Bioestadística:
Aplicaciones en Producción y Salud Animal

Segunda Edición



Carlos Solarte Portilla
Hernán Abdon García
Marco Antonio Imués Figueroa

Editorial
Universidad de Nariño

êditorial
Universidad de **Nariño**

Bioestadística:
Aplicaciones en Producción y Salud Animal

Segunda Edición

Bioestadística:
Aplicaciones en Producción y Salud Animal

Segunda Edición

Carlos Solarte Portilla
Hernán Abdón García
Marco Antonio Imués Figueroa

êditorial
Universidad de **Nariño**

Solarte Portilla, Carlos
Bioestadística: Aplicaciones en producción y salud animal / Carlos Solarte Portilla, Hernán Abdón García, Marco Antonio Imués Figueroa. -- 2ª. ed. -- San Juan de Pasto : Editorial Universidad de Nariño, 2026.

445 p. : il. byn., gráficas, tablas
Incluye bibliografía p. 380-385 y reseña de los autores p. 443
ISBN: 978-628-7864-97-9 Impreso
ISBN: 978-628-7864-96-2 Digital

1.Bioestadística 2. Salud animal--Estadísticas 3. Producción animal--Bioestadística 4. Salud animal--Bioestadística 5. Estadística vital I. García, Hernán Abdón II. Imués Figueroa, Marco Antonio

570.15195 S684b - SCDD-Ed. 23

Universidad de Nariño
FUNDADA EN 1904

Universidad de Nariño
ACREDITADA EN ALTA CALIDAD
RESOLUCIÓN MEN 000012 - ENERO 11 DE 2023

SECCIÓN DE BIBLIOTECA

Bioestadística: Aplicaciones en producción y salud animal

© Editorial Universidad de Nariño
© Carlos Solarte Portilla
Hernán Abdón García
Marco Antonio Imués Figueroa

ISBN impreso: 978-628-7864-97-9
ISBN digital: 978-628-7864-96-2
DOI: https://doi.org/10.22267/lib.udn.075
Segunda edición

Corrección de estilo: German Chaves Jurado

Diseño de cubiertas: Angie Gabriela Ordoñez y Nathaly Johana Rivadeneira

La imagen de portada fue creada mediante herramientas de inteligencia artificial generativa asistidas por ChatGPT, desarrollado por OpenAI. La composición final, selección y adecuación gráfica para esta publicación fueron realizadas por Angie Ordoñez y Nathaly Rivadeneira

Diagramación: Nathaly Johana Rivadeneira

Fecha de publicación: Junio 2026
San Juan de Pasto, Nariño, Colombia

Prohibida la reproducción total o parcial, por cualquier medio o con cualquier propósito, sin la autorización escrita de su Autor o de la Editorial Universidad de Nariño

Contenido

Presentación 12

Capítulo 1. Generalidades **15**

Estadística y método científico 20
Definiciones y conceptos 22
Datos 23
Variables 23
Escalas de medida 26
Población y Muestra 28

Capítulo 2. Presentacion, resumen y caracterizacion de la información **31**

Tablas de distribución de frecuencia 31
Construcción de Tablas de Distribución de Frecuencia 32
Gráficos estadísticos 40
Histograma 40
Polígono de Frecuencias 41
Ojiva o Polígono de Frecuencia Acumulada 42
Diagrama Circular 43
Curvas de Frecuencia 44
Medidas de tendencia central, posición y dispersión 46
Medidas de Tendencia Central 47
La Media o Promedio Aritmético 47
Promedio Ponderado 50
Media Geométrica 52
Media Armónica 53
Media Cuadrática 54
Cálculo de la Media Aritmética con datos agrupados 54
La Mediana (M_e) 56

Cálculo de la Mediana con datos agrupados 58
La Moda o Modo (M_o)..... 60
Cálculo de la Moda con datos agrupados..... 62
Medidas de Posición 63
Cuartiles..... 63
Cálculo de Cuartiles con datos agrupados..... 66
Deciles y Percentiles..... 69
Cálculo de Percentiles con datos agrupados..... 70
Medidas de Dispersión..... 71
La Varianza y la Desviación Estándar 71
Cálculo de la Varianza con datos agrupados 74
El Coeficiente de Variación o Variabilidad (CV)..... 76
El Rango (R)..... 77
El Rango Intercuartílico (RIQ)..... 77
Asimetría y curtosis..... 78
Coeficiente de Asimetría (C_A)..... 79
Coeficiente de Curtosis (C_{AP})..... 80
Diagramas de Caja..... 81

Capítulo 3. Conceptos básicos de probabilidad 84

Introducción 84
Conceptos básicos para el estudio de las probabilidades..... 85
Espacio muestral 88
Eventos 91
Operaciones con conjuntos 92
La Suma o Unión de Conjuntos (A ∪ B)..... 92
El Producto o Intersección de Conjuntos (A ∪ B) 93
Complemento de A (A^{-})..... 93
Probabilidad de un evento..... 96
Cálculo de probabilidades con espacios muestrales finitos 97
Sucesos Igualmente Probables 98
Métodos de Enumeración..... 102

Principio de la Multiplicación o del Conteo 103
Permutaciones 109
Combinaciones 110
Cálculo de Probabilidades cuando los resultados no son equiprobables 116
Otras reglas para calcular Probabilidades 117
Regla General de la Adición 117
Regla de la Multiplicación 118
Probabilidad Condicional 119
Eventos Independientes 127
Teorema de Bayes 127

Capítulo 4. Funciones de probabilidad para variables aleatorias discretas y continuas 136

Variable aleatoria 136
Rango de una Variable Aleatoria X 137
Variables Aleatorias Discretas 138
Función de Probabilidad para Variables Aleatorias Discretas 138
Función de Distribución Acumulativa 140
Valor Esperado o Esperanza Matemática de una Variable Aleatoria Discreta 143
Varianza de una Variable Aleatoria 144
Distribuciones de probabilidad especiales para variables discretas 145
La Distribución Binomial 145
La Distribución Poisson 150
La distribución Multinomial 156
Variables aleatorias continuas 157
La Distribución Normal 160
Propiedades de la Distribución Normal 161
Función de Distribución Acumulada de una Distribución Normal . 163
Distribución Normal Estándar 165
Teorema Central del Límite 174
Aproximación Normal de la Distribución Binomial 175

Capítulo 5. Distribuciones muestrales y estimación 178

Muestra aleatoria 178
Distribución de muestreo de la media muestral $\bar{Y}$ 184
Distribución de muestreo de la proporción muestral 186
Distribución muestral de la varianza S^2 188
Distribución Chi-cuadrada 188
Distribución Muestral de la Varianza S^2 191
Estimador 193
Estimador Insesgado 194
Eficiencia Relativa 194
Estimadores Insesgados de Varianza Mínima 195
Intervalos de confianza 195
Estimador por Intervalo 196
Intervalo de Confianza para la Media μ en muestras con Distribución Normal y Varianza conocida 197
Tamaño de la Muestra para estimar μ 199
Distribución *t* de Student 202
Intervalo de Confianza para $\mu\sigma$ con una Muestra pequeña y Varianza σ^2 desconocida 205
Intervalo de Confianza para la Diferencia de Medias 207
Intervalo de Confianza para la Diferencia de Medias ($\mu 1 - \mu 2$), con Varianzas σ^2_1 y σ^2_2 conocidas 208
Intervalo de Confianza para la Diferencia de Medias ($\mu 1 - \mu 2$) con Muestras pequeñas y Varianzas iguales $\sigma^2_1 = \sigma^2_2$ pero desconocidas .. 211
Intervalos de Confianza para Proporciones 213
Intervalo de Confianza para *p* con Muestra grande 214
Tamaño de Muestra para estimar *p* 215
Estimación de la Diferencia entre dos Proporciones 218
Intervalo de Confianza entre dos Proporciones $p1 - p2$ con Muestras grandes 218
Intervalo de Confianza para la Varianza 220
Intervalo de Confianza para σ^2 220
Intervalo de Confianza para el Cociente de dos Varianzas 222

Capítulo 6. Prueba o contraste de hipótesis 227

Región crítica o región de rechazo ... 232
Hipótesis unilaterales y bilaterales ... 234
Contrastes relacionados con la media μ ... 236
Contraste de Hipótesis para la Media Poblacional σ^2_1 con Varianza Poblacional σ^2_2 conocida ... 236
Relación entre Contraste de Hipótesis e Intervalo de Confianza ... 240
Contraste de Hipótesis para la Media Poblacional μ con Muestras grandes (n σ_2 30) y Varianza σ^2 desconocida ... 244
Contraste de Hipótesis para la Media Poblacional μ con Muestras pequeñas (n < 30) y σ^2 desconocida ... 246
Contrastes relacionados con la diferencia de medias ... 248
Contraste de Hipótesis para la Diferencia de Medias con Varianzas σ^2_1 y σ^2_2 conocidas ... 249
Contraste de Hipótesis para la Diferencia de Medias con Muestras grandes (n1, n2 < 30) y Varianzas σ^2_1 y σ^2_2 desconocidas ... 252
Contraste de Hipótesis para la Diferencia de Medias con Muestras pequeñas (n1, n2 < 30) y Varianzas desconocidas pero iguales ($\sigma^2_1 = \sigma^2_2$) 254
Contraste de Hipótesis para la Diferencia de Medias: datos pareados. 263
Cálculo del Tamaño de la Muestra para estimar la Diferencia de Medias .. 268
Muestras Independientes ... 268
Contrastes relacionados con proporciones ... 270
Contraste para la Proporción Poblacional: Muestras grandes ... 270
Contraste para la Diferencia entre dos Proporciones: Muestras grandes ... 273
Contraste de hipótesis para la varianza poblacional σ^2 ... 276
Contraste de hipótesis para la Igualdad de dos Varianzas σ^2_1 y σ^2_2 ... 279
Contraste de bondad de ajuste ... 282
Pruebas de independencia ... 289
Prueba de homogeneidad ... 293
Introducción al análisis de varianza de datos ... 296
Análisis de Varianza de un Factor ... 296
Procedimiento para el ANOVA ... 298

