %)	19,4	quintales
Límite inferior de $\mu$ ( $\alpha = 0,05$ )	117	quintales
Límite superior de $\mu$ ( $\alpha = 0,05$ )	156	quintales
Intervalo de confianza (99 %)	25,6	quintales
Límite inferior de $\mu$ ( $\alpha = 0,01$ )	111	quintales
Límite superior de $\mu$ ( $\alpha = 0,01$ )	162	quintales

**Cálculo de la producción de maíz a nivel de la comunidad:**

Estadísticos	Rendimiento (quintales.ha <sup>-1</sup> )	Producción (qq) estimada en 245 hectáreas
Límite inferior de $\mu$ ( $\alpha = 0,05$ )	117	28.665
Límite superior de $\mu$ ( $\alpha = 0,05$ )	156	38.220

**Interpretación:**

El rendimiento promedio de maíz es  $136 \pm 10$  quintales.hectárea<sup>-1</sup>, con lo cual se deduce que el parámetro  $\mu$  puede variar de 117 a 156 quintales.ha<sup>-1</sup> con una confianza del 95 %. Con estos datos se proyecta la producción de maíz por ciclo de cultivo, considerando una superficie de 245 hectáreas.

**Decisión:**

Con el 95 % de confianza, en las 245 hectáreas de maíz de la comunidad, se proyecta una producción entre 28.665 y 38.220 quintales.

## 7.2 MEDIAS MÓVILES

Una media móvil es un indicador de tendencia basado en la información subyacente de series temporales de datos donde se relaciona la información histórica de un período y los promedios de subconjuntos temporales. En base de este análisis se extrapola o interpola los datos y predice el comportamiento futuro. La extrapolación consiste en proyectar al futuro y la interpolación es estimar los valores intermedios (Alegre, 2022). Los promedios móviles se aplican en estudios demográficos, de compra - venta y en la bolsa de valores.

**Ejercicio 157:**

Usando las medias móviles, predecir las ventas de una empresa distribuidora de alimentos de la marca A, con base a la información acumulada de un año específico.

**Datos y análisis:**

La información mensual de ventas, en decenas, se indica en el Cuadro 189.

Cuadro 189. Ventas mensuales y anual, en decenas, de un producto alimenticio de la marca A.

2022	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.	$\Sigma V$
Ventas (decenas)	20	21	15	14	13	16	17	18	20	20	21	23	<u>218</u>

**Estimación 1:** Determinar las ventas proyectadas para enero del 2023 usando la media aritmética.

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{218}{12} = 18,2 \text{ decenas} \rightarrow 182 \text{ unidades}$$

**Estimación 2:** Determinar la tendencia de la demanda del producto usando una media móvil simple de los últimos tres meses:

Meses	Ventas del último trimestre (decenas)
Octubre	20
Noviembre	21
Diciembre	23
Total	64

$$\bar{Y} = \frac{21 + 23 + 20}{3} = 21,3 \text{ decenas} \rightarrow 213 \text{ unidades}$$

**Estimación 3:** Determinar la tendencia de la demanda usando una media móvil simple de dos periodos.

Meses	Ventas en los últimos dos meses (decenas)
Noviembre	21
Diciembre	23
Total	44

$$\bar{Y} = \frac{21 + 23}{2} = 22 \text{ decenas} \rightarrow 220 \text{ unidades}$$

**Estimación 4:** Pronosticar las ventas para enero del 2023 usando una media móvil de tres periodos, dando una ponderación del 50 % a diciembre, 40 % a noviembre y 10 % a octubre.

Meses	Ventas (decenas)	Ponderación	Valores ponderados
Octubre	20	0,10	2,0
Noviembre	21	0,40	8,4
Diciembre	23	0,50	11,5
Total	64		21,9

$$\bar{Y} = 20(0,10) + 21(0,40) + 23(0,50) = 21,9 \text{ decenas} \rightarrow 219 \text{ unidades}$$

**Cálculos de medias móviles:**

En el Cuadro 190, se expone las estimaciones de las medias móviles en  $k$  períodos y las desviaciones absolutas, así como la identificación de la predicción más segura.

*Cuadro 190. Ventas de un producto alimenticio y cálculo de las medias móviles con sus desviaciones absolutas*

Meses (2022)	Datos		Pronóstico			Desviaciones absolutas (D)		
	Promedio de ventas (decenas)	Media móvil k = 3	Media móvil k = 2	Media móvil ponderada (0,50-0,40-0,10)	Media móvil k = 3	Media móvil k = 2	Media móvil ponderada (0,50-0,40-0,10)	
Enero	20	--	--	--	--	--	--	
Febrero	21	--	--	--	--	--	--	
Marzo	15	--	20,50	--	--	--	--	
Abril	14	18,67	18,00	17,90	4,67	4,00	3,90	
Mayo	13	16,67	14,50	15,10	3,67	1,50	2,10	
Junio	16	14,00	13,50	13,60	2,00	2,50	2,40	
Julio	17	14,33	14,50	14,60	2,67	2,50	2,40	
Agosto	18	15,33	16,50	16,20	2,67	1,50	1,80	
Septiembre	20	17,00	17,50	17,40	3,00	2,50	2,60	
Octubre	20	18,33	19,00	18,90	1,67	1,00	1,10	
Noviembre	21	19,33	20,00	19,80	1,67	1,00	1,20	
Diciembre	23	20,33	20,50	20,50	2,67	2,50	2,50	
Pronóstico (decenas)		21,33	22,00	21,90				
Desviaciones absolutas promedios (decenas)					2,74	2,11	2,22	

**Interpretación:**

Media móvil con $K = 3$	Promedio 21,33 y desviación absoluta 2,74
Media móvil con $K = 2$	Promedio 22,00 y desviación absoluta 2,11
Media móvil ponderada	Promedio 21,90 y desviación absoluta 2,22

**Decisión:**

La media móvil con períodos  $k = 2$  (promedio de noviembre y diciembre) es más segura porque tiene la menor desviación absoluta, pronosticándose vender un promedio de 22 decenas ~ 220 unidades.

### 7.3 MODELACIÓN MATEMÁTICA

El análisis de regresión estudia la relación funcional que existe entre dos o más variables e identifica el modelo matemático que asocia la variable dependiente (Y) en función de una o más variables independientes (X) y estima los parámetros (Dagnino, 2014, p. 146). En un dispersograma se aprecia una nube de puntos, donde cada punto es una observación. En el modelo de causalidad:  $Y = f(X)$ , intervienen dos variables pero se desconoce el tipo de relación que puede ser lineal o no lineal (Bouza, 2018, p. 10). La relación es lineal si en el dispersograma los datos se distribuyen alrededor de una línea recta, la misma que se obtiene a través de la ecuación de la recta. En las relaciones no lineales, que es más frecuente en los análisis de las funciones de producción, los datos se distribuyen alrededor de una curva, cuya ecuación se determina a partir de alguno de los modelos no lineales como: cuadrática, cúbico, logarítmico y de otros tipos, todos desarrollados con fines predictivos (Dagnino, 2014, p. 148).

Los valores altos del coeficiente de correlación  $r$ , cercanos a 1 también tendrán altos coeficientes de determinación ( $R^2$ ) que son indicativos de la pertinencia de las variables que formarían parte de los modelos. Los modelos de predicción permiten comprender las relaciones entre los componentes del sistema productivo. Por ejemplo, con base a los datos de un huésped, fitopatógeno y elementos meteorológicos se puede predecir la evolución de un fitoparásito. En esta perspectiva, la vigilancia epidemiológica se enfoca en el monitoreo de los componentes: clima, ambiente, patógeno y circunstancias socioeconómicas para conocer los hechos y predecir el comportamiento fitosanitario.

Para aplicar la regresión lineal, los datos deben cumplir cinco supuestos: Linealidad, independencia, homocedasticidad, normalidad y no colinealidad (Bouza, 2018, p. 45; Rojo, 2007, p. 6). La linealidad se refiere a una propiedad abstracta definida entre las funciones X e Y como en los espacios en de relación causa – efecto. El concepto integra las propiedades de escalamiento y de aditividad. La independencia, en la teoría de probabilidades, indica que dos sucesos aleatorios son independientes entre sí, cuando la probabilidad de cada uno de ellos no está influida por lo que ocurra en otro suceso. Dos variables son estadísticamente independientes cuando el comportamiento de una de ellas no se ve afectado por los valores que toma la otra. La homocedasticidad se refiere a la homogeneidad de las varianzas, que expresa las dispersiones de los datos que tienen que ser razonablemente uniformes aun cuando las medias sean distintas.

La normalidad de las series de datos, en las variables independientes y dependientes tienen que cumplirse para que las conclusiones tengan validez. La normalidad de los errores permite estimar los intervalos de confianza para los coeficientes de regresión. La no colinealidad define que, en una muestra, las variables explicativas no pueden ser constantes. Los predictores ( $X_1, X_2, X_3$ ) se correlacionan formando una combinación lineal y no puede



haber autocorrelaciones. Por ejemplo: la producción de grano en gramos.planta<sup>-1</sup> no puede usarse para explicar el rendimiento en kg.ha<sup>-1</sup>.

Un modelo se ajusta a la serie de datos siempre que las diferencias entre los valores observados y los de predicción sean pequeñas y no presenten sesgo. El dispersograma y el Coeficiente de correlación, complementado con el coeficiente R<sup>2</sup>, facilitan la comprensión de la asociación entre variables. Este último se calcula como el cuadrado del coeficiente de correlación *r* de Pearson (paramétrico) o *r<sub>s</sub>* de Spearman (no paramétrico) e indica si los casos están cerca o se alejan de la línea de regresión (Cardona et al. 2014, p. 155).

En estudios de fitomejoramiento, el Coeficiente de determinación R<sup>2</sup> equivale al parámetro heredabilidad (Robles, 1986, p. 140; Pistorale et al., 2008, p. 261):

$$H = R^2 = \frac{\text{varianza explicada}}{\text{varianza total}} \quad (100)$$

La varianza explicada corresponde al cuadrado medio del factor genético en estudio: progenies, híbridos, líneas, variedades o clones. Otras formas de calcular R<sup>2</sup> se basa en la fórmula siguiente:

$$R^2 = 1 - \frac{\sum d^2}{\sum (Y - \bar{Y})^2} \quad [129]$$

Usando el programa Excel, se obtiene el Coeficiente de determinación con la función: =COEFICIENTE.R2(Xconocido; Yconocido).

El Coeficiente de determinación ajustado  $\bar{R}^2$  es una medida que define la proporción explicada por la varianza de la regresión, tomado en cuenta el número de variables explicativas (independientes), en modelos multivariantes. El  $\bar{R}^2$  penaliza la inclusión de *k* variables en un modelo matemático.

La fórmula del Coeficiente de determinación ajustado es la siguiente:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} [1 - R^2] \quad [130]$$

Dónde:

$\bar{R}^2$  = Coeficiente de determinación ajustado

n = Tamaño de la muestra

k = Número de variables explicativas.

R<sup>2</sup> = Coeficiente de determinación no ajustado

Las funciones básicas asociadas a los modelos matemáticos más comunes son:

Función básica: Y = f(X)	Función equivalente	Tipo de función
Y = β <sub>0</sub> + β <sub>1</sub> X	Y = a + b.X	Lineal
Y = β <sub>0</sub> + β <sub>1</sub> X + β <sub>2</sub> X <sup>2</sup>	Y = a + b.X + c.X <sup>2</sup>	Cuadrática
Y = β <sub>0</sub> + β <sub>1</sub> X + β <sub>2</sub> X <sup>2</sup> + β <sub>3</sub> X <sup>3</sup>	Y = a + b <sub>1</sub> .X + b <sub>2</sub> .X <sup>2</sup> + b <sub>3</sub> .X <sup>3</sup>	Cúbica
Y = β <sub>0</sub> + β <sub>1</sub> X + β <sub>2</sub> X <sup>2</sup> + ... + β <sub>n</sub> X <sup>n</sup>	Y = a + b <sub>1</sub> .X + b <sub>2</sub> .X <sup>2</sup> + ... + b <sub>n</sub> X <sup>n</sup>	Polinómica general
Y = a.b <sup>X</sup>		Exponencial
Y = log(X)		Logarítmica
Función multivariada:		
Y = f(X <sub>1</sub> , X <sub>2</sub> , ... X <sub>n</sub> )	Y depende de varias variables independientes	
Y <sub>1</sub> , Y <sub>2</sub> = f(X <sub>1</sub> )	Variables dependientes de una independiente	
Y <sub>1</sub> , Y <sub>2</sub> = f(X <sub>1</sub> , X <sub>2</sub> )	Variables dependientes de varias independientes	

Se han desarrollado modelos de pronóstico en todos los campos del conocimiento. A continuación se indican algunos ejemplos. Por ejemplo, el modelo de pronóstico de la incidencia de enfermedades en trigo, fue propuesto por Moschini et al. (2013, p. 16):

$$E_t = E_0 \cdot e^{rt}$$

Dónde:

$E_t$  = Nivel de enfermedad en el tiempo  $t$

$E_0$  = Nivel de enfermedad en el momento actual, en una muestra representativa

$rt$  = Tasa de crecimiento relativo de la enfermedad en el tiempo pronosticado

Un modelo de pronóstico de la incidencia de hojas enfermas en cafetos arábigos, propuesto por Sotomayor y Duicela (1995, p. 51), es el siguiente:

$$HE = -296 - 0,041(PP) + 2,5(HR) + 3,35(TMáx)$$

Dónde:

HE = Porcentaje de hojas enfermas en cafetos

PP = Precipitación mensual (mm)

HR = Humedad relativa mensual (%)

TMáx = Temperatura máxima mensual (°C)

Un modelo de predicción de la producción, en gramos.planta<sup>-1</sup>, en clones de café robusta, lo propone Duicela (2021, p. 79):

$$PCOP = 5,8(NR) + 15,5(NNR)$$

Dónde:

PCOP = Producción de café oro por planta (g)

NR = Número de ramas

NNR = Número de nudos por rama

La literatura especializada informa de muchos modelos linealizados de ajuste a curvas de progreso epidémico para el pronóstico de enfermedades en cultivos de alto valor comercial (Bombelli et al., 2013, p. 56), que ciertamente deben ser validados en las diferentes circunstancias agrosociales y ambientales.

En microeconomía agrícola se usa los modelos matemáticos que fundamentan la teoría de la función producción:  $Y = f(X)$ , que define la relación insumos (inputs) y el producto obtenido (outputs), en condiciones tecnológicas concretas (Sydsaeter y Hammond, 1996, p. 32; Doll & Orazem, 1978, p. 21). El modelo lineal en una función producción indica que aún falta por explorar más cuadrático se ajusta mayores dosis de insumo. El modelo cuadrático as explica mejor la realidad productiva agropecuaria.

Los insumos agropecuarios son los materiales usados en la producción agropecuaria, que pueden ser de naturaleza biológica y no biológica, dependen del sistema de producción (orgánica o convencional), del nivel tecnológico (intensivo o extensivo) o de los objetivos productivos. Los insumos pueden agruparse en categorías: (a) semilla/material de siembra, (b) fertilizantes, (c) plaguicidas y (d) equipo agrícola. Todos estos insumos pueden ser optimizados. El organismo oficial de regulación y control en Ecuador detalla los insumos de uso agrícola permitidos (Agrocalidad, 2020, pp.10-18). La selección de insumos y el uso adecuado en dosis óptimas, puede contribuir a mejorar significativamente la productividad, calidad e inocuidad de las cosechas.

7.3.1 MODELO LINEAL

La regresión de los valores de una variable con respecto a otra es significativa cuando se ajusta a los valores observados, identifica relaciones potencialmente causales o cuando no existen dudas sobre su relación causal. La regresión lineal estudia el efecto de una variable independiente sobre una única variable dependiente (Gómez & Gómez, 1983, p. 359).

El modelo lineal se representa por la forma.

$$Y = a + b(X) + \varepsilon$$

Dónde:

Y = Valor de la variable dependiente

a = Valor del intercepto (cuando sea pertinente)

b = Valor de la pendiente

X = Variable independiente

$\varepsilon$  = Error experimental

Las distribuciones de todos los valores posibles de  $a$  y  $b$  tienen una distribución normal con medias de  $a$  y de  $b$ , y desviaciones estándar  $S_a$  y  $S_b$  que se denominan errores estándar del intercepto y de la pendiente de la recta, respectivamente (Gráfico 53). Estos errores estándar se usan como en los errores típicos de las medias, para calcular intervalos de confianza y realizar las pruebas de hipótesis usando la distribución de  $t$ .

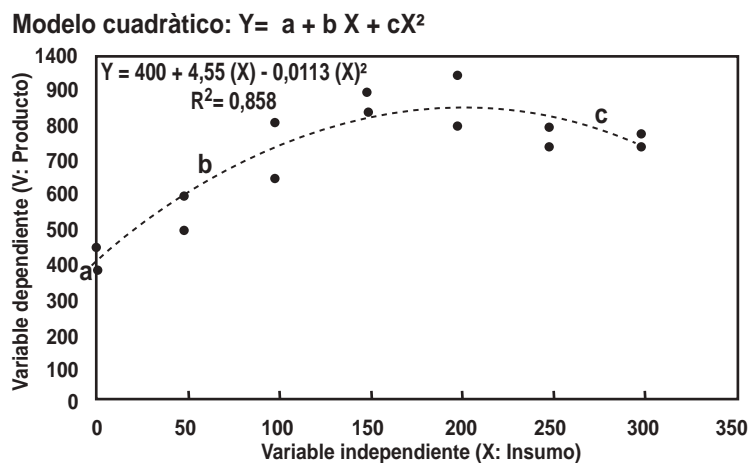
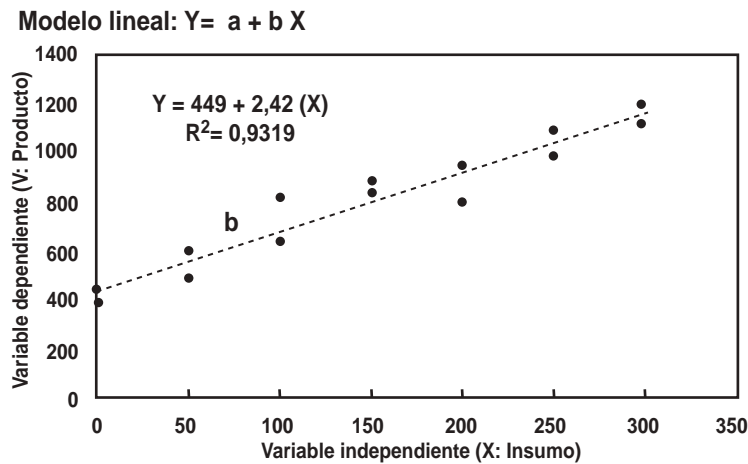


Gráfico 53. Modelos lineal y cuadrático que se ajustan a la función insumo producto

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

Cuando se analiza la respuesta productiva en un cultivo como el maíz, asumiendo que depende de las dosis de fertilizante nitrogenado,  $X = \text{g.planta}^{-1}$ . Si la dosis de nitrógeno es cero ( $X = 0$ ), hay un nivel de producción que va a variar entre repeticiones, configurando la dispersión. En  $X = 0$  habrá valores de  $Y$  que corresponden al intercepto  $a$  y a medida que se adiciona fertilizante ( $X_1, X_2, \dots X_n$ ), cambiarán los valores de  $Y_i$ .