Prueba de Hipótesis 299
Pruebas de Comparación Múltiple 305

Capítulo 7. Pruebas no paramétricas 310

Prueba del signo 312
Prueba del Signo para Datos pareados 316
Prueba de Rango con Signo de Wilcoxon 318
Prueba no paramétrica para comparar dos poblaciones con base en muestras aleatorias independientes 324
Prueba de Mann-Whitney 324
Prueba de Kruskal-Wallis para k Muestras Independientes 328

Capítulo 8. Principios básicos de las relaciones entre variables 332

Aspectos generales 332
Relaciones funcionales de dependencia 333
Regresión lineal simple 333
Prueba de Hipótesis sobre la Significancia del Modelo de Regresión . 344
Coeficiente de Determinación 346
Límites de Confianza 347
Correlación lineal 350
Regresión lineal simple con notación y operaciones matriciales 355
Análisis de regresión no lineal 359
Análisis de residuales. Validación del modelo 367
Regresión lineal múltiple 369
Predicción de respuestas futuras 377

Referencias bibliográficas 380
Apéndice 386
Acerca de los autores 443

Presentación

En la formación de los Biólogos y de los profesionales de las diversas disciplinas que tienen su sustento científico en la Biología, incluidas las aplicadas al sector pecuario, el desarrollo de teorías y métodos estadísticos ha contribuido efectivamente a la solución de incontables problemas prácticos y al entendimiento de múltiples fenómenos de la Producción y la Salud Animal. Igualmente, el desarrollo, incluso la comercialización de nuevos fármacos y suplementos alimenticios, la valoración genética de las poblaciones animales, la comprensión cada vez más clara de la variabilidad genética en poblaciones animales naturales y domésticas, el diseño de planes para controlar o erradicar epizootias y pandemias, para citar solo algunos casos, han sido posibles gracias al desarrollo de teorías y procedimientos estadísticos, ligados al avance explosivo en materia de software y de hardware.

Un ejemplo concreto, en las aplicaciones prácticas, hoy rutinarias, es el caso del Mejoramiento Genético clásico, mediante la selección animal. Un elemento tan práctico e imprescindible para la cría animal, no hubiese sido posible construir, sin las teorías estadísticas relativas al análisis genético planteadas por Lush (1945), las cuales fueron continuadas, corregidas y aumentadas por Henderson (1975, 1984, 1986 y 1988) y Schaeffer (1983), Schaeffer y Henerson (1983) y Dekkers et al (1996), como exponentes de primera importancia en este campo; sin embargo, estas formulaciones teóricas jamás se hubiesen convertido en aplicaciones prácticas, sin los algoritmos propuestos principalmente, por Boldman y Van Vleck, (1991), Boldman et al (1993) y Meyer (1988, 1989, 1991, 1993 y 1998), quienes a su vez, tampoco habrían tenido éxito, sin la participación de los Ingenieros de Sistemas. Aquí es fácilmente observable el esfuerzo investigativo, en

conjunto con especialistas en estadística, matemáticos, ingenieros y profesionales del agro, que han permitido proporcionar a los criadores, una de las herramientas que más ha impactado en el progreso de la producción y salud animal, ya que, en gran medida, gracias a toda esta labor interdisciplinaria, en la actualidad se dispone de ejemplares cada vez más productivos, menos contaminantes, más resistentes a enfermedades, de mejor desempeño productivo y por supuesto, más provechosos para la humanidad, pero aun así, la investigación básica y aplicada no se detiene en la búsqueda de más y mejores opciones.

Conscientes de la necesidad de incorporar nuevas soluciones ajustadas a la realidad de cada entorno, hoy en día, las carreras profesionales del sector pecuario incluyen en sus planes de formación el componente de investigación, como un elemento básico, de tal manera que los estudiantes desde fases muy tempranas de aprendizaje, se involucran en diversos proyectos investigativos y, es allí, donde la Estadística constituye en una poderosa ayuda para interpretar resultados, planear experimentos y formular de modo claro y preciso, las recomendaciones del caso. Obviamente, los estudiantes y profesionales del sector pecuario no tienen obligación de formarse como estadísticos, pero sí deben tener los fundamentos teóricos suficientes para entender los procedimientos y utilizar correcta y eficazmente las aplicaciones prácticas de la Estadística, en asuntos referidos a su desempeño.

Esta realidad condujo a escribir el presente libro dirigido, fundamentalmente, a los estudiantes de pregrado de las carreras del sector pecuario. En todo el texto, se han incluido abundantes ejemplos, en muchos casos, con situaciones reales objeto de frecuentes consultas por parte de docentes y estudiantes, al igual que experimentos efectivamente realizados por los mismos, proyectos de investigación y trabajos de grado. Fueron precisamente los estudiantes, los

principales motivadores, para que se concibiera la idea de recopilar en un solo texto, los principales fundamentos y aplicaciones de la Biostadística en Producción y Salud Animal.

En este punto, es necesario comentar que, en un principio se discutió ampliamente la utilidad de escribir este libro, dada la existencia en el mercado de muy buenos textos de Estadística; de hecho, en la bibliografía se presenta la reseña de algunas obras. Sin embargo, un análisis detallado de varias de ellas permitió concluir que muy pocas están exclusivamente referidas a producción y salud animal, con ejemplos donde se describan situaciones reales a las que se ven enfrentados en sus actividades diarias los estudiantes, investigadores, profesores y profesionales del sector pecuario y en ninguna se encontró casos propios de la región, como era apenas previsible.

Por otra parte, no debe perderse de vista, que el objetivo principal de los estudiantes, es la formación en su área de desempeño, por eso se utiliza, a lo largo de este texto, un lenguaje técnico relativo a las diversas profesiones pecuarias, con la intención de inducirlos, desde sus primeros contactos con el conocimiento disciplinar correspondiente y a la profundización en diversas temáticas, como la Nutrición, la Patología, la Epidemiología, la Genética o cualquier otra área de las mencionadas en los diferentes capítulos de este libro, con sus correspondientes ejemplos. De esta forma, se espera cumplir dos retos principales: ofrecer elementos de formación básicos en Estadística aplicada a Producción y Salud Animal y motivar a los estudiantes, para que, desde el inicio de su formación, incursionen en la actividad investigativa y en la profundización sobre temáticas específicas aplicadas a las profesiones del sector pecuario. Un buen indicador del cumplimiento de los objetivos antes descritos, serán las observaciones, sugerencias e inquietudes por parte de los lectores, las cuales siempre serán bien recibidas.

CAPÍTULO 1

Generalidades

En el mercado de los libros es posible encontrar varios textos de Bioestadística, que abordan conceptos y aplicaciones en diferentes disciplinas de las Ciencias Biológicas y, cuando se presentan discusiones en varios documentos científicos, es común utilizar el término *Bioestadística* como sinónimo de *Biometría*; así se asumió en la primera edición. Sin embargo, después que los autores hicieran un análisis sobre sus similitudes y diferencias, es bueno realizar algunas precisiones.

Etimológicamente, *Bioestadística* proviene de la raíz griega ***bios*** (vida) y de la raíz latina ***statisticum*** (estado), por lo cual, integra la Estadística con el análisis de los datos obtenidos en los fenómenos biológicos, facilitando la aplicación de los métodos y técnicas estadísticas para analizar diversas variables propios de los seres vivos, lo que indica una relación obligatoria entre la Biología y la Estadística. Por su parte, la palabra *Biometría,* proviene de dos raíces griegas, ***bios*** vida y ***metron*** medida, es decir, se refiere a las mediciones que se hacen en los seres vivos; en consecuencia, la Biometría se ocupa del desarrollo teórico y práctico, para llevar a cabo mediciones en seres vivos, para cuyo análisis hace uso de la Estadística. De lo anterior se deduce, las notorias diferencias entre las dos técnicas metodológicas, aun cuando en el lenguaje de los investigadores, existe la tendencia a utilizar los dos términos como equivalentes.

Es claro, entonces, que la Biometría dedica sus esfuerzos a desarrollar metodologías que aseguren la medición de los fenómenos biológicos, para la obtención de datos, con la exactitud y precisión requeridas, de manera que garanticen la mayor fidelidad con los sucesos reales y utiliza la Estadística para analizar los datos obtenidos, con métodos y modelos específicos de cada disciplina. Por su parte, la Bioestadística hace uso de las metodologías estadísticas, para desarrollar técnicas y modelos que permitan realizar los análisis tendientes a obtener conclusiones válidas para la población de donde se extrajeron los datos. También es claro que, en los dos casos, se tiene como base fundamental la Estadística.

En las Ciencias Biológicas, como en todas las ciencias, mediante el proceso investigativo, se trata de entender fenómenos observados en su estudio, por tanto, se requiere tener un conocimiento, lo más amplio posible, de un fenómeno particular y utilizar conceptos y procedimientos estadísticos que conduzcan a encontrar explicaciones válidas sobre el comportamiento del mismo y establecer las relaciones causa-efecto entre los componentes de los sistemas analizados, lo cual ayudará a tomar decisiones oportunas y acertadas.

El estudio, análisis y evaluación de los sistemas, permite identificar los componentes Causa y los componentes Efecto, que se miden o cuantifican a través de las respectivas variables (Causa-Efecto), cuyas relaciones sólo pueden ser establecidas mediante la aplicación de métodos y modelos de la Estadística.

Ahora bien, la palabra Estadística tiene muchas definiciones de tipo lingüístico, usualmente encontradas en enciclopedias y diccionarios, sin embargo, dado que en la época actual el uso de la Estadística abarca prácticamente todas las áreas del saber, entre las que pueden mencionarse las Ciencias Sociales, la Ingeniería, la Química y por supuesto, las Ciencias Biológicas en sus diferentes aplicaciones, se hace difícil, en cierto modo, expresar una definición única y categórica, así

que, antes que la definición, es más importante tener una clara idea, de la utilidad y conocer algunos aspectos importantes de la evolución histórica de la Estadística, que se describe a continuación.

De acuerdo con Steel y Torrie (1985), la Estadística debió comenzar como una Aritmética estatal para asistir a los gobernantes.

Los mismos autores señalan que, una aplicación de la probabilidad empírica para asegurar los embarques, se puede encontrar en la historia de la navegación flamenca del siglo XIV; si bien hoy en día se puede considerar especulativa, seguramente dio origen a la Estadística utilizada en la actualidad por las empresas aseguradoras.

En el siglo XVII se originó, la Teoría de las Probabilidades, desarrollada por Pascal y Fermat. Un hecho histórico trascendental para el desarrollo contemporáneo de la Estadística, fue la ecuación planteada para la curva normal de errores, publicada originalmente en 1733 por *de Moivre*. El mencionado texto, permaneció en el anonimato, hasta que Karl Pearson lo redescubrió en 1924. También es importante destacar, que dos astrónomos matemáticos, Laplace y Gauss, en el siglo XVIII, trabajando en forma independiente, llegaron a las mismas conclusiones que obtuvo de Moivre, sin que tuvieran conocimiento de sus trabajos. La curva normal, conserva gran importancia en salud y producción animal y se la conoce universalmente con los nombres de *Curva de Gauss* o *Curva Laplaciana*.

La utilización de la Estadística para el análisis de fenómenos biológicos comenzó aproximadamente en 1925, año en que el profesor Ronald Fisher publicó el libro titulado *Statistical methods for research workers*. Searle (1989) y Hofer (1998) presentaron un recuento histórico del uso de los componentes de varianza, en los que se puede apreciar que, después de Fisher,

Tippett clarificó y extendió el método ANOVA (analysis of variance) de estimación. Yates (1935) utilizó la metodología ANOVA en experimentos con cereales; Neyman & Iwaszkiewicz (1935) consideraron la eficiencia de los diseños de bloques aleatorizados y los diseños en cuadrado latino e hicieron uso extensivo de los modelos lineales. Quizás estos trabajos sean la primera aparición reconocible de los modelos mixtos.

Entre 1943 y 1987, se extendió el uso de los Modelos de Componentes de Varianza, para resolver problemas prácticos, especialmente por parte de los genetistas, quienes aplicaron esta técnica en humanos, ganado lechero, ganado de carne, maíz, trigo y aves, entre otros. La mayoría de estas aplicaciones involucraron datos desbalanceados. Una revisión histórica detallada de estos tópicos se puede encontrar en Solarte (2000).

En la actualidad se presenta un continuo crecimiento y auge de la Estadística, dadas las nuevas necesidades del hombre, que a diario requiere más respuestas a nuevos interrogantes, por la ampliación del espectro investigativo, donde la Estadística, propone métodos efectivos para describir con precisión, los parámetros económicos, políticos, sociales, biológicos y físicos, como una herramienta para analizar y relacionar dichos datos.

Quizá la definición que más se ajusta a los fines de este texto, es la propuesta de Steel y Torrie (1985), quienes enuncian *"La Estadística como la ciencia pura y aplicada, que crea, desarrolla y aplica técnicas, de modo que pueda evaluarse la incertidumbre de inferencias inductivas"*. De la anterior definición se deduce, que la Estadística se acepta como una ciencia, no como una disciplina de las Matemáticas, como se concebía anteriormente, así que, como ciencia pura, desarrolla conocimiento nuevo o teorías estadísticas, mientras que como ciencia aplicada, desarrolla métodos, para recopilar, representar, condensar y analizar los datos obtenidos con una técnica

particular, según los objetivos del estudio y cualquiera sea la disciplina interesada en alcanzarlos y, de esa manera, determinar el nivel probabilístico de la duda que se presenta en un proceso inferencial, esto es, de generalización sobre las conclusiones obtenidas en el estudio de una muestra, para ser aplicadas a la población de donde proviene, aplicando un método inductivo.

Sierra (1991) por su parte, define a la Estadística como una *"Ciencia formada por el conjunto de teorías y técnicas cuantitativas, que tienen por objeto, la organización, presentación, descripción, resumen y comparación de conjuntos de datos numéricos, obtenidos de poblaciones de individuos o fenómenos en su conjunto o bien de muestras que representan las poblaciones estudiadas, así como el estudio de su variación, propiedades, relaciones, comportamiento probabilístico y la estimación, inferencia o generalización de los resultados obtenidos de las muestras, respecto a las poblaciones que aquellas representan"*.

Otra manera de definir a la Estadística, utilizada con mayor frecuencia en varios textos especializados, es la siguiente: *"Conjunto de métodos que se utilizan para obtener, recolectar, presentar, analizar e interpretar datos numéricos, con el objeto de extractar conclusiones válidas, con base en las cuales se puedan tomar decisiones"*.

Algunos autores aseguran que, la Estadística no es una ciencia, sino un grupo de métodos con base científica. Para dirimir discusiones de esta naturaleza, el lector interesado puede remitirse a Bunge (1976) y Popper (1971).

Convencionalmente, la Estadística se divide en dos grandes ramas: ***Estadística Descriptiva*** y ***Estadística Inferencial***. La primera hace referencia al conjunto de métodos para resumir y presentar en forma clara una serie de observaciones, durante el estudio de un fenómeno natural; aquí se incluyen

los cuadros, los gráficos, las medidas de tendencia central y de dispersión o variabilidad. La Estadística Inferencial comprende, el conjunto de métodos probabilísticos necesarios, para obtener conclusiones válidas para una población, como resultado del análisis de una muestra extraída de ella.

Para los fines de este texto, tanto el análisis descriptivo como el inferencial, son temas de interés, sin embargo, la Estadística Inferencial, será de mayor importancia, puesto que la producción y la salud animal, son áreas con intenso trabajo investigativo en procura de lograr animales más productivos y eficientes, menos susceptibles a enfermedades y, por ende, más útiles a la humanidad.

Finalmente, es bueno resaltar que, el profesional de las Ciencias Pecuarias, debe cimentar una sólida formación en el campo investigativo, siendo uno de los aspectos de mayor relevancia en los planes de formación de recursos humanos, con miras a lograr niveles de producción internacionalmente competitivos, acorde con la actual tendencia mundial, donde la Estadística tiene una participación fundamental.

ESTADÍSTICA Y MÉTODO CIENTÍFICO

La investigación en Producción y Salud Animal tiene, como objetivo primordial, la generación de nuevo conocimiento o nuevas aplicaciones por lo que, utiliza de manera rigurosa, el *Método Científico*.

La aplicación del Método Científico parte del conocimiento disponible sobre un problema concreto; para ello se revisan profundamente los hechos y teorías existentes, se formulan hipótesis para ser contrastadas mediante la experimentación, se lleva a cabo una evaluación objetiva de las hipótesis y, finalmente, con base en resultados experimentales, mediante el estudio de una muestra, se emiten las conclusiones del caso, aplicables a la población.

Un punto de bastante complejidad es el referente a la evaluación objetiva de las hipótesis, especialmente cuando se trabaja con organismos vivos, ya que en muchas situaciones se desconocen, con la precisión deseada, las relaciones causa-efecto, dada la amplia variación natural a que están sometidas las especies animales. Para ello, los investigadores de las Ciencias Animales, estudian casos particulares, mediante la experimentación o el estudio de muestras, para extenderlos a situaciones generales, cumpliendo así el procedimiento de *Inferencia Incierta*.

La Estadística es una herramienta de mucha utilidad en el Método Científico. Su aplicación es palpable, en aspectos que van desde la elaboración de un plan para recolectar los datos o diseñar uno o más experimentos, hacer el análisis de los resultados, el resumen, la presentación y caracterización de los datos, hasta concluir con la evaluación de la incertidumbre de las inferencias.

Para visualizar con más claridad la utilidad de la Biometría en el Método Científico positivista, es conveniente analizar la Figura 1.1, adaptada de Martínez y Martínez, (1997).

A manera de resumen, se podría asegurar que el profesional pecuario debe incentivar su capacidad de análisis para recopilar información e interpretar datos numéricos procedentes de la observación de múltiples fenómenos que se presentan en los sistemas de producción con especies animales, al igual que para el manejo apropiado de las técnicas estadísticas útiles en la investigación científica.