El valor  $b$  corresponde al coeficiente de la regresión lineal y se calcula con la fórmula:

$$b = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{\Delta Y}{\Delta X}$$

**Ejercicio 158:**

En un ensayo se prueban frecuencias de aplicación de un Bioestimulante foliar en pastizales y se mide la productividad anual neta primaria, en gramos.parcela<sup>-1</sup>. ¿Elaborar un modelo lineal que estime la productividad esperada?

**Datos y análisis:**

El diseño de los tratamientos y de los datos de  $X$  e  $Y$ , se expone en el Cuadro 191. El modelo matemático que resultó del análisis de regresión es el siguiente:

$$Bms = 125 + 5,66 (Es)$$

Dónde:

Bms = Producción de biomasa como materia seca (g.parcela<sup>-1</sup>)

Es = Cantidad de bioestimulante aplicado en el año (mL.parcela<sup>-1</sup>).

El intercepto es  $a = 125 \text{ g.parcela}^{-1}$  y la pendiente es  $b = 5,663$ . El coeficiente de la regresión “b” indica que por cada mL de bioestimulante aplicado al pastizal, se incrementa la biomasa seca en  $5,66 \text{ g.parcela}^{-1}$ . En el Gráfico 54, se indica la relación tratamientos y producción de biomasa, el modelo lineal y el Coeficiente de determinación ( $R^2$ ). En el Cuadro 192, se expone el ADEVA de la regresión, que resultó significativa ( $p < 0,01$ ), por tanto,  $Y$  depende de  $X$  con un coeficiente de determinación de 96,4 %.

*Cuadro 191. Efecto de la aplicación de bioestimulante sobre la producción de pasto*

Repeticiones	Aplicaciones por año	Bioestimulante (mL.parcela <sup>-1</sup> ) (X)	Producción (g.parcela <sup>-1</sup> ) Y	Predicción de la producción Y'	Residuo Y - Y'
1	1	300	1500	1824	-324
2	1	300	1600	1824	-224
3	1	300	1400	1824	-424
4	1	300	2200	1824	376
1	2	600	3200	3523	-323
2	2	600	3900	3523	377
3	2	600	4200	3523	677
4	2	600	3700	3523	177
1	3	900	5800	5222	578
2	3	900	5350	5222	128
3	3	900	4900	5222	-322
4	3	900	4800	5222	-422
1	4	1200	7000	6921	79
2	4	1200	6500	6921	-421
3	4	1200	7200	6921	279
4	4	1200	6700	6921	-221
Media (g.parcela <sup>-1</sup> )			4372	4372	0

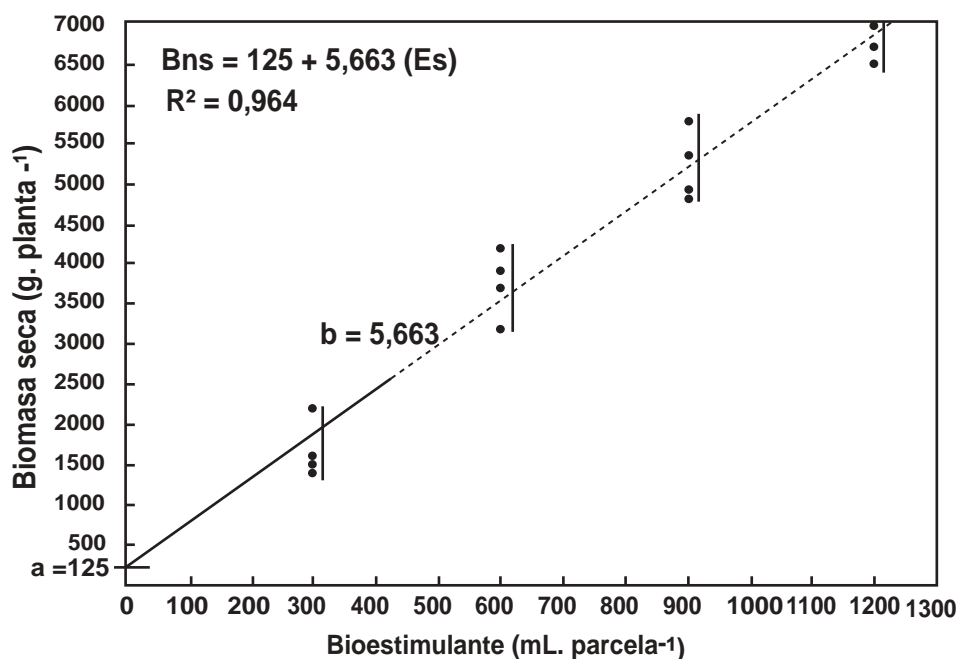


Gráfico 54. Producción neta de pasto en función de las dosis de bioestimulante

Cuadro 192. Análisis de la varianza de la regresión de la variable producción de biomasa en función de la aplicación de bioestimulante

Origen de la variación	GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p	Variación (%)
Regresión	1	57.715.031	57.715.031	13,13 **	4,60	8,86	0,003	96
Errores de la recta	14	2.154.813	153.915					4
Total	15	59.869.844						100

Nota: \*\* = La regresión entre las variables X, Y fue altamente significativa

El Coeficiente de determinación  $R^2 = 0,964$  se calcula elevando al cuadrado el Coeficiente  $r = 0,982$ . El Coeficiente  $R^2$  ajustado, se calcula con la fórmula siguiente:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} [1 - R^2] \rightarrow \bar{R}^2 = 1 - \frac{16-1}{16-1-1} [1 - 0,964] \rightarrow \bar{R}^2 = 0,96$$

El valor  $k$  se refiere al número de variables independientes usadas en la regresión. El Coeficiente de determinación está muy cercano a la unidad, por lo tanto, el análisis de la regresión es confiable, explica bien la relación de Y en función de X.

**Interpretación:**

Cada mililitro de bioestimulante/parcela posibilita incrementar la producción primaria neta en 5,66 g.parcela<sup>-1</sup>, con un  $\bar{R}^2 = 96,0\%$ , valor que coincide con la contribución a la variación indicada en el análisis de regresión.

**Decisión:**

El modelo lineal,  $Bms = 125 + 5,66(Es)$ , explica la causalidad de la producción de materia seca en función de las aplicaciones del bioestimulante. A mayor dosis de bioestimulante mayor producción. Se recomienda probar dosis más altas.

**Ejercicio 159:**

Realizar un análisis de regresión de la producción de azúcar blanca (Y= en toneladas) en función de la materia prima (X = toneladas de caña de azúcar), con la información de 12 días de operación del ingenio

**Datos y análisis:**

Los datos de caña de azúcar y azúcar blanca se exponen en el Cuadro 193.

Cuadro 193. Rendimiento azucarero en función de la materia prima

Materia Prima (t)	→300	100	290	330	350	380	430	480	515	550	630	650
Azúcar blanca (t)	→31,5	10,1	33,6	32,43	37,8	46	45	57,6	59	61,6	61,8	78

**Línea de tendencia:**

En el programa Excel, se elabora la línea de tendencia, en este caso ajustado al modelo lineal sin intercepto (Gráfico 55); así como, se calcula el Coeficiente de determinación,  $R^2 = 0,99 \sim 99 \%$ , y el Coeficiente de regresión,  $b = 0,111$  indicativo de que por cada tonelada de caña se produce 0,111 toneladas de azúcar blanca.

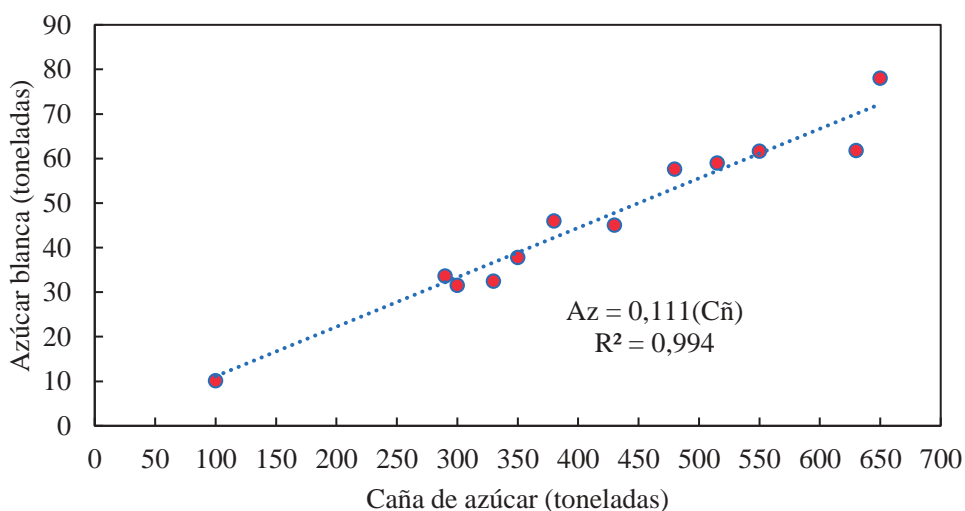


Gráfico 55. Producción de azúcar blanca en función de la materia prima

**Interpretación:**

El punto  $X_0$  coincide con  $Y_0$ , por tanto es una función donde no hay intercepto. El coeficiente de regresión indica que “una tonelada de caña de azúcar permite obtener 0,111 toneladas de azúcar blanca”.

De esta información, también se deduce otro parámetro productivo:

Si  $Az = 0,111(Cñ)$

¿Cuántas toneladas de caña se requiere para producir una tonelada de azúcar?

$$Cñ = \frac{Az}{0,111} = \frac{1}{0,111} = 9,009 \text{ unidades}$$

Para producir una tonelada de azúcar blanca se necesita 9,009 toneladas de materia prima.

El Coeficiente de determinación ajustado  $\bar{R}^2$  se calculó con la fórmula:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} [1 - R^2]$$

Dónde:

n = Pares de observaciones

k = Número de variables independientes

R<sup>2</sup> = Coeficiente de determinación sin corregir

$$\bar{R}^2 = 1 - \frac{12-1}{12-1-1} [1 - 0,994] \rightarrow \bar{R}^2 = 1 - \frac{11}{10} [1 - 0,994] \rightarrow 0,993 \sim 99,3 \%$$

El coeficiente M/P:  $\frac{\text{Materia prima}}{\text{Producto}} = \frac{1}{0,11} = 9,009$

Para tener una 1 tonelada de producto se requiere = 9,009 t de materia prima.

**Decisión:**

Por cada tonelada de caña de azúcar se produce 0,111 toneladas de azúcar blanca” con un coeficiente de determinación ajustado  $\bar{R}^2 = 99,3 \%$ .

7.3.2 MODELO POLINÓMICO DE GRADO 2

Una relación insumo: producto se concibe como una función: Y = f (X). Analizar con un modelo lineal simple genera incertidumbre, porque se deduce que la zona explorada de la variable independiente solo cubre una parte de las posibilidades del estudio. Se evidencia que hay una zona inexplorada de la variable X que no cubren las posibilidades de respuesta de la variable Y (Gráfico 56).

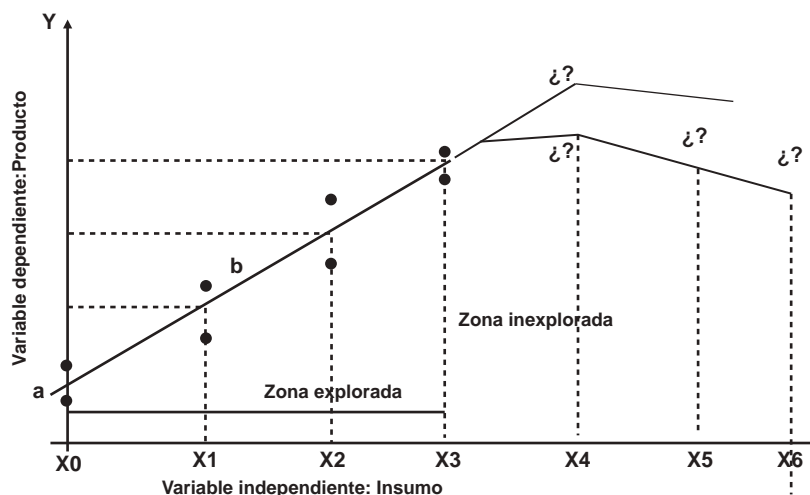


Gráfico 56. Incertidumbre de las respuestas posibles en la zona inexploradas

Un modelo polinómico de grado 2 explica de mejor forma este fenómeno de causalidad (Walpole et al., 2012, p. 451; Gómez & Gómez, 1983, p. 388). El modelo cuadrático se expresa como la función: Y = f (X), siendo su forma básica la siguiente:

$$Y = a + bX + cX^2$$

Dónde:

Y = Variable dependiente

X = Variable explicativa o independiente





La fórmula para determinar la cantidad de insumo requerida para maximizar la producción es la siguiente:

$$X_{Máx} = \frac{b}{2|c|} \quad [131]$$

Dónde:

$X_{Máx}$  = Cantidad de insumo requerida para obtener la producción máxima

b = Coeficiente de la expresión lineal

|c| = Valor absoluto del Coeficiente de la expresión cuadrática

La fórmula usada para determinar la cantidad de insumo requerida para maximizar el beneficio neto (optimizar el uso de insumos) es la siguiente:

$$X_{POE} = \frac{b - \frac{C_x}{P_y}}{2|c|} \quad [132]$$

Dónde:

$X_{POE}$  = Cantidad de insumo requerida para obtener una producción óptima económica

b = Coeficiente de la expresión lineal

|c| = Valor absoluto del Coeficiente de la expresión cuadrática

$\frac{C_x}{P_y}$  = Es la relación entre Costo unitario del insumo y precio unitario de venta del producto.

### Ejercicio 160:

Analizar los datos del ensayo “Efecto de distintas dosis de fertilizante nitrogenado sobre el rendimiento cultivo de maíz”.

#### Datos y análisis:

En los datos del ensayo de maíz, del Cuadro 194, se constata que usando la dosis 140 kg.ha<sup>-1</sup> de abono (tratamiento 3) se obtienen los rendimientos más altos. Analizando con los valores EAM, el tratamiento 3 equivale a rendimientos que varían de 92 a 100 %, mientras que en el testigo varían de 27 a 35 %. El análisis de regresión proporciona información requerida para maximizar la producción u optimizar el uso de insumos.

Cuadro 194. Datos de rendimiento del maíz en función de la fertilización nitrogenada

Repeticiones	Tratamientos	X1	Y
		Abono N (kg.ha <sup>-1</sup> )	Rendimiento (kg.ha <sup>-1</sup> )
1	1	0	2960
2	1	0	3188
3	1	0	2500
1	2	70	7400
2	2	70	6900
3	2	70	7046
1	3	140	9200
2	3	140	9200
3	3	140	8500
1	4	210	7500
2	4	210	7200
3	4	210	7350

Usando las herramientas de Excel se obtuvo la línea de tendencia, el modelo cuadrático (polinómico de grado 2) y el Coeficiente de determinación R<sup>2</sup> (Gráfico 58). El modelo cuadrático operacional es el siguiente:

$$REN = 2828 + 84,4(N) - 0,2984(N^2)$$

El intercepto tiene el valor 2828 kg.ha<sup>-1</sup>, esta producción se estima cuando no se aplica fertilizante (X = 0), el coeficiente b = 84,64 indica que hay un promedio de incremento de la producción por cada unidad de insumo que se adicione, sin embargo, este incremento tiene un límite, hay un punto de inflexión en la curva que indica una reducción progresiva de la producción, con un coeficiente c = - 0,2984.

En el Cuadro 195, se expone la predicción Y<sub>p</sub> de la producción en base a la aplicación del modelo cuadrático, obtenido a través de la regresión.

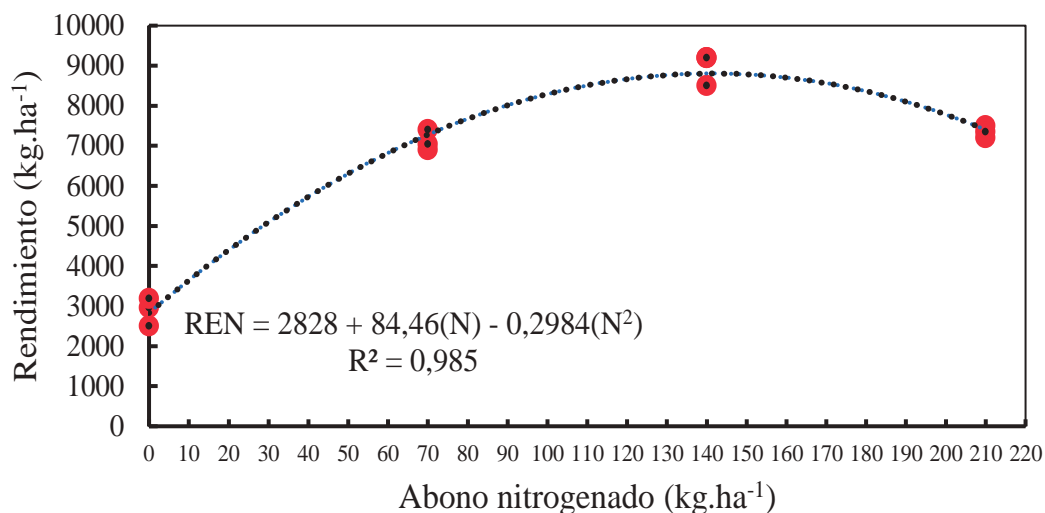


Gráfico 58. La producción de maíz en función de la fertilización nitrogenada

Cuadro 195. Relación entre variable explicativa (X) y productos resultantes observado (Y) y predicho (Y<sub>p</sub>)

Insumo (X) (variable independiente)	Producción (Y) observada (variable dependiente)	Producción Y <sub>p</sub> de predicción	Residuo (Y - Y <sub>p</sub> )
0	2960	2828	132
0	3188	2828	360
0	2500	2828	-328
70	7400	7278	122
70	6900	7278	-378
70	7046	7278	-232
140	9200	8804	396
140	9200	8804	396
140	8500	8804	-304
210	7500	7405	95
210	7200	7405	-205
210	7350	7405	-55
Media	6579	6579	0

**Maximización de la producción usando el cálculo diferencial:**

El modelo cuadrático se somete a la primera derivada y luego se iguala a cero (asocia al punto más alto de la curva de la función producción), resolviendo la ecuación:

$$REN = 2828 + 84,46(N) - 0,2984(N^2)$$

La primera derivada se iguala a cero y resuelve la ecuación para determinar el valor de X asociado a la producción máxima:

$$\frac{d(Y)}{d(X)} = 0 + 84,46 - 2(0,298N) = 0 \rightarrow 84,46 - 0,5968(N) = 0$$

Cálculo de  $X_{Máx} = N$ :

$$N = \frac{84,46}{0,5968} = 142 \text{ kg.ha}^{-1} \text{ de abono nitrogenado}$$

$Y_{Máx} = REN$  (Máximo). La producción máxima esperada se obtiene reemplazando  $N = 142 \text{ kg.ha}^{-1}$  en el modelo cuadrático de la función producción:

$$REN (\text{Máx}) = 2828 + 84,46(142) - 0,2984(142^2) = 8804 \text{ kg.ha}^{-1}$$

### **Maximización de la producción usando fórmula alternativa:**

Se resumen los coeficientes de la regresión cuadrática:

$$a = 2828, b = 84,46 \text{ y } c = -0,2984$$

La fórmula usada para determinar la cantidad de insumos asociada a la producción máxima es la siguiente:

$$X_{Máx} = \frac{b}{2|c|} \rightarrow X_{Máx} = \frac{84,46}{2(0,2984)} = 142 \text{ kg.ha}^{-1}. \text{ Cantidad de N requerida para lograr } Y_{Máx}.$$

La producción máxima esperada se obtiene reemplazando  $N = 142 \text{ kg.ha}^{-1}$  en la función producción:

$$REN (\text{Máx}) = 2828 + 84,46(142) - 0,2984(142^2) = 8804 \text{ kg.ha}^{-1}.$$

### **Optimización en el uso de insumos usando la primera derivada:**

Para optimizar un proceso productivo se requiere de datos económicos, como: el costo de un saco de 50 kg de abono (\$38,0 el saco equivale a  $C_x = \$0,76.\text{kg}^{-1}$ ) y si el precio proyectado de venta de un quintal de maíz (\$14,0 el quintal equivale a  $P_y = \$0,307.\text{kg}^{-1}$ ).

$$\text{El cálculo de la relación: } \frac{C_x}{P_y} \rightarrow \frac{C_x}{P_y} = \frac{0,76}{0,307} = 2,4722$$

Para determinar la cantidad de insumo asociada al máximo beneficio neto, se iguala la primera derivada a la relación  $\frac{C_x}{P_y}$  y se resuelve la ecuación resultante.

$$\frac{d(Y)}{d(X)} = 0 + 84,46 - 2(0,298N) = 2,4722$$

$$84,46 - 0,5968N = 2,4722 \rightarrow 84,46 - 2,4722 = 0,5968 (N) \rightarrow 81,988 = 0,5968 N$$

Nitrógeno  $POE = 81,988/0,5968 = 137 \text{ kg.ha}^{-1}$ . Esta es la cantidad de insumo ( $X_{POE}$ ) requerida para lograr el máximo beneficio neto ( $Y_{POE}$ ):

$$Y_{POE} = 2828 + 84,462 (137) - 0,2984 (137)^2 = 8799 \text{ kg.ha}^{-1}$$

### **Optimización de la producción usando fórmula alternativa**

La relación  $\frac{C_x}{P_y} = \frac{0,76}{0,307} = 2,4722$  se calcula con los datos actuales o proyectados de Costo unitario del insumo y precio unitario de venta del producto.

El cálculo de la cantidad de insumo ( $X_{POE}$ ) asociada al POE es el siguiente:

$$X_{POE} = \frac{b - \frac{C_x}{P_y}}{2|c|} \rightarrow X_{POE} = \frac{84,46 - 2,4722}{2(0,2984)} = 137 \text{ kg.ha}^{-1} \text{ de fertilizante nitrogenado .}$$

$$Y_{POE} = 2828 + 84,462 (137) - 0,2984 (137)^2 = 8799 \text{ kg.ha}^{-1}$$

**Decisión:**

La cantidad óptima de fertilizante requerido es  $X_{POE} = 137 \text{ kg.ha}^{-1}$ . La producción en el punto óptimo económico  $Y_{POE} = 8799 \text{ kg.ha}^{-1}$ .

7.3.3 MODELO POLINÓMICO DE GRADO 3

El análisis de regresión cúbica es una técnica usada cuando la relación entre una variable predictora y una variable de respuesta no es lineal y la curva tiende a mostrar una forma sigmoidea (Cardona et al., 2014, p. 154)). Los coeficientes del modelo cúbico (polinómica de grado 3), se determinan los valores máximos, mínimos y el punto de inflexión, tanto en producción como en beneficio neto.

**Ejercicio 161:**

Determinar las dosis óptima de insecticida para el manejo de una plaga insectil, en base al análisis del ensayo: Efecto de distintas dosis de insecticida sobre la mortalidad de insectos plaga en un ambiente específico.

**Datos y análisis:**

Los datos del ensayo y los cálculos de los componentes del modelo cúbico se exponen en el Cuadro 196.

*Cuadro 196. Componentes del modelo polinómico de grado 3 relacionado al número de insectos muertos en función de las dosis de insecticida*

Insecticida (mL.parcela <sup>-1</sup> ) (X)	Insectos muertos (%) Y	X <sup>2</sup>	X <sup>3</sup>	B <sub>0</sub>	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	Predicción Y'	Residuo Y - Y'
0,0	3	0,00	0,00	2,754	-7,362	17,11	-4,963	2,8	0,25
0,5	2	0,25	0,13	2,754	-7,362	17,11	-4,963	2,7	-0,73
1,0	8	1,00	1,00	2,754	-7,362	17,11	-4,963	7,5	0,46
1,5	14	2,25	3,38	2,754	-7,362	17,11	-4,963	13,5	0,54
2,0	16	4,00	8,00	2,754	-7,362	17,11	-4,963	16,8	-0,77
2,5	14	6,25	15,63	2,754	-7,362	17,11	-4,963	13,7	0,26

**Modelo matemático de grado 3:**

El modelo cúbico adquiere la forma:

$$Y = a + b_1.X + b_2.X^2 + b_3.X^3$$

El modelo matemático polinómico de grado 3, calculado con Excel es el siguiente:

$$M = 2,754 - 7,362(X) + 17,11(X^2) - 4,963(X^3)$$

Por lo tanto, los coeficientes del modelo son:

$$Y = \text{mortalidad (\%)}, X = \text{dosis de insecticida (mL.parcela}^{-1}\text{)}$$

$$b_0 = 2,724, b_1 = - 7,362, b_2 = 17,11 \text{ y } b_3 = - 4,963.$$

Cuando  $X = 0$ , la mortalidad esperada sería 2,75 % que se mantiene más o menos constante hasta la dosis 0,5 mL.parcela<sup>-1</sup>. La mortalidad se incrementa hasta 16 % con la dosis de 2,0 mL.parcela<sup>-1</sup> y luego la mortalidad tiende al estancamiento (Gráfico 59). Los puntos críticos para máxima y mínima mortalidad de los insectos plaga, así como del punto de inflexión o momento en que el insecticida empieza a perder efectividad, se determinan usando el cálculo diferencial.

***Cálculo de la primera derivada:***

Función mortalidad de los insectos plaga:

$$M = 2,754 - 7,362(X) + 17,11(X^2) - 4,963(X^3)$$

La primera derivada de la función es la siguiente:

$$\frac{dM}{dX} = 0 - 7,362 + 2(17,11)(X) - 3(4,963)(X^2) = 0$$

$$\text{Por lo tanto: } - 7,362 + 34,22(X) - 14,89(X^2) = 0$$

La solución de la ecuación cuadrática (con calculadora<sup>37</sup> para este propósito fue:

$X_1 = 0,240$  mL.parcela<sup>-1</sup>, indica la dosis de insecticida para tener la mínima mortalidad

$X_2 = 2,06$  mL.parcela<sup>-1</sup> es la dosis de insecticida asociada a la máxima mortalidad

Mortalidad de insectos plaga en el punto mínimo:

$$M = 2,754 - 7,362(0,24) + 17,11(0,24^2) - 4,963(0,24^3) = 1,9 \%$$

Mortalidad de insectos plaga en el punto máximo:

$$M = 2,754 - 7,362(2,06) + 17,11(2,06^2) - 4,963(2,06^3) = 16,8 \%$$

***Cálculo de la segunda derivada:***

La segunda derivada permite calcular el punto de inflexión, que indica el cambio de los valores de un tipo de concavidad a otra.

$$\frac{dM}{dX} = 34,22X - 2(14,89)(X) = 0$$

$$34,22 - 29,78(X) = 0 \rightarrow \frac{34,22}{29,78} = 1,15 \text{ mL.parcela}^{-1}.$$

A partir de esta dosis, la eficiencia del insecticida va reduciendo.

***Cálculo de la mortalidad en el punto de inflexión:***

$$M = 2,754 - 7,362(1,15) + 17,11(1,15^2) - 4,963(1,15^3) = 9,4 \%$$

---

<sup>37</sup> Symbolab. (noviembre 4, 2022). *Calculadora de ecuaciones cuadráticas*. <https://es.symbolab.com/solver/step-by-step/x>

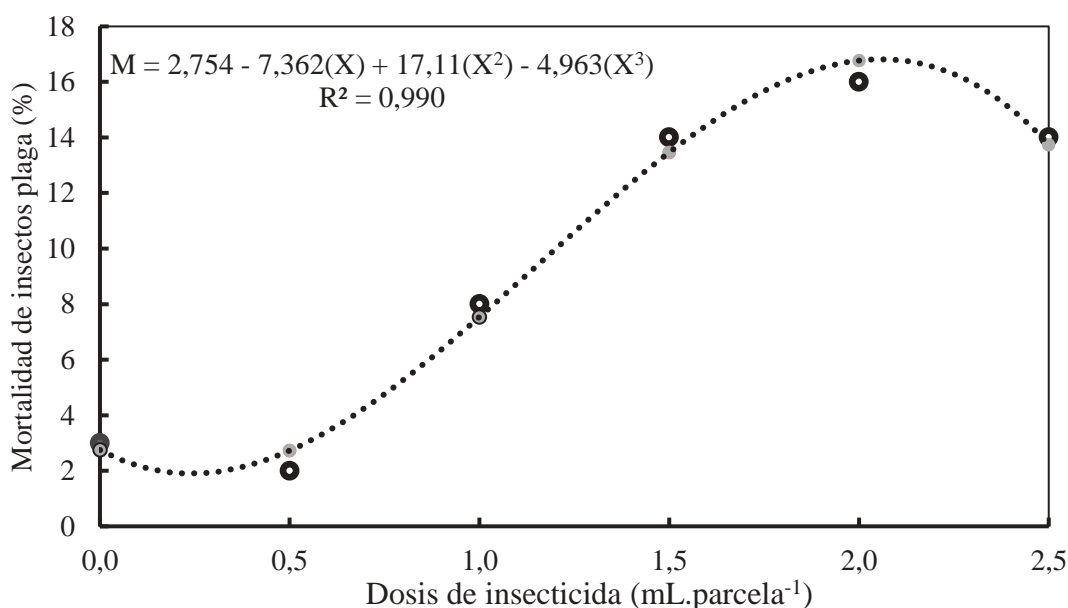


Gráfico 59. Modelo característico del comportamiento de una plaga en función del plaguicida

### 7.3.4 MODELO MULTIVARIADO

En la regresión múltiple, se ajusta un modelo que utiliza múltiples variables independientes para predecir la variable dependiente. El objetivo es encontrar la mejor combinación de variables independientes que explique la mayor cantidad de variabilidad en la variable dependiente. El modelo multivariado trata de explicar la causalidad de un hecho (variable dependiente (Y) a través de dos o más variables independiente ( $X_1, X_2, \dots, X_n$ ). Las series de datos deben cumplir condiciones como: tener sentido numérico, no haber variables redundantes, las variables deben tener justificación teórica y la relación entre variables debe ser lineal (Rojo, 2007, p. 4).

El modelo de regresión múltiple es el siguiente:  $Y = f(X_1, X_2, \dots, X_n)$ .

Operativamente, el modelo multivariado se expresa como:

$$Y = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_n X_n + \varepsilon \text{ (Walpole et al., 2012, p. 390).}$$

#### Ejercicio 162:

Elaborar un modelo matemático para la función  $Y = f(X_1 + X_2)$  con los datos de un ensayo donde se probaron dosis de fertilizante nitrogenado ( $X_1$ ) sobre la producción (g.parcela<sup>-1</sup>) y en un momento crítico se constató la incidencia de una plaga insectil  $W(X_2)$ .

#### Datos y análisis:

En el Cuadro 197, se exponen los datos de la producción (Y) en función de las dosis de fertilizante nitrogenado que varían de 0 a 200 g.parcela<sup>-1</sup> ( $X_1$ ) y de las incidencias de una plaga insectil que varían de 0 a 50 % ( $X_2$ ). Los valores de Y se transformaron al Escalar Al máximo (EAM), en por ciento, procedimiento que facilita la interpretación de los resultados.

En el Cuadro 198 se indica la matriz de correlaciones lineales ( $r$ ) cuyos valores denotan una fuerte asociación estadística entre las variables en estudio. La correlación entre EAM e Y tiene un coeficiente  $r = 1$ , interpretándose que las dos variables tienen la misma información.

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 197. Dosis de abono, infestación de una plaga, producción y valores escalares al máximo*

Dosis de abono (g.parcela <sup>-1</sup> )	Infestación de la plaga insectil W (%)	Producción (g.parcela <sup>-1</sup> )	Escalar Al Máximo (%)
X1	X2	Y0	EAM
0	50	300	68
0	40	250	57
0	21	280	64
100	14	360	82
100	17	340	77
100	9	380	86
200	8	440	100
200	7	430	98
200	3	420	95
300	4	420	95
300	2	430	98
300	0	420	95

*Cuadro 198. Matriz de correlaciones lineales bivariadas*

	X1	X2	Y	EAM
X1	1			
X2	- 0,838 **	1		
Y	0,901 **	- 0,856 **	1,0	
EAM	0,901 **	- 0,856 **	1,0 **	1,0

Nota:  $r_{0,01} = 0,623$  con GL = 14. \*\* = Hay correlación altamente significativa

La regresión indica alta significación estadística en el análisis de la causalidad (Cuadro 199). La producción (Y) depende de la adición de abono (X<sub>1</sub>) y de la incidencia de la plaga insectil W (X<sub>2</sub>) con un 99 % de confianza (p < 0,01). El 84,6 % de la variación en la producción es explicada a través de las variables independientes, proporción que coincide con el Coeficiente de determinación R<sup>2</sup>.

***Cálculo del Coeficiente de determinación ajustado:***

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} [1 - R^2] \rightarrow \bar{R}^2 = 1 - \frac{12-1}{12-2-1} [1 - 0,846] \rightarrow \bar{R}^2 = 0,812$$

El valor del Coeficiente de determinación ajustado  $\bar{R}^2$  es alto (81,2 %), por tanto, el análisis de la regresión es confiable.

*Cuadro 199. Análisis de regresión de la producción en función de las dosis de fertilizante e incidencia de la plaga W*

Origen de la variación	GL	SC	CM	F calculada	F <sub>0,05</sub>	F <sub>0,01</sub>	p	Variación (%)
Regresión	2	2099	1049,7	24,8 **	4,26	8,02	0,00022	84,6
Residuos	9	381	42,4					15,4
Total	11	2481						100

Nota: \*\* = alta significación estadística (p < 0,01)

**Modelo multivariado:**

El modelo que resultó del análisis de regresión múltiple<sup>38</sup> fue:

$$PP = 77,44 + 0,080X_1 - 0,323 X_2$$

Dónde:

PP = Producción por parcela (gramos)

X<sub>1</sub> = Dosis de fertilizante (gramos.parcela<sup>-1</sup>)

X<sub>2</sub> = Incidencia porcentual de la plaga insectil W

**Decisión:**

El modelo multivariado calculado es:  $PP = 77,44 + 0,080X_1 - 0,323 X_2$ . La producción está en función de la dosis de fertilizante y de la incidencia de la plaga insectil W. La regresión explica el 84,6 % de la variación en la producción.

---

<sup>38</sup> El modelo  $Y = f(X_1, X_2)$  se obtiene en el programa Excel usando la función regresión





## CAPÍTULO VIII: INVESTIGACIÓN APLICATIVA

La investigación aplicada o aplicada es un nivel donde se busca resolver los problemas prácticos de un campo específico en base a la integración de los conocimientos teóricos y científicos sistematizados o recomendaciones resultantes de investigaciones previas. La investigación pura o básica se centra en la generación de conocimiento teórico sin una aplicación directa, la investigación aplicada tiene como objetivo abordar cuestiones y desafíos del mundo real.