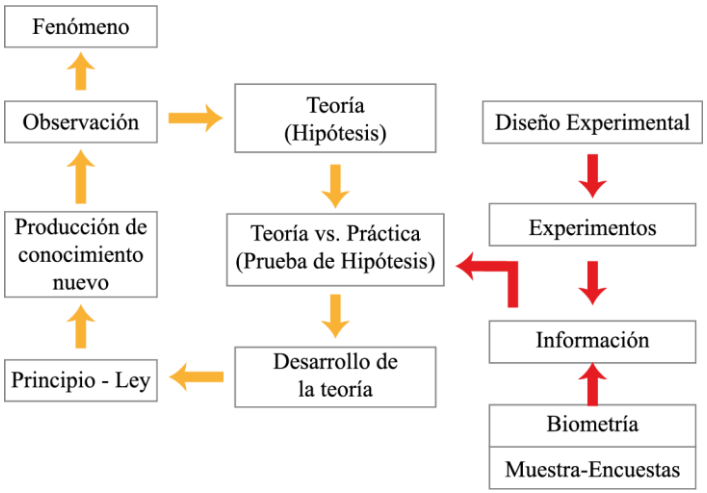


Figura 1.1. Ciclo de la Ciencia: Conexión entre la Biometría y el Método Científico (Adaptado de Martínez y Martínez, 1997).

Es necesario reiterar que, la Biometría es un instrumento de investigación, ya sea en Nutrición Animal, Genética, Farmacología, Patología y, en general, en cualquier campo de acción de los profesionales del agro y, sobre estos campos, se propondrán problemas específicos de investigación, por lo tanto, los instrumentos de investigación, como las técnicas y procedimientos biométricos, deben constituirse en herramientas de uso común para el profesional, de manera que sean utilizados en forma eficiente, al igual que las ayudas que proporciona el software especializado disponible. De este modo, los profesionales estarán capacitados, para presentar resultados en forma crítica, clara y objetiva y dispondrán de elementos, con base científica, en el momento de proponer soluciones.

DEFINICIONES Y CONCEPTOS

Para un desarrollo más ágil en la lectura y comprensión de este texto, es recomendable familiarizarse con ciertos términos que se utilizarán en las

temáticas siguientes. No es conveniente memorizar definiciones, sino comprender los conceptos, para su posterior análisis y aplicación; en algunas ocasiones será necesario diferenciar algunas expresiones que, en Estadística, significan una cosa y en el lenguaje cotidiano otra.

Datos. En Producción y Salud Animal, los fenómenos naturales, constituyen el espacio fundamental para proponer los trabajos de investigación. Cada vez que se observa un fenómeno, se mide o se cuenta los elementos que conforman un sistema, se obtiene un dato. Cuando se realizan análisis estadísticos debe tenerse en cuenta que, en la gran mayoría de situaciones, las observaciones estarán representadas en forma numérica, aunque también pueden ser características cualitativas; así por ejemplo, en Farmacología, las observaciones pueden ser los tiempos que tardan los animales en recuperarse luego de la aplicación de un antibiótico o el número de animales curados; en Mejoramiento Genético, las observaciones podrían estar constituidas, por el peso al nacer de los lechones provenientes de un cruzamiento entre dos razas; en Alimentación Animal, la ganancia de peso de los terneros sometidos a diferentes dietas; en el Manejo de Praderas, el rendimiento por hectárea de las variedades de pasto; en el estudio de Calidad de Aguas, la concentración de nutrientes; en la Identificación de Peces en un sistema natural, las observaciones o datos pueden ser la especie o el color. En general, las observaciones corresponden al ***fenómeno*** que se desea estudiar y los valores obtenidos de dichas observaciones, constituyen los ***datos*** que, en la mayoría de los casos, exhiben ***variación*** entre ellos.

Variables. En Estadística, el concepto de ***variable*** se refiere a cualquier característica o atributo de los individuos, elementos o componentes que conforman una población, como una estrategia para medir un fenómeno de interés en una investigación. Por ejemplo, en Nutrición se tendrá como objetivo

evaluar la calidad nutritiva de un alimento a través de la medición de variables, como la conversión alimenticia o consumo del alimento por parte de los animales. El profesional de la Acuicultura podría estar interesado en investigar la resistencia de una variedad del *Penaeus vannamei* al síndrome de la mancha blanca, midiendo la supervivencia, cuando se suministra un probiótico adicionado en el alimento. En virtud de que cualquiera de estas características, por regla general, presenta un valor diferente cuando se observa en diversas entidades, ella recibe el nombre de variable. Debido a que los valores que toma una variable, bien sean numéricos o cualitativos, provienen de factores fortuitos y su valor no es posible determinar o predecir con anticipación, ésta se denomina ***variable aleatoria***, al estar influenciada por al azar.

Las variables aleatorias pueden ser cuantitativas o cualitativas. Una ***variable cuantitativa*** está representada por valores de carácter numérico y expresa una magnitud. Por ejemplo, el peso al nacer de los bovinos, la alzada de los equinos, la edad al primer servicio de los caprinos, el número de galpones dedicados a la explotación de gallinas ponedoras en una región, el tamaño de los alevinos de trucha arco iris, el número de colonias de bacterias en un medio de cultivo, el número de animales enfermos y muchas otras más, que se puedan representar con datos numéricos. Las variables cuantitativas, a su vez, pueden ser continuas o discretas.

Una ***variable continua*** está representada por cualquier valor dentro de un intervalo, es decir que, teóricamente, puede ubicarse en cualquier punto de la recta real y puede tomar cualquier número real entre dos valores dados. La longitud y el peso de los animales, son ejemplos típicos.

Una ***variable discreta***, también llamada ***discontinua***, toma valores posibles, que no se pueden observar en una escala continua debido a la existencia de espacios entre estos datos y a la ausencia de valores intermedios; en otras

palabras, las variables discretas, se ubican en puntos precisos o determinados de la recta real, por lo que sólo toman valores enteros. Los conteos, producen datos que conforman variables discretas, por ejemplo, el número de pezones de una vaca, el número de productores de cuyes en un corregimiento, el número de lechones en un parto, el número de ovas que produce una hembra de tilapia, el número de animales recuperados después de un tratamiento con macrólidos y muchos otros.

Las ***variables cualitativas*** no se miden numéricamente, sino que representan cualidades y se asignan a una o varias categorías mutuamente excluyentes, es decir, una observación particular no es posible asignarla a más de una clase. Un ejemplo común, es la clasificación del ganado Holstein según el color del pelo, que puede ser negro o rojo, caso en el que habrá dos categorías; en el ganado Shortorn, con el mismo criterio, se pueden formar tres categorías: rojo, blanco o roano, este último también conocido como salinero; las especies ícticas para Acuicultura pueden ser agrupadas en dos categorías: de aguas frías y de aguas cálidas.

Para representar las variables en los análisis estadísticos, se emplean algunos símbolos, en lugar de escribir el nombre de ellas en cada oportunidad. Por ejemplo, si $\boldsymbol{Y}$ representa cualquier variable de interés, entonces $\boldsymbol{y_i}$ representará la i-ésima observación, es decir, cualquier observación, sin referirse a una en particular; cuando sea necesario hacer referencia a una observación en especial, se reemplazará el subíndice $\boldsymbol{i}$ por un número específico. El Ejemplo 1.1, que se ofrece a continuación, permite una mejor comprensión.

Ejemplo 1.1: En una camada de cuyes (*Cavia porcellus)* nacen vivos tres gazapos que pesan 89, 93 y 110 g, respectivamente. $\boldsymbol{Y}$ es la variable peso al nacimiento: $y_1 = 89$ g, $y_2 = 93$ g, $y_3 = 110$ g.

> Esta variable se clasifica como cuantitativa continua, puesto que, aun cuando los valores son números enteros, por definición, podría expresarse con decimales, tomando cualquier valor posible en la recta real. En general, se identifica el conjunto de observaciones como $y_1, y_2, y_3, \ldots, y_i, \ldots, y_n$, donde el sub-índice $\boldsymbol{n}$ representa la última observación y los puntos el resto de observaciones existentes. Aquí, $n = 3$, y_i corresponde a cualquier observación, $y_3 = 110$ g.

Escalas de Medida. Existen cuatro niveles generales de medición: nominal, ordinal, de intervalo y de razón. Estos niveles conducen a cuatro tipos de escalas:

- **Escala nominal.** Los datos obtenidos con escala nominal, implica que sean clasificados en categorías, por nombres o designaciones de las medidas de los fenómenos de interés, por lo que se utiliza para variables cualitativas. Por ejemplo, el sexo y la raza de los bovinos o las especies de microalgas. Si a las categorías se les asigna números que pueden ser 1, 2, 3 y 4, sólo es posible interpretarlos como diferentes, esto es $1 \neq 2 \neq 3 \neq 4$, sin que se pueda afirmar, que una es superior a otra; así una clasificación de las razas bovinas de carne en Hereford, Shorton, Angus y Limousine, identificadas como 1, 2, 3 y 4 no indica superioridad o inferioridad de alguna de ellas con respecto a cualquier atributo estudiado en esas razas.

- **Escala ordinal.** La medida con escala ordinal incluye las propiedades de la escala nominal, pero además las categorías pueden ser ordenadas en el sentido mayor que o menor que, indicando superioridad. Por ejemplo, por resistencia a la mastitis de las vacas Jersey, se podría sugerir un agrupamiento en tres categorías 1, 2 y 3, caso en el que el 3 es el número correspondiente a la mayor resistencia, sin embargo, ésta no ofrece ningún tipo de información

sobre la magnitud de las diferencias entre categorías. Si éstas vienen dadas en términos de alto, medio y bajo y se identificarán como 3, 2 y 1, respectivamente, esto no implica que exista una unidad determinada de diferencia entre cada categoría, sino simplemente que $3 > 2 > 1$.