La investigación aplicada usa el método científico y técnicas para obtener datos y evidencia empírica para comprender los fenómenos y resolver los problemas prácticos en áreas como la medicina, la ingeniería, la psicología, la educación, la agricultura, entre otros. El propósito de la investigación aplicada es lograr un impacto práctico en la mejora de procesos, en el desarrollo de nuevas tecnologías, en la optimización de productos o servicios, y proporcionar información para la formulación de políticas. La investigación aplicada es el proceso de verificación del desempeño biológico, de la significación estadística y del beneficio económico de las soluciones tecnológicas generadas a través de la investigación, en los niveles: descriptivos, relacionales, experimentales y predictivos.

En este nivel se validan tecnologías en varios ambientes, se maximizan los beneficios y minimizan los costos usando la programación lineal y se promueve el control de calidad, el análisis ambiental, el análisis de la sostenibilidad y la investigación participativa<sup>39</sup>.

### 8.1 VALIDACIÓN DE TECNOLOGÍAS

La validación de tecnologías es la verificación práctica, en condiciones próximas a la realidad socioeconómica, de las soluciones tecnológicas que proponen las instituciones de investigación para los “dominios de recomendación” concretos, previo a la entrega formal a la sociedad y al sector.

Una alternativa tecnológica para tener rápida adopción, por parte de los potenciales usuarios, debe cumplir tres condiciones: [1] comportamiento biológico resiliente, [2] decisión estadística confiable y [3] beneficio neto comparativamente superior. En esta etapa se realizan los ajustes a las recomendaciones, adaptadas a las circunstancias agroambientales de los territorios y socioeconómicas de las comunidades (Lores et al., 2008. p. 6).

En los ensayos de validación se comparan alternativas tecnológicas, propuestas nuevas o de mejora continua (TA), la tecnología de la “competencia” (TC) y la tecnología local del productor (TLP). Para poder recomendar con seguridad, la nueva tecnología (TA) debe ser sustancialmente superior en los ámbitos biológico, estadístico y económico a las prácticas vigentes, del productor (TLP) y de la competencia (TC). Cabe indicar que en este nivel no se contrasta con un testigo absoluto sino con la mejor práctica de los productores.

Los análisis estadísticos, este tipo de ensayos, comprenden tanto la estadística descriptiva como la inferencial, incluido el cálculo de las medias y de los intervalos de confianza; así como el análisis de varianza y la separación de medias mediante como las pruebas t de Student o la de Welch, DMS u otras técnicas. Un aspecto clave de la validación de tecnologías es el análisis económico de las alternativas tecnológicas propuestas, en los diferentes escenarios (ambientes y usuarios).

---

<sup>39</sup> La investigación participativa se apoya en técnicas de la andragogía que no son tratadas en este documento.

**Ejercicio 163:**

Analizar un ensayo de validación de tecnologías para la producción de maíz en varias localidades de la provincia de Manabí.

**Datos y análisis:**

La producción media en quintales de maíz al 12 % de humedad del grano, en un conjunto de 10 localidades de Manabí, probando tres alternativas tecnológicas que se expone en el Cuadro 200.

Cuadro 200. Comparación de tres alternativas tecnológicas para la producción de maíz en 10 localidades

Localidades	T1	T2	T3	Media por localidad (qq.ha <sup>-1</sup> )	Escala Al Máximo (%)
	Alternativa tecnológica	Tecnología de la competencia	Tecnología del productor (TLP)		
24 de Mayo	207	189	182	193	100
Chone	184	165	160	170	88
Junín	195	190	164	183	95
Paján	176	142	154	157	81
Pichincha	160	161	118	146	76
Portoviejo	190	180	122	164	85
Rocafuerte	185	151	168	168	87
Santa Ana	175	160	145	160	83
Sucre	171	168	165	168	87
Tosagua	178	170	135	161	83
Media por tecnologías (qq.ha <sup>-1</sup> )	182	168	151	167	87
Comparación con relación a TLP (%)	120	111	100		
Prueba t de Student	T1 > TLP p = 0,0003. Hay diferencia altamente significativa con 99 % de confianza (**)				
	T2 = TLP p = 0,0605. No hay diferencia estadística significativa (ns)				
	T1 > T2 p = 0,0043. Hay diferencia significativa con 95 % de confianza (*)				

Nota: La prueba de t corresponden a comparaciones apareadas cuando hay igual número de observaciones.

**Interpretación:**

La producción más alta corresponde a la alternativa tecnológica sujeta a la validación. Los valores EAM indican mayor producción en 24 de Mayo y Junín. La alternativa tecnológica (T1) tiene 20 % más de producción comparada con el testigo del productor (TLP) mientras que la tecnología de la competencia (T2) significó solo un 11 % adicional.

**Decisión:**

La alternativa tecnológica (T1), sujeta a validación, produjo 20 % más que las tecnologías de la competencia (T2). Hay que diseminar los resultados.

## 8.2 INTRODUCCIÓN A LA PROGRAMACIÓN LINEAL

La programación lineal (PL) es una técnica aplicada en la investigación de operaciones basada en métodos matemáticos de optimización que permite la representación de modelos lineales usada para reducir costos o maximizar los beneficios en los distintos campos empresariales (Doll & Orazem, 1978, p. 319-330).

La PL es un caso especial de programación matemática que se ocupa de la maximización o minimización de una función objetivo de tipo lineal en la que sus variables están sujetas a restricciones lineales de igualdad y desigualdad (Puente y Gavilánez, 2018, p. 8). La maximización o minimización de una función lineal se expresa como una ecuación de primer grado con potencia 1 (Doll & Orazem, 1978, p. 338).

La ecuación lineal es una igualdad matemática, cuya forma básica es  $Y = a + b(X)$  que se transforma en la función objetivo:  $Z = f(X,Y)$ . En este contexto, los elementos de la PL son la función objetivo y las restricciones. La función objetivo es la ecuación o inecuación algebraica a ser optimizada, procurando maximizar los beneficios o minimizar los costos. Por ejemplo:  $Z = 250 M + 350 N$ .

La función objetivo es una ecuación lineal que se propone maximizar o minimizar. Las restricciones también son ecuaciones o desigualdades lineales que representan las limitaciones del problema, como son los recursos limitados.

Las restricciones son las condiciones que deben cumplir los factores en estudio para ser optimizados (Puente y Gavilánez, 2018, p. 13). Por ejemplo, si una máquina solo se puede usar 16 horas al día. Los modelos de optimización deben ser manejables, resolubles y representativos de la situación original (Bermúdez, 2011, p. 85). La proposición teórica de la PL se sustenta en las premisas: continuidad, determinismo, proporcionalidad y aditividad.

**Continuidad.**- Indica que las variables deben ser continuas y pueden tomar cualquier valor real dentro del dominio permitido de números reales no negativos, aunque se fijen límites arbitrarios dentro de este conjunto. El modelo no aplica en situaciones que requieran el planteo de variables que puedan tomar valores negativos.

**Determinismo.**- No tiene consideraciones probabilísticas, enfocan su atención en las circunstancias críticas, donde las cantidades son determinadas y exactas.

**Proporcionalidad.**- La contribución de cualquier actividad a la función objetivo es directamente proporcional al nivel de dicha actividad. Si el nivel de una actividad se incrementa o disminuye, genera un cambio proporcional en la función objetivo.

**Aditividad.**- Hace referencia a la contribución a la función objetivo que está constituida por la suma de las contribuciones parciales. Esto implica que la contribución de cada actividad es independiente del nivel de las demás actividades.

El programa Excel tiene un asistente llamado SOLVER, que resuelve problemas de programación lineal (Bermúdez, 2011, p. 88). La activación del Solver en Excel, se resume en cinco pasos: (1) En el menú de Excel, ubicarse en datos y entrar en personalizar cinta de opciones, (2) Localizar complementos de Excel en el panel, (3) Marcar e ir a complementos de Excel, (4) Activar Solver (con un visto) y (5) Aceptar.

Las soluciones pueden basarse en el método matemático o en el método gráfico. En línea se encuentran disponibles varios programas para resolver problemas de PL<sup>40</sup>.

---

<sup>40</sup> El link de un programa en línea para resolver problemas PL es: <http://www.phpsimplex.com/>

**Ejercicio 164:**

Una empresa industrial puede producir dos tipos de productos, M y N, de alta demanda, en sus maquinarias A y B. ¿Cuál es la combinación óptima de productos M y N a ser producidas en las maquinarias A y B para obtener la máxima utilidad económica por día de trabajo?.

**Datos y análisis:**

La capacidad de producción y las restricciones de las maquinarias A y B en la producción de M y N, se exponen en el Cuadro 201.

Cuadro 201. Maquinarias A y B, productos M y N y restricciones de uso de las máquinas

Maquinarias	Producto M	Producto N	Restricción de uso de máquinas: máximo de horas de trabajo
A	1 hora/unidad	4 horas/unidad	16 horas/día
B	3 horas/unidad	2 horas/unidad	18 horas/día
Utilidad por unidad de producto	US\$ 250	US\$ 350	

**Método matemático**

El primer paso es determinar la función objetivo y restricciones como modelos lineales.

**Función objetivo:**

Z = Utilidad máxima por día de trabajo con las dos máquinas se indica en el modelo:

$$Z = 250 M + 350 N$$

**Restricciones:**

$$1 M + 4 N \leq 16 \quad \text{Para maquinaria A}$$

$$3 M + 2 N \leq 18 \quad \text{Para maquinaria B}$$

$$M > 0, N > 0$$

$$1 M + 4 N \leq 16: \quad \text{Para maquinaria A}$$

$$3 M + 2 N \leq 18: \quad \text{Para maquinaria B}$$

$$M \leq 16 - 4N$$

**Cálculos:**

$$3(16 - 4N) + 2N \leq 18 \rightarrow 48 - 12N + 2N \leq 18 \rightarrow -10N \leq -30 \rightarrow N = \frac{30}{10} = 3$$

Los valores de N = 3 y M = 4, se reemplazan en la función objetivo:

$$Z = 250 M + 350 N$$

$$Z = (250)(4) + 350 (3) = \$2050.\text{día}^{-1}$$

**Decisión:**

La máxima utilidad por día de trabajo es de \$2050, produciendo en 4 unidades de M.día<sup>-1</sup> en las máquinas A y 4 unidades de N.día<sup>-1</sup> en las máquinas B.

### Método gráfico

Para la solución gráfica del problema de PL se parte de los modelos lineales propuestos como función objetivo y restricciones y luego se traza sus funciones en el plano cartesiano, asumiendo que X e Y tomar los valores de las inecuaciones (Gráfico 60).

#### Análisis:

Para máquina A:

$$M + 4N \leq 16$$

Si  $M = 0$

$$0 + 4N = 16 \rightarrow N = 16/4 = 4$$

En el punto A(0, 4)

Si  $N = 0$

$$M + 4(0) = 16 \rightarrow M = 16$$

Punto B (16, 0)

Para máquina B:

$$3M + 2N \leq 18$$

Si  $M = 0$

$$3(0) + 2N = 18$$

$$2N = 18 \rightarrow N = 18/2 = 9$$

En el punto C (0, 9)

Si  $N = 0$

$$3M + 2(0) = 18 \rightarrow 3M = 18 \rightarrow M = 18/3 = 6$$

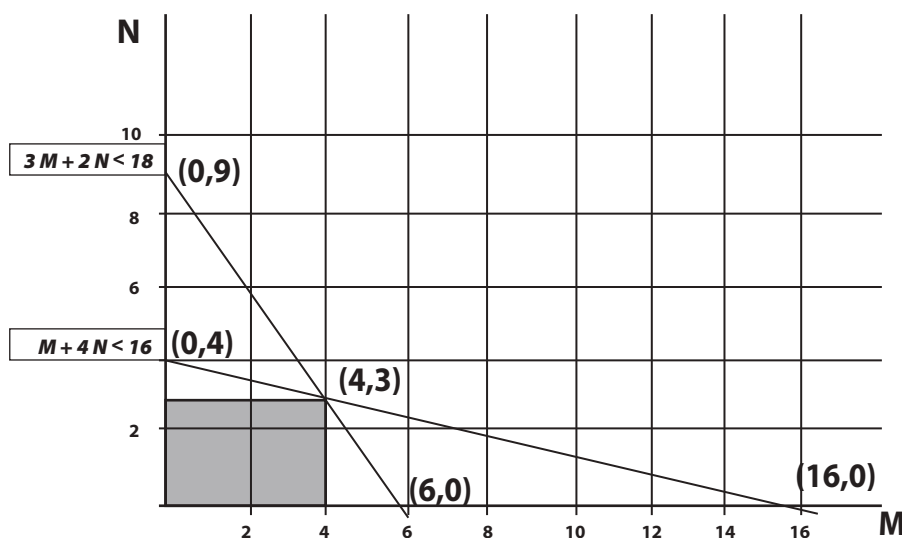


Gráfico 60. Maximización de beneficios usando máquinas A y B para producir M y N

#### Decisión:

Diariamente, en la maquinaria A se tiene que producir 4 unidades de M y en la maquinaria B, 3 unidades de N para lograr el máximo beneficio económico por día de trabajo, que alcanza el valor:  $Z = \$2050.\text{día}^{-1}$ .

### 8.3 INVESTIGACIÓN DEL AMBIENTE

En la investigación del ambiente se pueden usar las gráficas radiales, el cálculo de índices de sostenibilidad, el análisis de tendencias o de preferencias para describir, relacionar, comprobar y predecir situaciones en escenarios con intervención o sin la intervención de las personas (antropogénica).

Las mediciones de las variables pueden ser cuantitativas o cualitativas. Los datos cuantitativos se pueden transformar a cualitativas usando los cuartiles o quintiles, así como, los datos ordinales se pueden transformar a valores escalares al máximo (EAM) o a porcentajes con métodos como el de Townsend y Heuberger, para analizarlos como datos cuantitativos.

**Ejercicio 165:**

Comparar la situación de degradación de cinco elementos del ambiente, basado en la consulta con grupos focales de seis comunidades rurales.

**Datos y análisis:**

**Variables ambientales:**

Erosión del suelo agrícola, deforestación, contaminación del agua de consumo humano, riesgos de plaguicidas en los alimentos y pérdida de la biodiversidad.

**Escala de medición:**

Las variables se evalúan en consenso, usando una escala ordinal de 0 a 5: 0 = ausencia o condición ideal, 1 = bajo, 2 = medio bajo, 3 = medio, 4 = medio alto y 5 = alto. La variable cualitativa es del tipo “menor es mejor”.

Los datos agroambientales valorados en consenso con los actores locales, se exponen en el Cuadro 202.

*Cuadro 202. Mediciones del deterioro ambiental en cuatro comunidades rurales*

	Erosión	Deforestación	Contaminación del agua	Pérdida de la biodiversidad	Residuos de plaguicidas en alimentos
Comunidad A	4	4	4	5	4
Comunidad B	2	1	2	1	1
Comunidad C	3	2	4	2	5
Comunidad D	5	5	4	4	5

Nota: Las calificaciones en la escala de 0 a 5 tienen que ser consensuadas con grupos focales por comunidad.

**8.3.1 USO DE GRÁFICAS RADIALES EN ANÁLISIS AMBIENTAL**

Las gráficas radiales es una técnica de representación de datos ordinales siempre que las variables sea de un solo tipo, ya sea: “mayor es mejor” o “menor es mejor”. En ningún caso se deben mezclar las variables de distinto sentido, ya que imposibilitaría una correcta interpretación. Se usa las gráficas radiales para representar los perfiles sensoriales del café, del cacao y de otros alimentos, posibilita valorar la correspondencia de las calificaciones entre jueces sensoriales y medir el deterioro ambiental entre comunidades, así como los efectos durante un período de intervención (proyecto).

Los resultados de la situación de deterioro ambiental, indicados en el cuadro anterior, de cuatro comunidades rurales, en un momento concreto, se exponen en el Gráfico 61, donde se destaca que la deforestación está en niveles medio-altos y altos, en las comunidades A y D (urge formular un proyecto de reforestación). La contaminación del agua adquiere situaciones preocupantes en las comunidades A, B y C (se requiere un proyecto de capacitación). La erosión de los suelos agrícolas está en niveles preocupantes en las comunidades A y D (se necesita una agenda de capacitación).

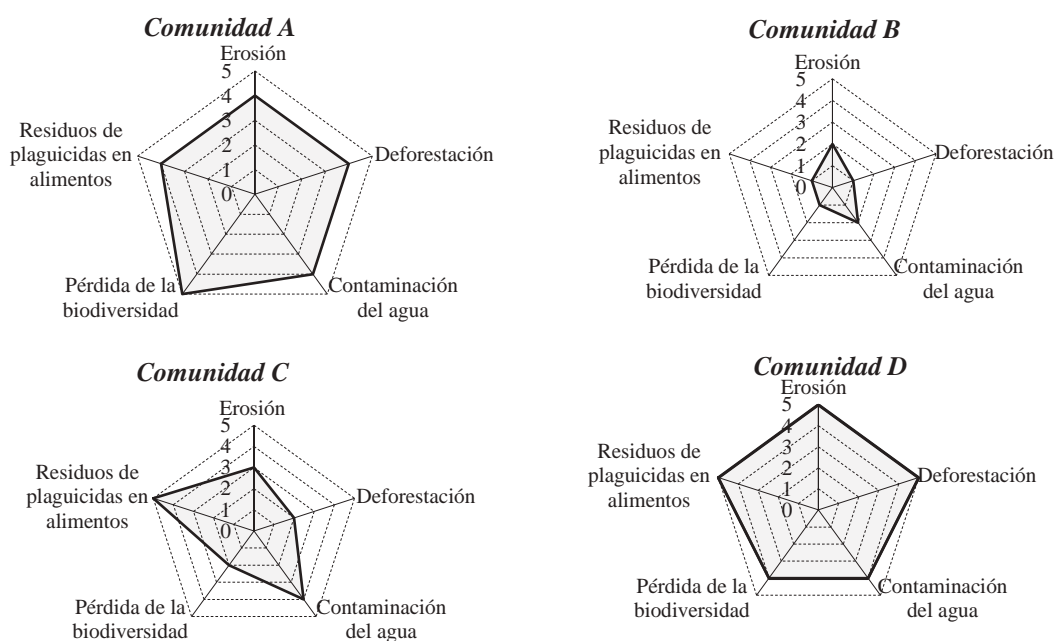


Gráfico 61. Análisis de variables ambientales en cuatro comunidades rurales en un momento 1

**Decisión:**

Se justifica la intervención en la reforestación en las comunidades A y D

**8.3.2 EVALUACIÓN DE IMPACTO AMBIENTAL**

Una evaluación del impacto ambiental (EIA) es un modelo de investigación aplicada que integra elementos de identificación, descripción, experimentación y predicción de los efectos de las actividades humanas, en los componentes como: agricultura, ganadería, silvicultura, manufactura, turismo y minería, englobando la biota y abiota, en la perspectiva de prevenir y mitigar sus efectos negativos.

Un efecto ambiental se refiere a la modificación de los factores bióticos y abióticos, sean: físico-químico (agua, suelo y aire), biótico (plantas y animales) o paisaje. El impacto ambiental se refiere al cambio en un parámetro ambiental en un período específico y en un área definida, como resultado de una actividad humana particular, comparado con la situación que habría resultado sin la acción (Perevochtchikova, 2013, p. 284).

La interpretación de los efectos e impactos se tiene que hacer en términos de conservación, salud de los ecosistemas y bienestar de las poblaciones. Un impacto ambiental genera cambios favorables o adversos sobre los ecosistemas, el clima y la sociedad a causa de actividades, como: extracción excesiva de recursos naturales, disposición inadecuada de residuos, emisión de contaminantes y uso inapropiado de plaguicidas, excesivo laboreo del suelo, entre otros. Los impactos directos e indirectos se pueden valorar en tres dimensiones: magnitud, importancia y significancia (Perevochtchikova, 2013, p. 287).

Un ejemplo de impacto del uso de prácticas agroecológicas en una finca cacaotera (con intervención) sobre la cantidad de lombrices de tierra, medido en escala ordinal de 1 a 10 (1 = ausente o muy bajo; 10 = muy alto) y un lote testigo con prácticas convencionales



en el uso de plaguicidas (Figura 35). En los dos escenarios se inicia en una magnitud muy alta (10) la misma que varía en los monitoreos anuales. Luego de 5 años de ejecución del proyecto agroecológico, la cantidad de lombrices de tierra se recupera al nivel 10 de la escala (muy alto) y en el testigo convencional con agroquímicos se reduce al nivel de 6 de la escala (alrededor de una cantidad mediana de lombrices)

El impacto ambiental puede valorarse con modelos de regresión lineal o no lineal, basado en muestreos aleatorios. Además, se pueden hacer comparaciones entre muestras de las áreas intervenidas con prácticas agroecológicas y las áreas convencionales usando pruebas estadísticas como la *t de Student* o la *t de Welch*.

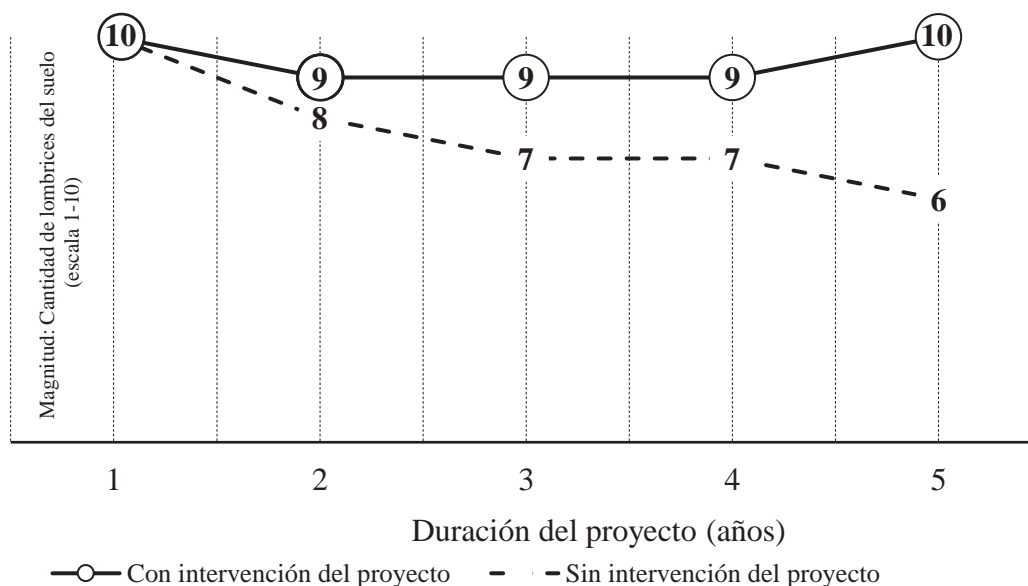


Figura 35. Impacto de un proyecto agroecológico sobre la sobrevivencia de lombrices en el suelo

La estadística tiene un rol fundamental la identificación, descripción y predicción de impactos ambientales (Perevochtchikova, 2008, p. 293; Dellavedova, 2016, p. 14). Se pueden comparar las situaciones del “antes” y del “después”. Los efectos de las actividades antrópicas se pueden valorar a través de parámetros como:

- Signo : favorable (+), desfavorables (-)
- Acumulación : acumulativo o no acumulativo (simple)
- Sinergia : reforzamiento de efectos simples o efectos contraproducentes
- Momento : de corto plazo o de largo plazo
- Persistencia : temporal o permanente
- Reversibilidad : reversible o no reversible
- Posibilidad : recuperable o no recuperable
- Periodicidad : recurrente o no recurrente
- Continuidad : continuo o discontinuo
- Localización : local, regional o global

La interpretación de los efectos debe hacerse en términos de conservación de los recursos, salud de los ecosistemas y bienestar de las comunidades. La valoración de la magnitud e importancia se realiza usando una escala tipo Likert (ejemplo: de 1 = muy baja a 10 = muy alta). La magnitud puede desagregarse en intensidad y afectación; y la importancia en duración e influencia, con mediciones en escala ordinal (Cuadro 203).

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

*Cuadro 203. Magnitud e importancia como variables asociadas a la medición del impacto ambiental*

Calificación	Magnitud		Calificación	Importancia	
	Intensidad	Afectación		Duración	Influencia
1	Baja	Baja	1	Temporal	Puntual
2	Baja	Media	2	Media	Puntual
3	Baja	Alta	3	Permanente	Puntual
4	Media	Baja	4	Temporal	Local
5	Media	Media	5	Media	Local
6	Media	Alta	6	Permanente	Local
7	Alta	Baja	7	Temporal	Regional
8	Alta	Media	8	Media	Regional
9	Alta	Alta	9	Permanente	Regional
10	Muy alta	Alta	10	Permanente	Nacional

**Ejercicio 166:**

En el proyecto “Propuesta de ordenamiento y uso turístico de la zona de playa de San Clemente” se miden las acciones antrópicas usando 17 criterios con escala ordinal: N = neutro, B = bajo impacto, M = impacto medio y A = alto impacto (Andrade et al., 2013).

**Datos y análisis:**

Los datos de campo se exponen en el Cuadro 204, relacionando actividades, acciones y factores ambientales con la aplicación de 17 criterios.

*Cuadro 204. Criterios ambientales para el ordenamiento territorial de la zona de playa de San Clemente*

	Aire	Agua	Suelo	Flora	Fauna	Social
[ST] turismo [AL] alojamiento [MV] movilización [IN] infraestructura	Calidad de aire (1) Niveles de ruido y vibraciones (2) Calidad agua superficial y subterránea (3) Calidad de agua de mar (4) Erosión (5) Afectación de hábitats (6) Calidad de suelo por presencia de desechos (7) Flora terrestre (8) Flora acuática (9) Fauna terrestre (10) Fauna acuática (11) Número de visitantes (12) Calidad visual y paisaje (13) Generación de empleo (14) Infraestructura de servicio público (15) Satisfacción al cliente (16) Calidad de vida de población (17)					
[ST] Consumo de productos locales	N N	N N	N N N	N N	N N	M N M N M M
[ST] Afluencia de pasajeros	N M	M M	B M M	N N	N N	A N A M N M
[ST] Intercambio cultural	N N	N N	N N N	N N	N N	M N N N M M
[AL] Demanda y consumo de agua potable	N N	M N	N N N	N N	N N	M N M M M A
[AL] Eliminación de aguas residuales	N N	N N	N N N	N N	N N	M M N N M M
turísticos	N N	N B	N N M	N N	N N	M M M M M M
[MV] Generación de desechos sólidos	M N	M N	N N N	N N	N N	M A B M M M
[MV] Evidencia de fogatas en sitios de visita	M N	N M	M M M	N N	N B	M M B N M M
[IN] Incremento de comercio	N N	N N	N N M	N N	N N	A N M M M M
[IN] Incremento de demanda de energía eléctrica	N M	N N	N N N	N N	N N	M N B M B M
[IN] Capacidad de carga turística	N M	N M	N N N	N N	N N	M M M N N M

Nota: Escala de impactos: N = neutro, B = bajo, M = medio y A = alto

**Resultados:**

En calidad del aire, las acciones no tienen impacto positivo ni negativo, es neutro en 82 %. El número de visitantes tienen impacto medio en 82 % y alto en 18 % (Cuadro 205).

*Cuadro 205, Impactos ambientales de las acciones antropogénicas en la “Propuesta de ordenamiento y uso turístico de la zona de playa de San Clemente”*

Criterios de evaluación ambiental	Calificación porcentual			
	Neutro	Bajo	Medio	Alto
Calidad de aire (1)	82	0	18	0
Niveles de ruido y vibraciones (2)	73	0	27	0
Calidad agua superficial y subterránea (3)	64	0	27	9
Calidad de agua de mar (4)	55	9	36	0
Erosión (5)	82	9	9	0
Afectación de hábitats (6)	82	0	18	0
Calidad de suelo por presencia de desechos (7)	45	0	45	9
Flora terrestre (8)	100	0	0	0
Flora acuática (9)	100	0	0	0
Fauna terrestre (10)	100	0	0	0
Fauna acuática (11)	91	9	0	0
Número de visitantes (12)	0	0	82	18
Calidad visual y paisaje (13)	55	0	36	9
Generación de empleo (14)	18	27	45	9
Infraestructura de servicio público (15)	45	0	55	0
Satisfacción al cliente (16)	18	9	73	0
Calidad de vida de población (17)	0	0	91	9

**Decisión:**

El mayor impacto negativo sobre el ambiente corresponde al número de visitantes.

## 8.4 ANÁLISIS DE LA SOSTENIBILIDAD DE LAS FINCAS

La sostenibilidad se refiere a la plena satisfacción de las necesidades actuales de la sociedad en su conjunto sin comprometer la capacidad de las generaciones futuras de satisfacer las suyas, garantizando un armonioso entre crecimiento económico, ambiente sano y bienestar social. Esta idea del desarrollo sostenible riñe con el concepto de plusvalía clásico que mide solo las ganancias económicas. Los retos a los que se enfrenta la población humana como el cambio climático o la escasez de agua solo se pueden abordar desde una perspectiva global, promoviendo el desarrollo sostenible (Sarandón y Flores, 2014, p.54).

La sostenibilidad ambiental se enfoca en preservar la biodiversidad sin tener que renunciar al progreso económico y socio-cultural. La sostenibilidad económica procura que todas las actividades productivas sean rentables. La sostenibilidad socio-cultural busca la cohesión de la sociedad en armonía con el ambiente y potenciando las capacidades de los pueblos de generar ciencia, arte y cultura (Anzules et al., 2015, p. 265). El principal objetivo es perpetuar al ser humano como especie, satisfaciendo sus necesidades presentes y futuras, mediante el uso responsable de los recursos naturales. En estos procesos de construcción de desarrollo sostenible, la estadística juega un rol fundamental en el diseño metodológico y la confianza en la toma de decisiones.

#### 8.4.1 DIMENSIONES, INDICADORES Y SUBINDICADORES

En las dimensiones económica (E), ambiental (A) y socio-cultural (S), se definen los indicadores y subindicadores que deben ser calificados para determinar los índices de sostenibilidad: IE, IA e IS (Sarandón y Flores, 2009, p. 22).

##### 8.4.1.1 DIMENSIÓN ECONÓMICA (E)

###### E1. Autosuficiencia alimentaria

- E11. Diversificación agro productiva de la finca
- E12. Huerto familiar con especies aromáticas, medicinales y alimentarias
- E13. Frutales diversos en policultivos

###### E2. Ingreso económico neto mensual de la familia

- E21. Ingresos por venta de productos agropecuarios
- E22. Ingresos por actividades artesanales o pequeño negocio
- E23. Prestación de servicios como asalariado agrícola
- E24. Estacionalidad de los ingresos

###### E3. Estabilidad económica de la familia

- E31. Ingresos permanentes por ventas de productos agropecuarios
- E32. Ingresos permanentes por ventas de productos de artesanías u otros agregados
- E33. Variación drástica de precios en insumos, herramientas y equipos agrícolas.

##### 8.4.1.2 DIMENSIÓN AMBIENTAL (A)

###### A1. Conservación del suelo

- A11. Prácticas de conservación del suelo en la finca
- A12. Sistemas agroforestales
- A13. Incorporación de materia orgánica

###### A2. Conservación del agua

- A21. Protección de las fuentes hídricas con cobertura vegetal
- A22. Prevención de la contaminación de las fuentes hídricas
- A23. Uso eficiente del agua de riego
- A24. Buen uso del agua de consumo humano

###### A3. Conservación de la biodiversidad

- A31. Protección de la fauna nativa
- A32. Protección de la flora nativa
- A33. Mantenimiento de franja biológica

##### 8.4.1.3 DIMENSIÓN SOCIO CULTURAL (S)

###### S1. Satisfacción de necesidades básicas

- S11. Acceso permanente a alimentos
- S12. Acceso permanente a la educación
- S13. Acceso permanente a la salud
- S14. Adecuadas condiciones de la vivienda

###### S2. Aceptabilidad del sistema de producción

- S21. Mercado permanente para el cultivo predominante
- S22. Uso racional de abonos químicos
- S23. Uso racional de pesticidas
- S24. Predisposición a la mejora continua de la finca

S3. Integración social a sistemas organizativos

S31. Participación activa en la organización de productores

S32. Aprobación de la gestión de los órganos directivos

8.4.2 MEDICIÓN DE LA SOSTENIBILIDAD

La medición de la sostenibilidad, para valorar cada subindicador, en las dimensiones económica, ambiental y socio-cultural, en este modelo, se realiza usando la escala ordinal de 0 a 4 (Sarandón y Flores, 2009, p. 24), aunque pueden usarse escalas ordinales de mayor magnitud, como la de 0 a 10. En todo caso, los criterios o subindicadores deben caracterizarse en una sola dirección tipo de variable, por ejemplo: “mayor es mejor”.

La escala ordinal se describe a continuación:

Escala	Descripción de la calificación ordinal
0	No hay aseguramiento de la sostenibilidad
1	Baja sostenibilidad
2	Regular sostenibilidad
3	Buena sostenibilidad
4	Alta sostenibilidad

8.4.3 INDICADORES DE LA SOSTENIBILIDAD

La importancia de los subindicadores que equivale al peso de los criterios/variables, se los categoriza, en dos niveles:

- 1) Menos importante
- 2) Más importante

Los pesos asignados para cada criterio denotan la importancia que da el investigador a las variables que se expresan en los criterios de evaluación. Esta ponderación de la importancia de los criterios debe definirse con la participación de los actores en territorio.

En el Cuadro 206, se indican los criterios y la importancia asignados por los actores locales.

Cuadro 206. Criterios de evaluación de las dimensiones económica, ambiental y socio cultural e importancia

Criterios de evaluación	Importancia (Peso del criterio)
<b>DIMENSIÓN ECONÓMICA (E)</b>	
E1. Autosuficiencia alimentaria	2
E2. Ingreso económico neto mensual/familia	2
E3. Estabilidad económica de la familia	2
<b>DIMENSIÓN AMBIENTAL (A)</b>	
A1. Conservación del suelo	1
A2. Conservación del agua	2
A3. Conservación de la biodiversidad	1
<b>DIMENSIÓN SOCIO CULTURAL</b>	
S1. Satisfacción de necesidades básicas	2
S2. Aceptabilidad del sistema de producción	1
S3. Integración social a sistemas organizativos	1

**Cálculo de los índices de sostenibilidad:**

Los índices de sostenibilidad parciales IE, IA e IS se calculan usando la fórmula general de la media ponderada. La propuesta de Sarandón y Flores (2009, p. 25), para la evaluación de la sostenibilidad de una finca o de una comunidad rural conformada por un conjunto de fincas, se fundamenta en este modelo.

$$\bar{X}_p = \frac{X_1p_1 + X_2p_2 + \dots + X_np_n}{p_1 + p_2 + \dots + p_n}$$

Las herramientas estadísticas pueden aplicarse en distintos momentos, a lo largo de un período, sea corto, mediano o largo, en la perspectiva de conocer la evolución de los índices de sostenibilidad económica (IE), ambiental (IA) y sociocultural (IS), así como, el índice de sostenibilidad general (IG). La interpretación de los índices puede basarse en la escala ordinal (en este caso el valor más alto es 4).

De manera práctica, los índices IE, IA e IS pueden convertirse a porcentaje para facilitar su interpretación. El índice de sostenibilidad general (IG) corresponde al promedio de los índices IE, IA e IS valorados con la escala ordinal de 1 a 4. Los datos ordinales se transforman a porcentajes usando la fórmula de Townsend & Heuberger.

El modelo general para calcular el índice de sostenibilidad se basa en la media aritmética.:

$$IG = \frac{IE + IA + IS}{3} \quad [133]$$

**Regla de decisión:**

Si  $IG \geq 70\%$  hay sostenibilidad segura

Si  $IG$  de 50 a 70 % hay riesgos en la sostenibilidad, urge tomar medidas

Si  $IG < 50\%$  no hay sostenibilidad, urge intervención

**Índice de sostenibilidad económica (IE)**

Las calificaciones de los subindicadores o criterios económicos se indican en el Cuadro 207.