- **Escala de Intervalo.** Una medida que genera datos con escala de intervalo se realiza cuando se asignan ciertos números a los objetos o acontecimientos estudiados que, además de poseer las características de la escala ordinal, permiten interpretar la diferencia entre dos medidas, es decir, se puede indicar la cantidad en que un evento se diferencia de otro, por lo que se aplica para variables cuantitativas. En esta escala, el punto cero y la unidad de medida son arbitrarios. Las variables con escalas de medición en intervalo más conocidas, son las utilizadas para medir la temperatura. Las escalas Celsius y Fahrenheit, por ejemplo, tienen un punto cero arbitrario, pero esto no indica ausencia de temperatura, puesto que es posible registrar valores por encima o por debajo de ese punto; además, las unidades de medida tienen un tamaño diferente.

- **Escala de Razón.** Es la escala más potente de las cuatro existentes, ya que posee todas las propiedades de las escalas anteriores y un punto cero absoluto que, obviamente, se aplica para variables cuantitativas, por lo tanto, es posible indicar cuántas veces un objeto es más grande que otro y el tamaño de la diferencia. El peso vivo, la alzada, el tiempo de recuperación, el rendimiento en canal, entre otros atributos, se miden en escala de razón.

Según la escala de medición aplicada a las variables estudiadas, éstas se pueden clasificar, como nominales, ordinales, de intervalo o de razón. Es importante que, el investigador identifique la escala de medida de las variables, puesto que, para cada una de ellas, existen procedimientos

estadísticos adecuados, que posibilitan el mejor uso de la información contenida en los valores del atributo.

Población y Muestra. Etimológicamente, el término ***Población*** proviene del latín *populus* (pueblo) y éste del griego *polós* (mucho). En los textos de Estadística, se encuentran definiciones más o menos comunes de población, también llamada Universo. En Estadística, la población se define, como el conjunto de todos los valores posibles, que puede asumir una variable medida en cualquier escala. Estos valores, no necesariamente, serán diferentes entre sí, ni tampoco estarán obligatoriamente representados en número finito. También es aceptado el concepto de población, como el conjunto de elementos o individuos, reunidos bajo una característica común y definidos en el espacio y en el tiempo.

Los conceptos antes mencionados, bastan para establecer con claridad, el significado del término población, sin embargo, dado que en este caso, el objeto de trabajo, en las mayoría de los casos, lo constituyen las poblaciones biológicas, es conveniente analizar la definición expresada por Silva (1982) para una población biológica, en los siguientes términos: *"Sistema constituido por partes, que se integran armónicamente entre sí, para adaptarse a las continuas alteraciones del ambiente en sus múltiples aspectos".*

La anterior definición es válida, tanto para las poblaciones animales como para las vegetales y de microorganismos, sobre las cuales se desarrollan los trabajos de investigación en producción y salud animal.

Las medidas calculadas sobre la población se llaman ***Parámetros*** y el total de elementos que la constituyen, esto es, el tamaño de una población se denota con la letra ***N*** (mayúscula). Si se conocen todos los posibles valores de una variable en la población, teóricamente podría ser estudiada y descrita perfectamente.

En general, ***Muestra*** es toda parte ***representativa*** de una población, lo que significa, que la muestra debe conservar las características esenciales de la población. Desde el punto de vista del Método Científico, la muestra se considera como una parte extraída correctamente de una población, para luego ser observada científicamente, con el objetivo de obtener resultados válidos para la población de donde se obtuvo, esto indica que, además de representar a todos los grupos o categorías de individuos presentes en una población, debe ser obtenida de manera aleatoria, en donde todos tienen la misma posibilidad de ser seleccionados.

Estadísticamente, es factible determinar los límites de error y la probabilidad sobre la validez de los resultados obtenidos con la muestra, cuando éstos se generalizan en toda la población. Teniendo en cuenta que, para disponer de una muestra representativa, es necesario incorporar el Principio de Aleatoriedad y las Leyes de las Probabilidades, con el propósito de evitar sesgos en la selección de las observaciones; en la práctica constituye un proceso de cierta complejidad. Los procedimientos generales para obtener muestras representativas de una población, se indicarán en el Capítulo 4.

Las medidas, obtenidas con datos de la muestra, se llaman ***Estadígrafos*** o ***Estadísticos*** y el número de elementos que constituyen la muestra, se denota por ***n***, que en la gran mayoría de los casos será menor que ***N,*** salvo que la muestra tenga el tamaño de la población, como sucede en ocasiones realmente excepcionales.

A manera de resumen, en concordancia con los principales conceptos básicos expuestos hasta este punto, un análisis de la Figura 1.2 proporciona una visión global de los mismos y una rápida comprensión de los objetivos del texto.

Los ***Parámetros***, convencionalmente, se identifican con letras mayúsculas o con símbolos griegos. Así, el tamaño de la población se identifica con la letra

N, la media se reconoce universalmente como $\boldsymbol{\mu}$, la varianza como $\boldsymbol{\sigma^2}$, la desviación estándar con $\boldsymbol{\sigma}$, el coeficiente de regresión con $\boldsymbol{\beta}$, el coeficiente de correlación $\boldsymbol{\rho}$ y los correspondientes estadígrafos como ***n,*** $\bar{y}$, S^2, *b, r.*

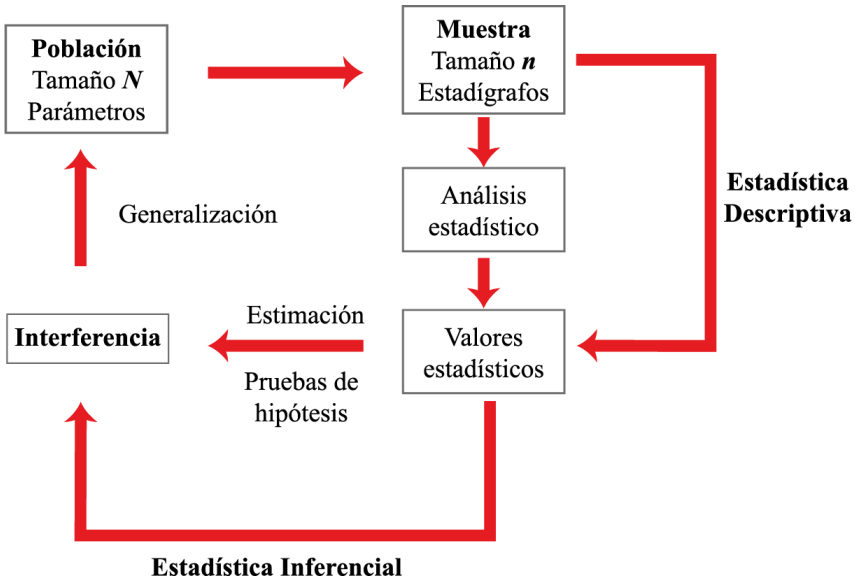


Figura 1.2 Ciclo Inferencial

CAPÍTULO 2

Presentación, resumen y caracterización de la información

La Estadística Descriptiva utiliza dos métodos para presentar y resumir los datos, uno ***gráfico*** y otro ***numérico***, con el fin de caracterizar la información generada en el estudio de una muestra.

Es conveniente comenzar, por agrupar los datos mediante procedimientos, como la construcción de las tablas de distribución de frecuencias y el diseño de gráficos estadísticos, que incluyen histogramas, barras simples, barras compuestas, barras proporcionales, diagramas circulares o pasteles, polígonos de frecuencia, ojivas y gráficos de dispersión, entre los más importantes.

Las ***Tablas de Frecuencia*** y los ***Gráficos*** son de utilidad para interpretar, presentar y describir los datos, pero tienen limitaciones, cuando se quiere establecer comparaciones entre conjuntos de datos. El cálculo de las ***Medidas de Tendencia Central y de Dispersión***, obviamente, son métodos numéricos que contribuyen a solucionar, en parte, estos inconvenientes.

TABLAS DE DISTRIBUCIÓN DE FRECUENCIAS

Para construir una Tabla de Distribución de Frecuencias, es oportuno indicar que, las mismas se pueden elaborar con cualquier tipo de variable y para este proceso, es necesario precisar algunos conceptos de interés.

Se entiende por ***Frecuencia,*** el número de veces que una característica o un valor aparece en un conjunto de observaciones o en un rango de valores y ***Distribución de Frecuencias,*** es el método estadístico para describir el comportamiento de un conjunto de datos; éstos se ubican en categorías o clases y se indica la frecuencia correspondiente en cada una de ellas, esto es, el número de datos incluidos en cada clase.

Para agrupar los datos en ***Categorías,*** es conveniente seguir y establecer una serie de reglas que se resumen a continuación. Sin embargo, es importante indicar que, en la actualidad existe software especializado, para desarrollar estos procedimientos estadísticos en forma rápida, eficaz y confiable. En lo que respecta a la presentación, resumen, caracterización de la información y análisis estadístico, en general, los paquetes estadísticos Stat Graphics, SPSS, Minitab, R Studio y SAS, entre los más conocidos, son herramientas eficientes y otros, como MS-Excel, pueden ser de gran utilidad.

Construcción de Tablas de Distribución de Frecuencias. Los pasos que se describen a continuación, son utilizados en este libro para desarrollar dicho proceso, aclarando que se pueden encontrar diferencias con los propuestos por otros autores, que pueden resultar igualmente aceptados y eficientes.

a. Calcular el Número de Categorías (*NC*). Depende, fundamentalmente, de la cantidad de datos disponibles y del tipo de variable. Con base en estas consideraciones, existen algunas normas para lograr un número óptimo de categorías, como la Regla de Sturges, citada por Milton (1994); la tabla sugerida por Milton (1994), basada justamente en esta regla; el procedimiento propuesto por Carvajal et al (1993):

- Regla de Sturges: $NC = 1 + 3{,}3\ (log\ n)$
- Método sugerido por Carvajal et al: $NC = \sqrt{n}$

- Tabla reportada por Milton, (1994) según el tamaño de la muestra (Tabla 1 del Apéndice).

Es conveniente tener en cuenta, que, cualquiera de estos métodos es válido; escoger muchas clases, puede ocasionar, que algunas de ellas queden sin elementos, originando clases vacías, por el contrario, si se determinan pocas clases, se pierde claridad de la información por exceso de agrupamiento. Se recomienda establecer entre 5 y 10 categorías, aclarando que se constituye en una guía, sin ser un número definitivo, pudiendo variar de acuerdo con la conformación adecuada la tabla. En muestras con alto número de datos es más recomendable la Regla de Sturges, puesto que los otros generan alto número de clases.

b. Calcular el Intervalo de Clase (*IC*). El cálculo del Intervalo de Clase implica seguir estos pasos:

- Establecer el Rango (*R*): *R = valor máximo - valor mínimo.*

- Dividir el Rango entre el número de categorías previamente escogido. En caso que el resultado de esta división no sea entero, puede aproximarse a la unidad más cercana. Para ello se aplica la siguiente fórmula:

  $$IC = \frac{R}{NC}$$

  Con esta operación se encuentra la longitud mínima requerida, para cubrir la amplitud o intervalo de cada clase y, de esa manera, establecer los límites de éstas. Este valor debe tener el mismo número de decimales que los datos.

c. Definir el *Límite inferior* (*Li*) y el *Límite superior* (*Ls*) de cada intervalo. El Límite inferior de la primera clase es el valor mínimo, el

resto de los límites de clase se obtiene mediante la suma sucesiva del Intervalo de Clase a partir del primer valor. Debe tenerse en cuenta que, el dato mayor, no es necesariamente el Límite superior de la última clase, pero debe estar incluido dentro de ésta.

El Límite superior de una clase es el mismo valor del Límite inferior de la siguiente, pero se recomienda que, ninguna observación particular coincida con los valores de los límites, y para esto puede utilizarse varias estrategias que se indican más adelante.

Para definir los límites de cada intervalo, es importante tener en cuenta, que éstos deben ser mutuamente excluyentes, para lo cual, es necesario establecer criterios para evitar la ambigüedad, es decir, que no debe quedar duda, sobre la pertenencia de las observaciones a cada una de las clases, estableciendo intervalos abiertos en el Límite inferior y cerrados en el Límite superior, es decir de la forma ( ].

Un método alternativo para determinar los límites de clase, consiste en establecer el Límite inferior de cada clase, sumando una unidad al Límite superior de la clase anterior, de esta manera, no se generan dudas para el agrupamiento de los datos.

d. Asignar la Frecuencia a Cada Categoría. Para ello, se deben tener en cuenta las siguientes definiciones:

- La ***frecuencia absoluta*** **(f_i),** es el número de datos, perteneciente a una muestra analizada, que se encuentran entre los límites de cada clase.

- La ***frecuencia absoluta acumulada*** **(f_a),** es la suma de todas las frecuencias absolutas anteriores al Límite superior de una clase dada y se obtiene realizando sumas sucesivas de ésta. La frecuencia

absoluta acumulada, correspondiente a la última categoría, coincide con el número total de observaciones.

- La ***frecuencia relativa*** **(F_i),** es el porcentaje de cada frecuencia absoluta, con respecto al total de datos de la población o muestra analizada, es decir, el resultado de dividir el valor de la frecuencia absoluta entre el número total de datos, multiplicado por 100.

- La ***frecuencia relativa acumulada*** **(F_a),** es la suma de todas las frecuencias relativas anteriores al Límite superior de una clase dada, la cual, es el resultante de la suma sucesiva de ésta para las clases anteriores. La frecuencia relativa de la última categoría debe ser igual a 100.

Ejemplo 2.1: Solarte y García (2001) recolectaron varias muestras aleatorias, constituidas por 50 cuyes (*Cavia porcellus*) machos, en la Granja Botana, de propiedad de la Universidad de Nariño, ubicada en el municipio de Pasto, con el fin de analizar la variable peso al destete. Al pesar cada individuo, se obtuvo los siguientes resultados, expresados en gramos:

218	225	205	258	139	257	278	289	258	209
258	230	189	258	209	129	289	297	209	157
188	231	305	288	285	309	218	303	187	208
177	230	285	186	304	258	289	285	287	258
205	143	277	258	308	310	148	308	258	128

Con los anteriores datos, correspondientes a una de las muestras, se construyó una Tabla de Distribución de Frecuencias, cuyo procedimiento se detalla a continuación.

Según la Regla de Sturges: $NC = 1 + 3{,}3\ (log\ 50) = 6{,}61 \approx 7$
Según la Fórmula de Carvajal et al: $NC = \sqrt{50} = 7{,}07 \approx 7$
Según Milton (Tabla 1 del Apéndice): $NC = 6$

Cualquiera de las tres opciones es válida; en este caso se escogió $NC = 6$.

Al desarrollar los pasos descritos anteriormente, se tiene:

- $R = valor\ máximo - valor\ mínimo = 310 - 128 = 182$
- $IC = \frac{R}{NC} = \frac{182}{6} = 30{,}33 \approx 31$
- El Límite inferior de la primera clase, se estableció con la observación menor, para este ejemplo es 128 y, posteriormente, se sumó sucesivamente el valor de *IC* (31), hasta completar todos los límites, de la siguiente manera:

Clases	Límite inferior	Límite superior
1	128	159
2	159	190
3	190	221
4	221	252
5	252	283
6	283	314

- Calcular las frecuencias en cada categoría. Para esto, simplemente se cuentan los datos que caen dentro de cada intervalo.

  Clase uno: 128, 129, 139, 143, 148, 157
  Subtotal: 6 observaciones

Clase dos: 177, 186, 187, 188, 189
Subtotal: 5 observaciones
Clase tres: 205, 205, 208, 209, 209, 209, 218, 218
Subtotal: 8 observaciones.
Clase cuatro: 225, 230, 230, 231
Subtotal: 4 observaciones
Clase cinco: 257, 258, 258, 258, 258, 258, 258, 258, 258, 277, 278
Subtotal: 11 observaciones.
Clase seis: 285, 285, 285, 287, 288, 289, 289, 289, 297, 303, 304, 305, 308, 308, 309, 310
Subtotal: 16 observaciones.

Total: 50 observaciones.

En este caso se tomó intervalos de la forma ( ]; por ejemplo, suponiendo que en la muestra apareciera el número 221, éste se debe clasificar en el intervalo 221-252 y no en el intervalo 190-221.

Utilizando el método alternativo para agrupar este mismo conjunto de datos, la tabla quedaría de la siguiente manera:

Clases	**Límite inferior**	**Límite superior**
1	128	158
2	159	189
3	190	220
4	221	251
5	252	282
6	283	313

En agrupaciones de este tipo, especialmente para variables continuas, es recomendable tener en cuenta los límites reales de clase, definidos como el punto medio entre el Límite superior de una clase y el Límite inferior de la siguiente, así:

Clases	Límite inferior	Límite superior
1	127,5	158,5
2	158,5	189,5
3	189,5	221,5
4	221,5	251,5
5	251,5	282,5
6	282,5	313,5

Se recomienda revisar en detalle la Tabla 2.1, donde se presenta la información completa de los resultados para la tabla de frecuencias en este ejemplo.

Tabla 2.1

Peso al destete de cuyes machos (*Cavia porcellus*) en la Granja Botana de la Universidad de Nariño, Pasto, Colombia (Solarte y García, 2001).

Clase	Límites de clase: Peso al destete (g)	Frecuencia absoluta f_i	Frecuencia absoluta acumulada f_a	Frecuencia relativa F_i	Frecuencia relativa acumulada F_a
1	128-159	6	6	12	12
2	159-190	5	11	10	22
3	190-221	8	19	16	38
4	221-252	4	23	8	46
5	252-283	11	34	22	68
6	283-314	16	50	32	100
Total		50		100	

Para diseñar tablas de este tipo, además de las consideraciones de procedimiento descritas anteriormente, deben tenerse en cuenta los siguientes aspectos:

- Toda la tabla de resultados debe llevar un título, el cual debe ser claro y completo, de tal manera que responda a las preguntas ¿Qué es?, ¿Dónde?, ¿Cuándo? (Carvajal et al, 1993).

- En la primera columna se indica la categoría, en la segunda la variable con su unidad de medida y los límites de clase correspondientes, en el resto de columnas se consignan las respectivas frecuencias.

- Deben diseñarse, según el modelo de la Tabla 2.1, sin líneas verticales y las horizontales sólo aquellas que enmarcan los títulos de columna y los resultados finales de la última fila.