Cuadro 207. Grado de importancia y calificaciones de las variables económicas

Subindicadores o variables económicas	Importancia (1 - 2)	Calificación (Escala 0 - 4)
E11. Diversificación agro productiva de la finca	2	2
E12. Huerto con especies aromáticas, medicinales y alimentarias	2	2
E13. Frutales diversos en policultivos	2	3
E21. Ingresos agropecuarios	2	2
E22. Ingresos por actividades artesanales o pequeño negocio	2	1
E23. Demanda de mano de obra agrícola o prestación de servicios	2	3
E24. Estacionalidad de los ingresos	2	3
E31. Ingresos por ventas de productos agropecuarios	2	3
E32. Ingresos por ventas de artesanías u otros agregados	2	3
E33. Precios de insumos, herramientas y equipos agrícolas	2	0

**Cálculos:**

$$IE = \frac{\sum(E_n * p)}{\sum p}$$

$$IE = \frac{(2 * 2) + (2 * 2) + (3 * 2) + (2 * 2) + (1 * 2) + (3 * 2) + (3 * 2) + (3 * 2) + (3 * 2) + (0 * 2)}{2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2}$$

**HERRAMIENTAS ESTADÍSTICAS PARA LA INVESTIGACIÓN AGROPECUARIA**

$$IE = \frac{4 + 4 + 6 + 4 + 2 + 6 + 6 + 6 + 6 + 0}{20}$$

$$IE = \frac{44}{20} = 2,2 \rightarrow IE(\%) = \frac{2,2}{4} (100) = 55 \%$$

**Índice de sostenibilidad ambiental (IA)**

Las calificaciones de los subindicadores ambientales se indican en el Cuadro 208.

*Cuadro 208. Grado de importancia y calificaciones de las variables ambientales*

Subindicadores o variables ambientales	Importancia (1 - 2)	Calificación (Escala 0 - 4)
A11. Prácticas de conservación del suelo	1	3
A12. Sistemas agroforestales	1	3
A13. Remoción del suelo para cultivos	1	0
A21. Protección de las fuentes hídricas con cobertura vegetal	2	3
A22. Prevención de contaminación de las fuentes hídricas por pesticidas	2	2
A23. Uso eficiente del agua de riego	2	1
A24. Buen uso del agua de consumo humano	2	2
A31. Protección de la fauna nativa	1	3
A32. Protección de la flora nativa	1	3
A33. Mantenimiento de franja biológica o bosque de regeneración	1	3

**Cálculos:**

$$IA = \frac{\sum(A_n * p)}{\sum p}$$

$$IA = \frac{(3 * 1) + (3 * 1) + (0 * 1) + (3 * 2) + (2 * 2) + (1 * 2) + (2 * 2) + (3 * 1) + (3 * 1) + (3 * 1)}{1 + 1 + 1 + 2 + 2 + 2 + 2 + 1 + 1 + 1}$$

$$IA = \frac{3 + 3 + 0 + 6 + 4 + 2 + 4 + 3 + 3 + 3}{14}$$

$$IA = \frac{31}{14} = 2,21 \rightarrow IA(\%) = \frac{2,21}{4} = 55,3 \%$$

**Índice de sostenibilidad sociocultural (IS)**

Las calificaciones de los subindicadores socio culturales se indican en el Cuadro 209.

*Cuadro 209. Grado de importancia y calificaciones de las variables socio culturales*

Subindicadores o variables socio-culturales	Importancia (1 - 2)	Calificación (Escala 0 - 4)
S11. Disponibilidad de alimento	2	3
S12. Acceso a la educación	2	2
S13. Acceso a la salud	2	2
S14. Condiciones de la vivienda	2	2
S21. Sistema café como cultivo predominante	1	3
S22. Uso intensivo de abonos químicos	1	2
S23. Uso intensivo de pesticidas	1	3
S24. Predisposición a la mejora continua de la finca	1	4
S31. Participación activa en la organización de productores	1	3
S32. Aprobación de la gestión de los órganos directivos	1	2

**Cálculos:**

$$IS = \frac{\sum(S_n * p)}{\sum p}$$

$$IS = \frac{(3 * 2) + (2 * 2) + (2 * 2) + (2 * 2) + (3 * 1) + (2 * 1) + (3 * 1) + (4 * 1) + (3 * 1) + (2 * 1)}{2 + 2 + 2 + 2 + 1 + 1 + 1 + 1 + 1 + 1}$$

$$IS = \frac{6 + 4 + 4 + 4 + 3 + 2 + 3 + 4 + 3 + 2}{14}$$

$$IS = \frac{35}{14} = 2,5 \quad \rightarrow \quad IS (\%) = \frac{2,5}{4} = 62,5 \%$$

**Índice de sostenibilidad general (IG)**

Los índices de sostenibilidad por dimensiones se resumen:

Índice de sostenibilidad económica, E = 55 %

Índice de sostenibilidad ambiental, IA = 55,3 %

Índice de sostenibilidad socio cultural, IS = 62,5 %.

$$IG = \frac{IE+IA+IS}{3} \quad \rightarrow \quad IG (\%) = \frac{55 + 55,3 + 62,5}{3} = 57,6 \%$$

La sostenibilidad (factores ambiental, económico y socio cultural) puede representarse en forma matemática o gráfica, destacando las el balance o desbalance, situación que facilitará identificar intervenciones para mejorar los índices IE, IA e IS.

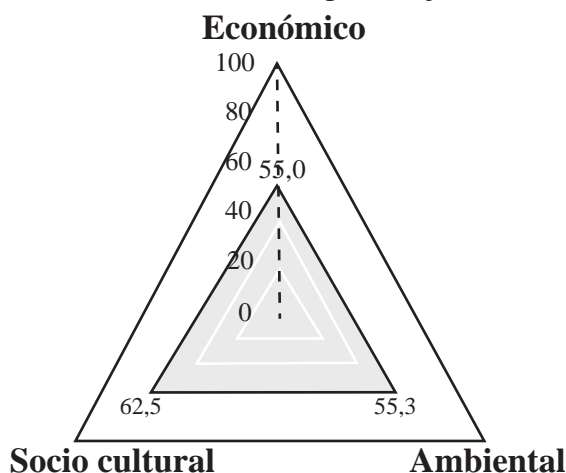


Figura 36. Triángulo de la sostenibilidad a partir de mediciones en fincas

**Decisión:**

El índice general de sostenibilidad IG = 57,6 % es un indicativo de que hay riesgos. Resulta imperativo tomar medidas para mejorar el índice general de sostenibilidad.





## REFERENCIAS BIBLIOGRÁFICAS

- Agrocalidad [Agencia de Regulación y control fito y zoonosanitario, EC]. (2020). *Manual técnico para el registro y control de fertilizantes, enmiendas de suelo y productos afines de uso agrícola*. 5ta. ed. Ministerio de Agricultura y Ganadería. 85 p.
- Aguilar-Barojas, S. (2005). Fórmulas para el cálculo de la muestra en investigaciones de salud. *Salud en Tabasco*, 11(1-2), 333-338.
- Alder, H. L. & Roessler, E. B. (1976). *Introduction to probability and Statistics*. Sixth edition. W. H. Freeman and Company.
- Andrade, P., Carlier, M., Guevara, S. y Osorio, V. (2013). *Propuesta de ordenamiento y uso turístico de la zona de playa de San Clemente*. [Artículos de tesis de grado. Licenciatura en turismo]. Escuela Superior Politécnica del Litoral.
- Andujar, A., Barranco, P., Belda, J. E., Cabello, T. y Carreño, R. (1997). Análisis de eficacia de productos fitosanitarios. *Phytoma, la revista profesional de protección vegetal*, (92), 32-40.
- Angulo, N. y Corujo, M. (2008). Tipo de investigación: Exploratoria-descriptiva. Universidad de Carabobo. *Revista Educación en Valores*, 1(9), 9-18.
- Anzules, V., Soria, N. y Basantes, E. (2015). Evaluación correlacional de la asociación cacao-leguminosas en la sostenibilidad del cacao (*Theobroma cacao* L.), en Santo Domingo de los Tsáchilas. *Ciencia*, 17(2), 263-273.
- Arroyo, P., Carrete, L. y García, S. I. (2008). Construcción de un índice de satisfacción para clientes de supermercados mexiquenses: Una investigación exploratoria. *Contaduría y administración*, (225), 59-78.
- Badii, M. H. y Guillen, A. (2009). Decisiones estadísticas: bases teóricas. *Daena: International Journal of Good Conscience*, 5(1), 185-207. ISSN 1870-557X.
- Badii, M. H., Castillo, J. y Wong, A. (2008). Uso de Análisis de Covarianza (ANCOVA) en investigación científica. *InnOvaciOnes de NegOciOs*, 5(1), 25-38. <http://eprints.uanl.mx/12489/1/A3.pdf>
- Badii, M. H., Guillen, A. y Abreu, J. L. (2021). Estadísticas para estimación del riesgo. *Daena: International Journal of Good Conscience*, 16(1), 1-21.
- Badii, M. H., Guillen, A., Cerna, E. y Valenzuela, J. (2011). Nociones Introductorias de Muestreo Estadístico. *Daena: International Journal of Good Conscience*, 6(1), 89-105.
- Badii, M. H., Guillen, L. A., Araiza, E., Cerna, J. y Landeros, J. (2012). Métodos No Paramétricos de Uso Común. *Daena: International Journal of Good Conscience*, 7(1), 132-155.
- Badii, M. H., López, F, Quiróz, H. y Pazhakh, A.R. (2004). Muestreo como un requisito fundamental en las ciencias. *InnOvaciOnes de NegOciOs* 1(1), 33-53.
- Báez, Y., Limón, J., Tlapa, D. y Rodríguez, M. (2010). Aplicación de Seis Sigma y los Métodos Taguchi para el incremento de la resistencia a la prueba de jalón de un diodo emisor de Luz. *Información Tecnológica*, 21 (1), 62-76.
- Balzarini, M., Bruno, C. Córdoba, M y Teich, I. (2015). *Herramientas en el análisis multivariado*. Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba.
- Bermúdez, Y. (2011). Aplicaciones de programación lineal, entera y mixta. Ingeniería Industrial. *Actualidad y Nuevas Tendencias*, II(7), 85-104.
- Bohórquez, B. P. y Ramírez, N. (2020). *Desarrollo y caracterización de un material biodegradable a partir de desechos cerveceros y micelio del hongo *Pleurotus ostreatus**. [Tesis de Ingeniería Química, Universidad de Los Andes]. <http://hdl.handle.net/1992/45043>

- Bombelli, E., Moschini, R., Wright, E. R., López, M.V., Fabrizio, M. C., Barberis, G. y Rivera, M. C. (2013). Modelado para la Predicción de Enfermedades en Cultivos de Alto Valor Comercial. *Proyecciones*, 11, 47-59.
- Bouza-Herrera, C. (2018). *Modelos de regresión y sus aplicaciones*. Serie reportes técnicos: modelos cuantitativos para aplicaciones en estudios de salud humana y el medio ambiente. Facultad de Matemática y Computación, Universidad de La Habana.
- Briceño-Yen, H., Álvarez-Benaute, L.M. y Valverde-Rodríguez, A. (2021). *Formulación de proyectos de investigación en Ciencias Agrarias*. Universidad Nacional “Hermilio Valdizan”, Huánuco, Perú. <https://www.unheval.edu.pe>
- Cardona, D., Rivera, M., González, J. y Cárdenas, E. (2014). Estimación y predicción con el modelo de regresión cúbica aplicado a un problema de salud. *Ingeniería solidaria*, 10(17), 153-160.
- Cifuentes, J. E. (2016). El método científico y la nueva filosofía de la ciencia: aportes y perspectivas. *Rastros Rostros*, 18(33), 61-70.
- Condo, L. A. y Pazmiño, J. M. (2015). *Diseño experimental en el desarrollo del conocimiento científico de las ciencias agropecuarias*. Tomo 1. Escuela Superior Politécnica de Chimborazo.
- Corchuelo, B. y Quiroga, A. (2014). *Análisis microeconómico II*. Manuales UEX. Universidad de Extremadura. 109 p.
- Corral, L. (2019). *Estadística y técnicas experimentales para la investigación biológica*. Universidad Politécnica Salesiana. Abya Yala.
- Corral, Y. (2009). *Validez y confiabilidad de los instrumentos de investigación para la recolección de datos*. Facultad de Ciencias Económicas y Sociales. Universidad de Carabobo.
- Dagnino, J. (2014). Regresión lineal. *Revista chilena de anestesia*, 43(2), 143-149. <https://revistachilenadeanestesia.cl/PII/revchilanestv43n02.14.pdf>
- De Boef, W. S. y Thijssen, M. H. (2007). *Herramientas de trabajo participativo con cultivos, variedades y semillas*. Wageningen International.
- De la Fuente, S. (2011). *Análisis de correspondencias simples y múltiples*. Universidad Autónoma de Madrid: Facultad de Ciencias Económicas y Empresariales.
- Dellavedova, M. G. (2016). *Guía metodológica para la elaboración de una evaluación de impacto ambiental*. Facultad de arquitectura y urbanismo. Universidad Nacional de La Plata.
- Di Rienzo, J. A., Casanoves, F., González, L. A., Tablada, E. M., Díaz, M., Robledo, C. W. y Balzarini, M. G. (2009). *Estadística para las ciencias agropecuarias*. 7ma. ed. Brujas.
- Doll, J. & Orazem, F. (1978). *Production economics: Theory with applications*. 2nd ed. John Wiley y Sons. Singapore. 470 p.
- Duicela, L. A. (2021). *Productividad y estabilidad ambiental de clones de café robusta en distintas localidades cafetaleras del Ecuador*. [Tesis doctoral, Universidad del Zulia]. DOI:10.13140/RG.2.2.24412.51845
- Duicela, L. A., Andrade, J., Farfán, D. S. y Velázquez, S. (2018). Calidad organoléptica, métodos de beneficio y cultivares de café robusta (*Coffea canephora* Pierre ex Froehner) en la Amazonía del Ecuador. *Revista Iberoamericana de Tecnología Poscosecha*, 19(2).
- Duicela, L. A., Velázquez, S. y Farfán, D. S. (2017). Calidad organoléptica de cafés arábigos en relación a las variedades y altitudes de las zonas de cultivo, Ecuador. *Revista Iberoamericana de Tecnología Poscosecha*, 18(1), 67-77.
- Espinoza, L. A. y Ochoa, J. M. (2020). El nivel de investigación relacional en las ciencias sociales. *Acta Jurídica Peruana*, 3(2), 93-111.

- García, A. (2011). *Elementos de Bioestadística*. 3ra. ed. Universidad de Extremadura.
- Geilfus, F. (1997). *80 herramientas para el desarrollo participativo: Diagnóstico, planificación, monitoreo y evaluación*. Prochamate-IICA. El Salvador.
- Gil, S. y Rodríguez, E. (2000). *Introducción al método experimental y teoría de errores*. Universidad Nacional San Martín.
- Gómez, K. A. & Gómez, A. A. (1983). *Statistical procedures for agricultural research*. Second edition. John Wiley & Sons.
- Gordón, R. y Camargo, I. (2015). Selección de estadísticos para la estimación de la precisión experimental en ensayos de maíz. *Agronomía Mesoamericana*, 26(1), 56-63.
- Gulland, J. A. (1966). *Manual de Métodos de Muestreo y Estadísticos para la Biología Pesquera*. FAO. <http://www.fao.org/docrep/x5684s/x5684s00.htm#Contents>
- Hernández, A.B., De la Paz, M. y García, L.A. (2015). La metodología de Taguchi en el control estadístico de la calidad. *Revista de la Escuela de Perfeccionamiento en Investigación Operativa*, 23(37):65-83.
- Hernández, R., Fernández, C. y Baptista, M. (2010). *Metodología de la investigación*. 5ta. ed. Mc Graw-Hill-Interamericana editores.
- Izaguirre, R. C. (2014). Enfoque filosófico dialéctico-materialista de la investigación científica. *Humanidades Médicas*, 14(1), 127-144. [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1727-81202014000100009&lng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1727-81202014000100009&lng=es).
- Jarque, C. y Bera, A. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review*, 55 (2), 163-172. <http://dx.doi.org/10.2307/1403192>
- Jiménez, E. (2017). *Obtención de pulpa de celulosa a partir de residuos de agavaceas: potencial elaboración de papel tipo artesanal*. [Tesis doctoral en ciencias ambientales, Universidad Autónoma del Estado de Hidalgo]. <https://repository.uaeh.edu.mx/bitstream/123456789/18450>
- Lagos, T. C. y Criollo, H. (2001). Eficiencia relativa del diseño látice con respecto a los diseños bloques al azar e irrestrictamente al azar en un ensayo de rendimiento de maíz de clima medio. *Revista de ciencias agrícolas*. 141-149. <https://revistas.udenar.edu.co/index.php/rfacia/article/view/1005>
- Lind, D. A., Marchal, W. G. y Wathen, S. A. (2012). *Estadística aplicada a los negocios y la economía*. 15ª ed. McGraw-Hill.
- Little, T. M. y Hills, F. J. (1987). *Métodos estadísticos para la investigación en agricultura*. 1ra. ed. Trillas.
- Lores, A., Leyva, A. y Varela, M. (2008). Los Dominios de Recomendaciones: Establecimiento e importancia para el análisis científico de los agroecosistemas. *Cultivos Tropicales*, 29(3), 5-10.
- Lucero, M. (2021). *Técnicas de fermentación del cacao CCN-51 y Nacional usando la enzima pectintranseliminasa como precursor sensorial de la almendra y licor de cacao*. [Tesis de magister en agroindustria, Escuela Superior Politécnica Agropecuaria de Manabí]. <https://repositorio.espam.edu.ec/xmlui/handle/42000/1365>
- Manterola, C., Grande, L., Otzen, T., García, N., Salazar, P. y Quiroz, G. (2018). Confiabilidad, precisión o reproducibilidad de las mediciones. Métodos de valoración, utilidad y aplicaciones en la práctica clínica. *Revista chilena de infectología*, 35(6), 680-688. <https://dx.doi.org/10.4067/S0716-10182018000600680>