En Biometría, el uso de la Estadística implica tener claro el concepto sobre el método a aplicar y su utilidad en el problema a resolver, al igual que el procedimiento y, fundamentalmente, la interpretación de los resultados, que conlleva a formular las conclusiones. En el caso de las tablas de frecuencia, se pueden generar varias conclusiones, dependiendo del objetivo del análisis o de las características de la población que se desea destacar. En el Ejemplo 2.1, algunas conclusiones, a partir de la información consignada en la Tabla 2.1, pueden ser las siguientes:

- El mayor número de cuyes (16) incluidos en la muestra, se encuentran en el rango de 283-318 g de peso, los cuales representan el 32% de la población.

- El grupo con menor número de *C. porcellus* (6) en la muestra estudiada, tiene un peso entre 128-159 g de peso, que representan el 12% de la población.

- El 38% de los animales, tienen un peso inferior a 221 g de peso, lo cual implica, que esta población tiene un bajo número de individuos pequeños.

GRÁFICOS ESTADÍSTICOS

El Método Gráfico, es otra herramienta de la Estadística Descriptiva, consistente en representar en forma gráfica, mediante figuras geométricas dibujadas a escala, los datos agrupados en las Tablas de Frecuencia, con la intención de proporcionar una visión de fácil comprensión, con referencia a la variable estudiada, sobre aquellas características de la muestra, que no son posibles de observar con la representación numérica, por este motivo, el gráfico debe ser sencillo y claro.

Para diseñar la figura, es necesario construir previamente la respectiva Tabla de Distribución de Frecuencias, por lo tanto, el título del gráfico debe concordar con el de la tabla. En las publicaciones científicas, la mayoría de editores de revistas especializadas, recomiendan no duplicar la información, es decir, si se decide presentar los datos en tablas, no se debe incluir el gráfico para el resumen de los datos, así que el investigador, decidirá sobre la forma que proporcione mayor claridad. En este texto, se presentan gráficos basados en las Tablas de Frecuencia, sólo como una estrategia pedagógica, sin que ello implique, una recomendación persistente.

Entre los gráficos más utilizados en Estadística, se pueden mencionar, los histogramas y polígonos de frecuencia, las ojivas y los diagramas circulares, estos últimos, comúnmente denominados pasteles.

Histograma. Para este tipo de gráfico, se construye un rectángulo para cada categoría, donde la base corresponde al intervalo y la altura a la frecuencia, que puede ser absoluta o relativa. Los datos agrupados de la Tabla 2.1, representados en un Histograma, se pueden apreciar en la Figura 2.1, donde, en el eje *X* se colocaron los intervalos y en el eje *Y*, las frecuencias absolutas.

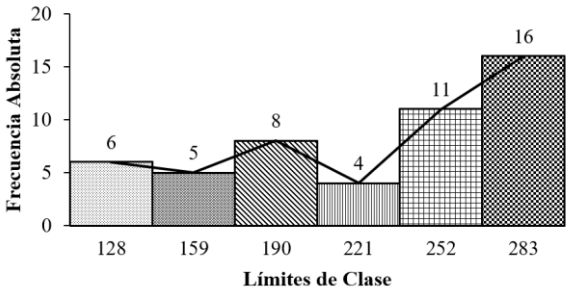


Figura 2.1 Peso al destete de cuyes machos (*Cavia porcellus*) en la Granja Botana de la Universidad de Nariño, Pasto (Solarte y García, 2001).

La interpretación de la información, registrada en los gráficos estadísticos, genera conclusiones similares a las que se formularon para los datos agrupados que los generaron, con la particularidad, que los gráficos muestran características generales de la población, frecuentemente inadvertidos con métodos numéricos. Para la Figura 2.1, una de las conclusiones podría ser:

- Los animales, incluidos en la muestra, tienen un peso heterogéneo, al distribuirse con frecuencias similares en todas las clases, con una leve tendencia a existir mayor número de animales con mayor peso.

Polígono de Frecuencias. Este tipo de gráfico se construye utilizando los valores de frecuencia absoluta, para ello, se unen, mediante líneas rectas, los puntos medios localizados en la parte superior de cada rectángulo del Histograma, también indicado en la Figura 2.1.

La interpretación y las conclusiones surgidas de la observación de un Polígono de Frecuencias, son similares al Histograma, sin embargo, brinda mayor facilidad para identificar la tendencia de los datos; en el caso del Ejemplo 2.1, es más visible la presencia de un mayor número de animales grandes.

Ojiva o Polígono de Frecuencia Acumulada. Esta se diseña, para representar las frecuencias acumuladas absolutas o relativas. Para su construcción, se forman puntos, haciendo coincidir el límite inferior de cada clase con el valor de la frecuencia acumulada respectiva, luego se unen estos puntos con líneas rectas. En el eje *X,* se ubican los intervalos; en el eje *Y,* las frecuencias acumuladas, tal como se aprecia en la Figura 2.2.

La Figura 2.2 corresponde a la ojiva denominada ***"menor que"***, la que debe interpretarse como el número de observaciones inferiores al límite superior de cada clase. Para el Ejemplo 2.1, no hay ninguna observación menor que 128, seis son menores que 159, 11 menores que 190 y así sucesivamente, hasta concluir que, 50 son menores de 314 gramos. Esta gráfica también puede construirse, con la frecuencia relativa acumulada, en cuyo caso la interpretación hace referencia a porcentaje de individuos que se encuentran por debajo del Límite superior de la clase que se está leyendo, así podría decirse, como una de las conclusiones, que el 46% de los animales tienen un peso menor que 252 gramos.

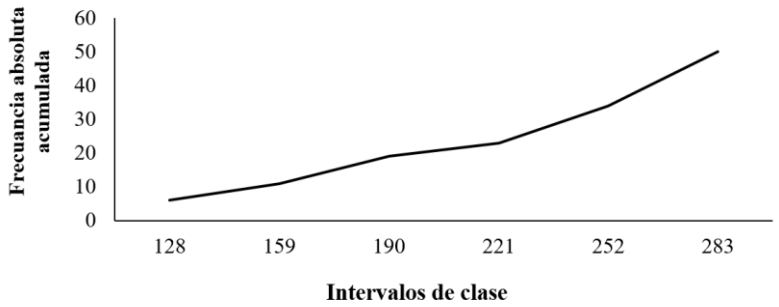


Figura 2.2 Peso al destete de cuyes *machos* (*Cavia porcellus*) en la granja Botana de la Universidad de Nariño, Pasto, Colombia (Solarte y García, 2001), representado en una ojiva.

Diagrama Circular. También conocido, en el lenguaje coloquial, como torta o pastel, el cual se construye, con la frecuencia relativa, para mostrar la proporción gráfica que ocupa el número de elementos que pertenecen a cada una de las clases; este tipo de gráfica permite mostrar, de manera clara, tales valores en una visión global.

En la Figura 2.3 se puede apreciar los valores de la frecuencia relativa, calculados en la Tabla 2.1, representados en un diagrama circular, donde las divisiones distinguibles del círculo, corresponden cada una de las categorías y es necesario establecer convenciones para identificar cada una de ellas.

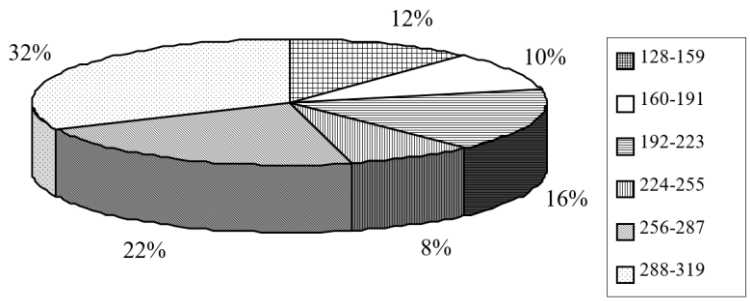


Figura 2.3 Peso al destete de cuyes machos (*Cavia porcellus*) en la Granja Botana de la Universidad de Nariño, Pasto, Colombia (Solarte y García, 2001), representado en un Diagrama Circular.

A partir de un Diagrama Circular, es posible interpretar y concluir a cerca de los valores de frecuencia relativa da cada una de las clases, por ejemplo, es claro que el 22% de los cuyes en estudio tienen un peso entre 256 y 287 gramos. Sin embargo, lo más importante, es la percepción global que se pueda tener del total de la muestra estudiada, como que los animales con mayor peso representan más de la mitad de la población (54%).

A manera de resumen, con respecto a la presentación de la información en forma gráfica, se deben tener en cuenta las siguientes recomendaciones:

a. Cuando la variable analizada es continua, es preferible utilizar el Histograma, el Polígono o el Gráfico de Dispersión.

b. Para variables discretas, se recomienda realizar la representación con Histogramas o con Diagramas Circulares.

c. En el Ejemplo 2.1, los datos fueron representados mediante un Diagrama Circular, sólo como ilustración didáctica.

d. Todos los gráficos deben llevar un título, identificando con la palabra "Figura" y un número de secuencia; sin embargo, no es necesario aclarar el tipo de figura. En el ejemplo se hizo la aclaración, exclusivamente con fines didácticos.