- Martínez de Lejarza, I. y Martínez de Lejarza, J. (2010). *Inferencia: Contrastes no paramétricos*. Curso. Universidad de Valencia. <http://ocw.uv.es/ciencias-sociales-y-juridicas/inferencia/1mats06.pdf>
- Martínez, M. A., Almudena, E., Sánchez, A., Toledo, E. y Faulín, J. (2020). *Bioestadística amigable*. 4ta. ed. Universidad de Navarra.
- Martínez, M. y Sáenz, C. (2003). *Principios de genética mendeliana*. 2da. ed. Universidad Michoacana de San Nicolás de Hidalgo.
- Matas, A. (2018). Diseño del formato de escalas tipo Likert: un estado de la cuestión. *Revista Electrónica de Investigación Educativa*, 20(1), 38-47. <https://doi.org/10.24320/redie.2018.20.1.1347>
- Melo, O., López, L. y Melo, S. (2020). *Diseño de experimentos: Métodos y aplicaciones*. 2da. ed. Universidad Nacional de Colombia. 675 p.
- Mendel, G. (1866). *Experiments in plant hybridization*. Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865, Abhandlungen, 3-47. <http://www.esp.org/foundations/genetics/classical/gm-65.pdf>
- Millet, C. (19 de enero de 2022). *Protocolo de puntuación SCA*. Mare Terra Coffee. <https://mareterracoffee.com/es/blog/protocolo-de-puntuacion-sca/>
- Molina, J. G. y Rodrigo, M. F. (2010). *Estadísticos de forma de la distribución*. Curso 2009-2010. Universidad de Valencia.
- Mondragón, M. A. (2014). Uso de la correlación de Spearman en un estudio de intervención en fisioterapia. *Revista Movimiento Científico*, 8(1), 98-104. <https://dialnet.unirioja.es/servlet/articulo?codigo=5156978>
- Monsalve, S. y Zumaeta, Y. (2011). *Comparación entre la inseminación artificial profunda y tradicional en novillas Brahman blanco en la finca León, en Icononzo (Tolima)*. [Tesis en Medicina veterinaria]. Facultad de Ciencias Agropecuarias, Universidad de La Salle.
- Montes, F. (2007). *Introducción a la probabilidad*. Universidad de Valencia. <https://www.uv.es/montes/probabilitat/manual.pdf>
- Moschini, R., Martínez, M.I. y Sepulcri, M.G. (2013). Sistemas de pronóstico de enfermedades. Capítulo XXI, 2da. ed. En G.M. Murphy y R. H. Hurtado (Ed.) (pp. 409-441). *En Agrometeorología*. Universidad Nacional de la Plata.
- Muñoz, G. y Fernández, J. (1993). *Descriptores varietales: arroz, fréjol, maíz y sorgo*. FAO. <https://agris.fao.org/agris-search/search.do?recordID=DO2001100670>
- Muñoz, M. T. (2010). Uso de plaguicidas y discapacidad intelectual en estudiantes de escuelas municipales, Provincia de Talca, Chile. *Revista Facultad Nacional de Salud Pública*, 28(1), 29-35.
- Naranjo, F., Ríos, A.J., Pantoja, Y.V. y Tapia, M. (2020). Diseños ortogonales de Taguchi fraccionados. *Ingeniería, investigación y tecnología*, 21(2), 1-12. <https://doi.org/10.22201/ii.25940732e.2020.21n2.011>
- Navarro, J. R. y Vargas, J. C. (2015). Eficiencia relativa del diseño de bloques completos al azar para ensayos de arroz en Bagaces, Guanacaste, Costa Rica. *InterSedes*, XVI (34), 1-9. <https://www.scielo.sa.cr/pdf/is/v16n34/2215-2458-is-16-34-00061.pdf>
- Olivares, B. y Franco, E. (2015). Diagnóstico agrosocial de la comunidad indígena de Kashaama: Un estudio empírico en el estado de Anzoátegui, Venezuela. *Revista Científica Guillermo de Ockham*, 13(1), 87-95. <https://doi.org/10.21500/22563202.1691>



- Oliver, A., Galiana, L. y Bustos, V. (2018). Validación de la Escala de Satisfacción con la Vida y su relación con las dimensiones del Autoconcepto en universitarios peruanos. *Persona*, 21(2), 29-44. <http://dx.doi.org/10.26439/persona2018.n021.3018>
- Organización de las Naciones Unidas [ONU]. (2015). *Objetivos de Desarrollo Sostenible*. <https://www.un.org/sustainabledevelopment/es/development-agenda/>
- Perevochtchikova, M. (2013). La evaluación del impacto ambiental y la importancia de los indicadores ambientales. *Gestión y política pública*, 22(2), 283-312.
- Pérez, R. (2015). Sistema Internacional de Unidades (SI). *Rev. Obstet. Ginecol. Venez.*, 75(1), 49-74. <https://ve.scielo.org/pdf/og/v75n1/art07.pdf>
- Pistorale, S. M., Abbott, L. A. y Andrés, A. (2008). Diversidad genética y heredabilidad en sentido amplio en agropiro alargado, *Thinopyrum ponticum*. *Ciencia e investigación agraria*, 35(3), 259-264.
- Ponce, J. M. (2018). El método dialéctico en la formación científica de los estudiantes de pedagogía. *Actualidades Investigativas en Educación*, 18 (3), 147-167. <https://dx.doi.org/10.15517/aie.v18i3.33214>
- Puente, M. y Gavilánez, O. D. (2018). *Programación lineal para la toma de decisiones*. Escuela Superior Politécnica de Chimborazo.
- Quesada, J. y Figuerola, J. (2010). Potencia de una prueba estadística: aplicación e interpretación en ecología del comportamiento. *Etología*, 22, 19-37. <http://www.ebd.csic.es/jordiplataforma/subidas/Etologia2010.pdf>
- Quezada-Lucio, N. (2021). *Metodología de la investigación*. Macro EIRL.
- Quinaluisa, C., Navarrete, G., Franco, D., González, Bustamante, A., Sotomayor, I. y Medhi, S. (2021). Selection of self-compatible progenies of high productivity cocoa in the Ecuadorian coast. *Ciencia y Tecnología*, 14(1), 11-21. <https://revistas.uteq.edu.ec/index.php/cyt/article/view/448>
- Reyes, P. 1980. *Diseño de experimentos aplicados*. 2da. ed. Trillas. México. 344 p.
- Robles, R. (1986). *Genética elemental y fitomejoramiento práctico*. Limusa.
- Rodríguez, A. y Pérez, A. (2017). Métodos científicos de indagación y de construcción del conocimiento. *Revista EAN*, 82, 179-200. <https://doi.org/10.21158/01208160.n82.2017.1647>
- Rojas, B. (1962). El diseño de San Cristóbal para el experimento de fertilizantes. XI Reunión. International Society of Sugar Cane. *Technologist Proceedings*. <https://issct.org/resources/proceedings/>
- Rojas, M. (2006). *Manual de redacción científica*. Universidad Nacional de San Marcos. <https://mrojas.perulactea.com/wp-content/uploads/2008/04/mrc.pdf>
- Rojas, M. (2015). Tipos de Investigación científica: Una simplificación de la complicada incoherente nomenclatura y clasificación. *Revista electrónica de Veterinaria, REDVET*, 16 (1), 1-14. <https://www.redalyc.org/pdf/636/63638739004.pdf>
- Rojo, J. M. (2007). *Regresión lineal múltiple*. Instituto de Economía y Geografía. [http://humanidades.cchs.csic.es/cchs/web\\_UAE/tutoriales/PDF/Regresion\\_lineal\\_multiple\\_3.pdf](http://humanidades.cchs.csic.es/cchs/web_UAE/tutoriales/PDF/Regresion_lineal_multiple_3.pdf)
- Ruiz-Ramírez, J., Hernández-Rodríguez, G. E., Ruiz-Bello, R. y Sánchez-Viveros, G. (2015). Relación de la eficiencia del diseño experimental en la calidad de los experimentos de caña de azúcar. *Terra Latinoamericana*, 33(3), 247-250.
- Salazar-Cuque, O. R. (2010). *Principios de estadística aplicados a las ciencias ambientales y agrícolas*. Universidad Rafael Landívar.

- Sarandón, S. J. y Flores, C. C. (2009). Evaluación de la sustentabilidad en agroecosistemas: Una propuesta metodológica. *Agroecología* 4, 19-28.  
<https://revistas.um.es/agroecologia/article/view/117131>
- Sarandón, S. J. y Flores, C. C. (2014). *Agroecología: bases teóricas para el diseño y manejo de agroecosistemas sustentables*. Libros de cátedra. Facultad de Ciencias Agrarias y Forestales. Universidad Nacional de La Plata.
- Siegel, S. (1990). *Estadística no paramétrica aplicada a las ciencias de la conducta*. 3ra. ed. Trillas
- Sotomayor, I. A. y Duicela, L. A. (1995). *Control integrado de las principales enfermedades del café en el Ecuador*. Instituto Nacional de Investigaciones Agropecuarias.  
<https://repositorio.iniap.gob.ec/bitstream/41000/1617/1/Control%20de%20enfermedades%20café.pdf>
- Suárez, M. O. y Tapia, F. A. (2018). *Interaprendizaje de estadística básica*. 2da. ed. Universidad Técnica del Norte, Facultad de Ciencias Económicas y Administrativas.
- Supo, J. (2013). *Cómo validar un instrumento*. Bioestadístico.  
[http://www.cua.uam.mx/pdfs/coplavi/s\\_p/doc\\_ng/validacion-de-instrumentos-de-medicion.pdf](http://www.cua.uam.mx/pdfs/coplavi/s_p/doc_ng/validacion-de-instrumentos-de-medicion.pdf)
- Supo, J. (2017). *Portafolio de aprendizaje para la docencia en investigación científica*. Bioestadístico EIRL. <https://idoc.pub/documents/idocpub-pnxk8gdyxg4v>
- Supo, J. y Cavero, H. (2014). *Fundamentos teóricos y procedimentales de la investigación científica en ciencias sociales*. Universidad Nacional del Altiplano – Puno y Universidad Andina Néstor Cáceres Velásquez – Juliaca.
- Sydsaeter, K. y Hammond, P. (1996). *Matemáticas para el análisis económico*. Prentice Hall. Universidad Nacional de San Marcos. 774 p.
- Universidad de Granada. (2016). *Acceder a la estadística sin limitaciones: Un entorno virtual de autoaprendizaje de la estadística*. Estadística. <http://wpd.ugr.es/~bioestad/>
- Urdaneta, O. y Urdaneta, M. (2016). Pruebas paramétricas versus pruebas no paramétricas y sus aplicaciones en la investigación odontológica. *Acta Odontológica Venezolana*, 54 (1).  
<https://www.actaodontologica.com/ediciones/2016/1/art-6/>
- Vásquez, E. R. y Caballero, A. (2011). Inconsistencia del Coeficiente de Variación para expresar la variabilidad de un experimento en un modelo de Análisis de Varianza. *Cultivos Tropicales*, 32(3), 42-45. <http://scielo.sld.cu/pdf/ctr/v32n3/ctr06311.pdf>
- Villasmil, P., Martínez, E. y Segura, G. (1974). El diseño San Cristóbal y su uso en ensayos de fertilización en caña de azúcar. *Revista de la Facultad de Agronomía, Universidad del Zulia*, 3(4), 7-25.
- Walpole, R. E., Myers, M. H. y Myers, S. L. (2012). *Probabilidad y estadística para ingeniería y ciencias*. 9na. ed. Pearson.
- Yengle, C. (2012). Aplicación del análisis de componentes principales como técnica para obtener índices sintéticos de calidad ambiental. *UCV Scientia*, 4(2), 145-153.  
<https://dialnet.unirioja.es/servlet/articulo?codigo=4369401>
- Zapata, A. y Sarache, W.A. (2014). Mejoramiento de la calidad del café soluble utilizando el método Taguchi. *Revista chilena de ingeniería*, 22(1), 116-124. <http://dx.doi.org/10.4067/S0718-33052014000100011>

**ANEXOS**





## Anexo 1: Glosario estadístico

**Aleatorización.**- Técnica experimental para “esparcir” el error experimental al azar entre individuos, grupos o tratamientos, asegurando que todos los elementos tengan la misma oportunidad de conformar una muestra.

**Análisis de varianza.**- Técnica estadística que permite desagregar las respuestas medias de las distintas fuentes de variación en un experimento.

**Arreglo factorial completo.**- Tipo de diseño experimental que explora todas las posibles combinaciones de los niveles y factores en estudio.

**Arreglo factorial incompleto.**- Tipo de diseño experimental que explora organizadamente solo una fracción de las posibles combinaciones de los niveles y los factores en estudio.

**Arreglo ortogonal.**- Matriz de distribución de niveles en estudio por cada factor en una tabla de doble entrada, para posibilitar el análisis racional y por separado de las variables de respuesta.

**Bloqueo.**- Variación controlada de un factor externo, cuando se evidencia una gradiente que se mantiene en una sola dirección en un ensayo

**Característica de calidad.**- Medida de un atributo o variable que presenta variabilidad medible con el uso de escalas.

**Censo.**- Es el registro individual de los datos de interés, en todos los elementos u objetos que constituyen una población. Los datos de un censo, por ejemplo, del peso en gramos por planta, genera un parámetro que es un valor característico, único y constante como  $\mu$ .

**Coefficiente de correlación.**- Es una medida de la intensidad de la relación entre dos variables y puede tomar valores entre -1,00 y 1,00.

**Densidad poblacional.**- El número de individuos por unidad de área o de volumen.

**Desviación típica.**- Medida de dispersión en torno a la media. En una distribución normal, el 68% de los datos se encuentra dentro  $\pm$  una desviación típica alrededor de la media y el 95% de los casos se encuentra dentro  $\pm$  dos desviaciones típicas.

**Diseño de parámetros.**- Acción de determinar las características de calidad para tener una respuesta óptima.

**Diseño de tratamientos.**- Indicación de los niveles por cada factor en estudio o de sus combinaciones a ser probadas en un experimento.

**Diseño experimental.**- Técnica de planeación que viabiliza la aplicación de una prueba estadística para explicar las relaciones de causalidad con un grado de confianza específico a partir de la manipulación deliberada de niveles cada factor en estudio.

**Dispersión de la población.**- Patrón de distribución en el plano cartesiano de los individuos dentro de los límites de su población.

**Distribución de frecuencias.**- Información sobre los valores concretos que adopta una variable y sobre el número de veces que se repiten. Se puede representar en porcentaje.

**Efecto principal.**- Variación de la respuesta media causada por el cambio de nivel en un factor en estudio.

**Error estándar.**- Es una estimación de la cantidad en que varía un estadígrafo de muestra a muestra.

**Error experimental.**- Incertidumbre en los estadígrafos atribuida a factores externos, ajenos a los tratamientos.

**Estadística descriptiva.**- Tiene como objetivo caracterizar, describir y extraer conclusiones sobre una muestra de datos.

**Estadística inferencial.**- Implica estimar los parámetros de la población a partir de datos muestrales, asumiendo un riesgo y error en las decisiones.

**Estadística.**- Ciencia que se ocupa de la ordenación y análisis de los datos de muestras representativas y de la realización de inferencias sobre las poblaciones de las que proceden.

**Estadístico.**- Propiedad descriptiva de una muestra.

**Experimento factorial completo.**- Los tratamientos cubren todas las posibles combinaciones de los niveles en todos los factores en estudio.

**Experimento factorial fraccionado.**- Los tratamientos en estudio solo cubren una fracción seleccionada de las posibles combinaciones de niveles por factor.

**Experimento multifactorial.**- Los tratamientos resultan de la combinación de varios factores de interés a ser probados en un ensayo.

**Factor de ruido.**- Variable externa que tiene efectos sobre la respuesta experimental que no se puede controlar.

**Factor en estudio.**- Variable independiente de tipo cuantitativo o cualitativo que se manipula para reproducir los efectos en el comportamiento de las variables de respuesta.

**Fiabilidad.**- Es la precisión de las medidas, en estabilidad, equivalencia y consistencia o coherencia. Un instrumento es fiable cuando aplicado varias veces en circunstancias similares permite obtener medidas consistentes.

**Frecuencia absoluta.**- Número de veces que aparece repetida un dato o un grupo de datos en el conjunto de las observaciones realizadas.

**Frecuencia relativa.**- Cociente entre la frecuencia absoluta de la clase y el número total de datos.

**Grados de libertad.**- Comparaciones posibles dentro de cada fuente de variación en un experimento.

**Hipótesis estadística.**- La hipótesis es una conjetura o proposición que el investigador ofrece como respuesta teórica al problema de investigación, la misma que tiene que probarse mediante una prueba estadística

**Índice de asimetría.**- Grado en el que los datos de una serie se reparten por encima y por debajo de la media aritmética. Si el coeficiente de asimetría tiende a cero, la distribución normal es simétrica, si el coeficiente es mayor que 0,50 la curva tiende a mostrar un sesgo a la derecha (asimetría positiva) y si es mayor que - 0,50 la curva tiende a mostrar sesgo a la izquierda (asimetría negativa).

**Índice de curtosis.**- Grado de apuntamiento de una distribución de frecuencias. Para una distribución normal óptima, el coeficiente o índice de curtosis se aproxima a cero. Una curtosis positiva mayor que 0,50 indica que los datos están concentrados alrededor de la media dando una forma puntiaguda, mientras que una curtosis negativa  $> - 0,50$  indica que los datos están dispersos, no concentrados, dando la forma de una curva aplanada.

**Individuo.-** Cada uno de los elementos que componen una población o una muestra. Es el ente observable que tiene un conjunto de características medibles.

**Interacción.-** Expresión de una variable de respuesta en términos de aditividad, antagonismo o sinergia relacionada al efecto de los distintos niveles por factor en estudio.

**Media.-** Medida de tendencia general de una variable X en un muestra n de sujetos o individuos.

**Mediana.-** Medida de tendencia central que divide la serie de datos de la variable X en dos mitades iguales (50% de datos a cada lado).

**Medidas de dispersión.-** Miden la variabilidad de un conjunto de datos. Las medidas más utilizadas son: rango, varianza y desviación típica.

**Medidas de posición.-** Indican la posición relativa que ocupa un sujeto en una distribución de frecuencias. Las medidas más utilizadas son: los centiles, los deciles y los cuartiles.

**Medidas de tendencia central.-** Son medidas estadísticas que se usan para describir cómo se puede resumir la localización de los datos. Las medidas de tendencia central indican la agrupación los datos, siendo las más usadas: media aritmética, la mediana y la moda.

**Moda.-** Valor de la variable X que más aparece en un conjunto de datos (mayor frecuencia absoluta).

**Muestra.-** Subconjunto de n elementos de una población.

**Muestreo aleatorio simple.-** Método de extracción de muestras que garantiza que todos los elementos de la población tienen la misma probabilidad de ser elegidos para la muestra en cualquier momento de la extracción.

**Muestreo.-** Es un procedimiento estadístico por medio del cual se toma una muestra, asegurándose que sea una parte representativa de la población para que los estadígrafos tengan validez y haya confianza en la estimación de los parámetros.

**Parámetro.-** Propiedad descriptiva de una población.

**Percentil.-** Puntuación que deja debajo de sí un porcentaje acumulado de sujetos de la distribución.

**Población estadística.-** Conjunto de los datos de los elementos con condiciones físicas, químicas, biológicas o de otra naturaleza que son comunes a un objeto de estudio (p.e.: en un bosque, la población estadística son los datos de altura de los árboles).

**Población finita.-** Es el conjunto limitado en número de elementos, individuos u objetos con condiciones físicas, químicas, biológicas o de otra naturaleza, comunes a un objeto de estudio. Ejemplos: las vacas lecheras de un establo para ordeño, las macetas con plantas ornamentales en una vivienda, las *cajas petri* de un experimento de laboratorio y los alumnos de un aula de clase.

**Población infinita.-** Conjunto de elementos, individuos u objetos reales muy grandes en número, donde identificarlos y medirlos implicaría un costo elevado. Ejemplo: un bosque de 10, 20, 100 o más hectáreas (hay muchos árboles, donde medir la altura de fuste de todos resultaría muy costoso).

**Población real.-** Conjunto de elementos, individuos, objetos o sucesos que tienen valor para la sociedad o un segmento de ella (p.e.: bosque = conjunto de árboles).

**Prueba F.-** Prueba de hipótesis de Fisher utilizada para determinar si dos muestras son homocedásticas o comparar medias de tratamientos con un nivel de significación concreto.

**Rango.-** Diferencia entre los valores máximo y mínimo de una variable numérica. El valor máximo menos el valor mínimo.

**Regresión lineal.-** Estimación de los coeficientes que explican la dependencia de una variable en función de una o más variables independientes, posibilitando pronosticar una asociación estadística lineal.

**Regresión no lineal.-** Estimación de los coeficientes que explican el comportamiento no lineal de una variable dependiente en función de una independiente.

**Robustez del diseño.-** Condición que prevé una reducida variación no explicada en un experimento, asegurando una alta confianza en los resultados.

**Tipo de variable.-** Respuesta esperada del comportamiento de una variable en estudio, que pueden ser de tres tipos: Mayor es mejor, menor es mejor o nominal es mejor.

**Tratamiento.-** Un nivel simple de un factor en estudio o una combinación de niveles por factores en estudio que se someten a una prueba estadística.

**Variable cualitativa.-** La que implica variaciones por clasificación o jerarquización, entre una observación y otra, en las poblaciones o muestras.

**Variable cuantitativa.-** La que implica cambios en la magnitud, entre una observación y otra, como resultado de conteos y mediciones numéricas, en poblaciones o muestras.

**Variable de respuesta.-** Variable experimental, cuantitativa o cualitativa, que muestra el efecto de la manipulación de los factores en estudio

**Variable dependiente.-** Es la variable que está en función de otra (o depende de ...), cuya magnitud cambia cuando varía la variable independiente.

**Variable independiente.-** Es la que origina los cambios en la magnitud de otra.

**Variable.-** Cualquier carácter, característica, rasgo o atributo que pueda mostrar valores numéricos o categóricos diferenciados entre una observación y otra.

**Anexo 2:** Valores críticos del Coeficiente de correlación de Spearman

Pares de observaciones →	4	5	6	7	8	9	10	12	14
$\alpha = 0,05$	1,00	0,90	0,829	0,714	0,643	0,60	0,564	0,506	0,456
$\alpha = 0,01$	-	1,00	0,943	0,893	0,833	0,783	0,746	0,712	0,645

Pares de observaciones →	16	18	20	22	24	26	28	30
$\alpha = 0,05$	0,425	0,399	0,377	0,359	0,343	0,329	0,317	0,306
$\alpha = 0,01$	0,601	0,564	0,534	0,508	0,485	0,465	0,448	0,432

Nota: Pares de observaciones (X, Y) y el nivel de significación

Fuente: Siegel (1990, p. 318)

Anexo 3: Lectura de valores críticos  $p$  de  $Z$  en la distribución normal

<b>z</b>	<b>0,00</b>	<b>0,01</b>	<b>0,02</b>	<b>0,03</b>	<b>0,04</b>	<b>0,05</b>	<b>0,06</b>	<b>0,07</b>	<b>0,08</b>	<b>0,09</b>
<b>0,0</b>	0,500	0,504	0,508	0,512	0,516	0,520	0,524	0,528	0,532	0,536
<b>0,1</b>	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
<b>0,2</b>	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
<b>0,3</b>	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
<b>0,4</b>	0,655	0,659	0,663	0,666	0,670	0,674	0,677	0,681	0,684	0,688
<b>0,5</b>	0,691	0,695	0,698	0,702	0,705	0,709	0,712	0,716	0,719	0,722
<b>0,6</b>	0,726	0,729	0,732	0,736	0,739	0,742	0,745	0,749	0,752	0,755
<b>0,7</b>	0,758	0,761	0,764	0,767	0,770	0,773	0,776	0,779	0,782	0,785
<b>0,8</b>	0,788	0,791	0,794	0,797	0,800	0,802	0,805	0,808	0,811	0,813
<b>0,9</b>	0,816	0,819	0,821	0,824	0,826	0,829	0,831	0,834	0,836	0,839
<b>1,0</b>	0,841	0,844	0,846	0,848	0,851	0,853	0,855	0,858	0,860	0,862
<b>1,1</b>	0,864	0,867	0,869	0,871	0,873	0,875	0,877	0,879	0,881	0,883
<b>1,2</b>	0,885	0,887	0,889	0,891	0,893	0,894	0,896	0,898	0,900	0,901
<b>1,3</b>	0,903	0,905	0,907	0,908	0,910	0,911	0,913	0,915	0,916	0,918
<b>1,4</b>	0,919	0,921	0,922	0,924	0,925	0,926	0,928	0,929	0,931	0,932
<b>1,5</b>	0,933	0,934	0,936	0,937	0,938	0,939	0,941	0,942	0,943	0,944
<b>1,6</b>	0,945	0,946	0,947	0,948	0,949	0,951	0,952	0,953	0,954	0,954
<b>1,7</b>	0,955	0,956	0,957	0,958	0,959	0,960	0,961	0,962	0,962	0,963
<b>1,8</b>	0,964	0,965	0,966	0,966	0,967	0,968	0,969	0,969	0,970	0,971
<b>1,9</b>	0,971	0,972	0,973	0,973	0,974	0,974	0,975	0,976	0,976	0,977
<b>2,0</b>	0,977	0,978	0,978	0,979	0,979	0,980	0,980	0,981	0,981	0,982
<b>2,1</b>	0,982	0,983	0,983	0,983	0,984	0,984	0,985	0,985	0,985	0,986
<b>2,2</b>	0,986	0,986	0,987	0,987	0,987	0,988	0,988	0,988	0,989	0,989
<b>2,3</b>	0,989	0,990	0,990	0,990	0,990	0,991	0,991	0,991	0,991	0,992
<b>2,4</b>	0,992	0,992	0,992	0,992	0,993	0,993	0,993	0,993	0,993	0,994
<b>2,5</b>	0,994	0,994	0,994	0,994	0,994	0,995	0,995	0,995	0,995	0,995

Nota: El valor  $Z$  equivale a distancias expresadas en desviaciones estándar a partir de  $\mu = 0$

<https://personales.unican.es/salvadol/tablnorm.xls>

- Caso: El valor  $Z = 1,96$ , en la prueba de dos colas, significa  $1 - \frac{\alpha}{2}$ , que equivale a 0,975 (2,5 % de equivocarse en cada cola,  $\alpha = 5$  % total)
- Caso: El valor  $Z = 2,57$ , en la prueba de una cola, significa  $1 - \alpha$ , que equivale a 0,995 (0,5 % de equivocarse en cada cola,  $\alpha = 1$  % total)

**Anexo 4:** Lectura de valores críticos Chi cuadrada para varios niveles de significación

Grados de libertad	Percentiles en distribución Chi-cuadrado									
	0,001	0,005	<b>0,01</b>	0,02	0,025	<b>0,05</b>	0,1	0,2	0,3	0,4
1	10,828	7,879	6,635	5,412	5,024	3,841	2,706	1,642	1,074	0,708
2	13,816	10,597	9,210	7,824	7,378	5,991	4,605	3,219	2,408	1,833
3	16,266	12,838	11,345	9,837	9,348	7,815	6,251	4,642	3,665	2,946
4	18,467	14,860	13,277	11,668	11,143	9,488	7,779	5,989	4,878	4,045
5	20,515	16,750	15,086	13,388	12,833	11,070	9,236	7,289	6,064	5,132
6	22,458	18,548	16,812	15,033	14,449	12,592	10,645	8,558	7,231	6,211
7	24,322	20,278	18,475	16,622	16,013	14,067	12,017	9,803	8,383	7,283
8	26,124	21,955	20,090	18,168	17,535	15,507	13,362	11,030	9,524	8,351
9	27,877	23,589	21,666	19,679	19,023	16,919	14,684	12,242	10,656	9,414
10	29,588	25,188	<b>23,209</b>	21,161	20,483	<b>18,307</b>	15,987	13,442	11,781	10,473
11	31,264	26,757	24,725	22,618	21,920	19,675	17,275	14,631	12,899	11,530
12	32,909	28,300	26,217	24,054	23,337	21,026	18,549	15,812	14,011	12,584
13	34,528	29,819	27,688	25,472	24,736	22,362	19,812	16,985	15,119	13,636
14	36,123	31,319	29,141	26,873	26,119	23,685	21,064	18,151	16,222	14,685
15	37,697	32,801	30,578	28,259	27,488	24,996	22,307	19,311	17,322	15,733
16	39,252	34,267	32,000	29,633	28,845	26,296	23,542	20,465	18,418	16,780
17	40,790	35,718	33,409	30,995	30,191	27,587	24,769	21,615	19,511	17,824
18	42,312	37,156	34,805	32,346	31,526	28,869	25,989	22,760	20,601	18,868
19	43,820	38,582	36,191	33,687	32,852	30,144	27,204	23,900	21,689	19,910
20	45,315	39,997	37,566	35,020	34,170	31,410	28,412	25,038	22,775	20,951
21	46,797	41,401	38,932	36,343	35,479	32,671	29,615	26,171	23,858	21,991
22	48,268	42,796	40,289	37,659	36,781	33,924	30,813	27,301	24,939	23,031
23	49,728	44,181	41,638	38,968	38,076	35,172	32,007	28,429	26,018	24,069
24	51,179	45,559	42,980	40,270	39,364	36,415	33,196	29,553	27,096	25,106
25	52,620	46,928	44,314	41,566	40,646	37,652	34,382	30,675	28,172	26,143

Nota: Usar el nivel de significación y los grados de libertad

<https://statologos.com/tabla-de-distribucion-de-chi-cuadrado/>

- Caso: En un ensayo con grados de libertad (GL) = 10 y  $\alpha = 0,05$   
 $\chi^2 = 18,307$
- Caso: En un ensayo con grados de libertad (GL) = 10 y  $\alpha = 0,01$   
 $\chi^2 = 23,209$



**Anexo 5:** Lectura de los valores críticos t de Student para distintos grados de libertad y niveles de significación

Grados de libertad	Probabilidades (arriba 2 colas, abajo 1 cola)									
	50,0%	25,0%	20,0%	10,0%	5,0%	2,0%	1,0%	0,20%	0,10%	
	25,0%	12,5%	10,0%	5,0%	2,5%	1,0%	0,5%	0,10%	0,05%	
1	1,000	2,414	3,078	6,314	12,7062	31,821	63,6567	318,309	636,619	
2	0,816	1,604	1,886	2,920	4,3027	6,965	9,9248	22,327	31,599	
3	0,765	1,423	1,638	2,353	3,1824	4,541	5,8409	10,215	12,924	
4	0,741	1,344	1,533	2,132	2,7764	3,747	4,6041	7,173	8,610	
5	0,727	1,301	1,476	2,015	2,5706	3,365	4,0321	5,893	6,869	
6	0,718	1,273	1,440	1,943	2,4469	3,143	3,7074	5,208	5,959	
8	0,706	1,240	1,397	1,860	2,3060	2,896	3,3554	4,501	5,041	
<b>10</b>	0,700	1,221	1,372	1,812	<b>2,2281</b>	2,764	<b>3,1693</b>	4,144	4,587	
12	0,695	1,209	1,356	1,782	2,1788	2,681	3,0545	3,930	4,318	
15	0,691	1,197	1,341	1,753	2,1314	2,602	2,9467	3,733	4,073	
20	0,687	1,185	1,325	1,725	2,0860	2,528	2,8453	3,552	3,850	
<b>30</b>	0,683	1,173	1,310	1,697	<b>2,0423</b>	2,457	<b>2,7500</b>	3,385	3,646	

Nota: Usar el nivel de significación y los grados de libertad (n - 1).

<https://es.scribd.com/document/351442327/Distribucion-t-de-Student-en-Excel#>

- *Caso:* En un ensayo con 10 grados de libertad y  $\alpha = 0,05 \sim 5\%$   
 $\alpha/2 = 0,025 \rightarrow 2,5\%$   
 Valor t = 2,2282
- *Caso:* En un ensayo con 30 grados de libertad y  $\alpha = 0,05 \sim 5\%$   
 $\alpha/2 = 0,025 \rightarrow 2,5\%$   
 Valor t = 2,0423

**Anexo 6: Lectura de los valores críticos de F con  $\alpha = 0,05$**

Grados de libertad del error	$\alpha = 0,05$										
	Grados de libertad del numerador										
	1	2	3	4	5	6	7	8	9	10	
1	161,446	199,499	215,707	224,583	230,160	233,988	236,767	238,884	240,543	241,882	
2	18,513	19,000	19,164	19,247	19,296	19,329	19,353	19,371	19,385	19,396	
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,785	
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	
8	5,318	4,459	4,066	3,838	3,688	3,581	3,500	3,438	3,388	3,347	
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544	
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494	
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450	
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412	
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378	
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348	

Nota: Usar los grados de libertad de los tratamientos o repeticiones (numerador) y los grados de libertad del error (denominador).

<https://statologos.com/f-valor-critico-excel>

- *Caso:* En un ADEVA de 5 tratamientos:  
 $\alpha = 0,05 \sim 5\%$  de probabilidad de equivocarse  
 Grados de libertad de tratamientos (GL) =  $t - 1 = 4$   
 Grados de libertad del error (GLE) = 12  
 Valor de F = 3,259

Anexo 7: Lectura de Amplitud Estudentizada de Tukey (Q) con  $\alpha = 0,05$  y  $0,01$

Grados de libertad del error	$\alpha$	Número de tratamientos								
		2	3	4	5	6	7	8	9	10
5	0,05	3,64	4,60	5,22	5,67	6,03	6,33	6,58	6,80	6,99
	0,01	5,70	6,97	7,80	8,42	8,91	9,32	9,67	9,97	10,24
6	0,05	3,46	4,34	4,90	5,31	5,63	5,89	6,12	6,32	6,49
	0,01	5,24	6,33	7,03	7,56	7,97	8,32	8,61	8,87	9,10
7	0,05	3,34	4,16	4,68	5,06	5,36	5,61	5,82	6,00	6,16
	0,01	4,95	5,92	6,54	7,01	7,37	7,68	7,94	8,17	8,37
8	0,05	3,26	4,04	4,53	4,89	5,17	5,40	5,60	5,77	5,92
	0,01	4,74	5,63	6,20	6,63	6,96	7,24	7,47	7,68	7,87
9	0,05	3,20	3,95	4,42	4,76	5,02	5,24	5,43	5,60	5,74
	0,01	4,60	5,43	5,96	6,35	6,66	6,91	7,13	7,32	7,49
10	0,05	3,15	3,88	4,33	4,65	4,91	5,12	5,30	5,46	5,60
	0,01	4,48	5,27	5,77	6,14	6,43	6,67	6,87	7,05	7,21
11	0,05	3,11	3,82	4,26	4,57	4,82	5,03	5,20	5,35	5,49
	0,01	4,39	5,14	5,62	5,97	6,25	6,48	6,67	6,84	6,99
12	0,05	3,08	3,77	4,20	4,51	4,75	4,95	5,12	5,27	5,40
	0,01	4,32	5,04	5,50	5,84	6,10	6,32	6,51	6,67	6,81
13	0,05	3,06	3,73	4,15	4,45	4,69	4,88	5,05	5,19	5,32
	0,01	4,26	4,96	5,40	5,73	5,98	6,19	6,37	6,53	6,67
14	0,05	3,03	3,70	4,11	4,41	4,64	4,83	4,99	5,13	5,25
	0,01	4,21	4,89	5,32	5,63	5,88	6,08	6,26	6,41	6,54
15	0,05	3,01	3,67	4,08	4,37	4,60	4,78	4,94	5,08	5,20
	0,01	4,17	4,83	5,25	5,56	5,80	5,99	6,16	6,31	6,44
16	0,05	3,00	3,65	4,05	4,33	4,56	4,74	4,90	5,03	5,15
	0,01	4,13	4,78	5,19	5,49	5,72	5,92	6,08	6,22	6,35

Nota: Usar el número de tratamientos y los grados de libertad del error del experimento.  
<https://www.lifeder.com/prueba-de-tukey/>

- *Caso:* Ensayo de 5 tratamientos con 10 grados de libertad del error:  
 $\alpha = 0,05 \sim 5\%$  de probabilidad de equivocarse  
 $Q = 4,65$   
 $\alpha = 0,01 \sim 1\%$  de probabilidad de equivocarse  
 $Q = 6,14$





Carrera de  
**INGENIERÍA  
AGRÍCOLA**

**Solubles  
Instantáneos C.A.**  
DESDE 1960



*“Un día llegará en  
que el pensamiento  
estadístico será tan  
necesario para  
ejercer la ciudadanía  
con eficiencia, como la  
capacidad de leer y  
escribir ”*

HERBERT GEORGE WELLS

ISBN: 978-9942-606-24-2



9 789942 606242

ISBN: 978-9942-606-25-9



9 789942 606259