Curvas de Frecuencia. En Estadística, los datos que se procesan, por lo general, pertenecen a muestras aleatorias extraídas de una población, por lo tanto, el Histograma y el Polígono de Frecuencias, describen las variaciones que experimenta una variable o característica en dicha muestra, las cuales pueden utilizarse, para formular conclusiones inferenciales, siempre que los análisis se hayan realizado sobre una muestra adecuada. Si se aumenta el número de observaciones y al mismo tiempo se disminuye la amplitud de los intervalos, el Polígono de Frecuencias estará formado por pequeños segmentos que se aproximan a una curva, tal como se aprecia en la Figura 2.4. Este tipo de curva ideal recibe el nombre de ***Curva de Frecuencia***.

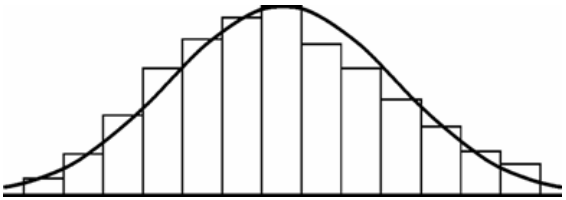


Figura 2.4. Representación de una Curva de Distribución de Frecuencias.

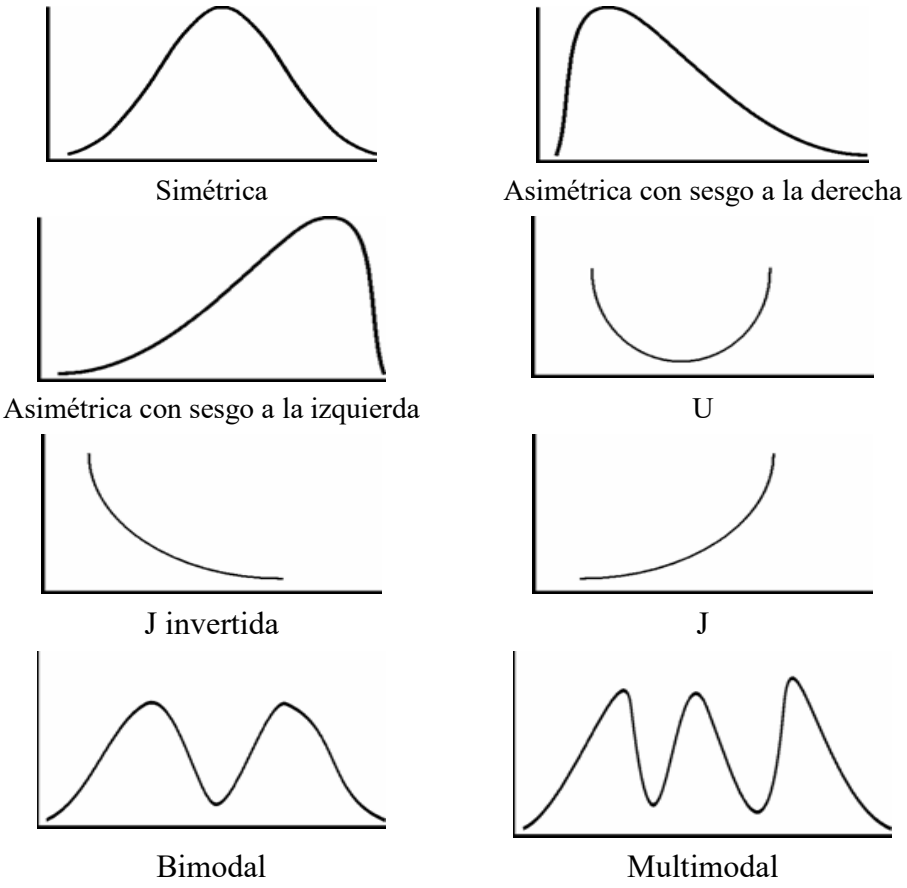


Figura 2.5 Formas de Curvas de Distribución de Frecuencia más comunes en Estadística.

Las Curvas de Frecuencia, son de gran utilidad en el análisis estadístico, al proporcionar métodos simplificados, para describir características básicas de las poblaciones, así como en la inferencia estadística, para conocer la distribución de la variable en la población de donde se obtuvo la muestra. En la Figura 2.5 están representadas las formas de curvas más comunes.

La ***Curva Simétrica*** se denomina ***Distribución Normal*** o ***Curva Normal***; en ella, las frecuencias de clase son mayores en el centro y disminuyen hacia los extremos de manera similar. Esta curva tiene gran importancia en Estadística, puesto que proporciona, una adecuada representación de las distribuciones de una gran cantidad de variables, tales como, los datos meteorológicos de temperatura y precipitación pluvial, mediciones efectuadas en organismos vivos, entre las que pueden mencionarse, el peso y la talla, calificaciones en pruebas de aptitud, rendimiento en cultivos y muchas otras.

Las ***Curvas Asimétricas***, con sesgo a la derecha o a la izquierda, son modelos de distribuciones asimétricas, bastante comunes en datos económicos y comerciales.

La ***Curva J***, llamada así por su forma, es frecuente en muchos estudios biométricos, sobre todo en aquellos, que tienen que ver con el crecimiento de organismos vivos o de crecimientos poblacionales en condiciones naturales.

La forma de las curvas, también proporciona información importante, por ejemplo, en el caso de las *bimodales*, la aparición de dos modas, puede implicar un cambio en el proceso durante el estudio de un fenómeno particular o que los datos provienen de dos fuentes diferentes.

MEDIDAS DE TENDENCIA CENTRAL, POSICIÓN Y DISPERSIÓN

Se mencionó anteriormente, que los métodos para presentar la información en forma gráfica o en Tablas de Frecuencia, son útiles, cuando se trata de ofrecer al lector, una visión general de los resultados obtenidos en el estudio

de una muestra, pero no son tan eficientes, cuando se tratan de establecer, las características más importantes de los datos en lo referente a su localización y variabilidad.

Los métodos numéricos permiten extender los análisis y posibilitan comprender, de modo más claro, la naturaleza de la distribución de cualquier variable de interés en Producción o Salud Animal. Estos métodos se orientan, básicamente, a describir el centro de la distribución de las observaciones y la forma como éstas varían, en torno a ese valor central. Dicho de otra manera, permiten calcular las ***Medidas de Tendencia Central, Medidas de Posición*** y las ***Medidas de Espaciamiento o Dispersión*** de los datos.

Medidas de Tendencia Central

Las más utilizadas son, la *Media o Promedio, la Mediana y la Moda o Modo*. A continuación, se hará referencia a cada una de ellas, con las correspondientes formas de cálculo, con datos agrupados y sin agrupar.

La Media o Promedio Aritmético. Es la Medida de Tendencia Central más utilizada; se representa con la letra griega μ (mu) para la población; se utiliza el símbolo $\bar{y}$ (y barra) para representar la medida tomada en la muestra, de tal manera que μ es el parámetro, cantidad fija y desconocida, mientras $\bar{y}$ es el estadígrafo, el cual puede tomar diferentes valores, al calcularse con diversas muestras obtenidas de la misma población objeto de estudio. La Media Aritmética se expresa en las mismas unidades de la variable original, sean litros, gramos, kilogramos, hectáreas, etc.

Conviene aclarar, que existen otros tipos de medias diferentes a la aritmética, tal como se verá posteriormente, pero el uso generalizado de la palabra Media

o Promedio ha hecho que se tome este término para referirse al Promedio Aritmético, sin indicarlo expresamente.

La Fórmula de Cálculo de la Media Aritmética, es la siguiente:

$$\mu = \frac{\sum_{i=1}^{N} y_i}{N} \qquad (2.1)$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad (2.2)$$

Donde:

$\sum_{i=1}^{N} y_i$ = Sumatoria de los y_i, desde la primera hasta la última observación.
y_i = Las observaciones.
N y n = tamaño de la población y muestra, respectivamente.

Aunque no existen normas precisas para calcular los decimales del valor de la Media, algunos autores recomiendan utilizar un decimal más que los datos originales.

Además, es importante tener en cuenta las siguientes características de esta medida de localización:

- La Media, no se encuentra dentro del conjunto de datos analizado; es un valor que se calcula con todos los datos.
- El Promedio Aritmético, es calculado en *forma rígida*, por lo tanto, implica que existe un sólo promedio para cada conjunto de datos.
- El Promedio, es sensible a valores extremos.
- Por estar definido algebraicamente, el Promedio permite tratamiento algebraico.

- La sumatoria de las desviaciones con respecto a la Media, es igual a cero, esto es:

$$\sum(y_i - \bar{y}) = 0$$

- Es un estadígrafo que se utiliza con datos, cuya escala es de intervalo o razón.

- El Promedio Aritmético, puede ser empleado en el análisis de distribuciones simétricas.

- El Promedio es la base de la inferencia estadística.

Ejemplo 2.2: Las observaciones que se indican a continuación, corresponden a la longitud estándar (en cm), de una muestra de 10 ejemplares de bagre del Patía (*Rhamdia quelem*), según reporte de Ortega y Rodríguez (2004).

Ejemplares	1	2	3	4	5	6	7	8	9	10
Talla (cm)	35,0	27,0	34,5	34,9	34,0	32,0	22,0	37,0	26,0	33,0

De acuerdo con la fórmula (2.3.2), la Media se calcula así:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$= \frac{35{,}0 + 27{,}0 + 34{,}5 + 34{,}9 + 34{,}0 + 32{,}0 + 37{,}0 + 26{,}0 + 33{,}0}{10}$$

$$= \frac{315{,}4}{10} = 31{,}54\ cm.$$

Como se observa, el valor de la Media, se calculó con un decimal más que los datos originales.

> Este resultado se interpreta como la Media Aritmética de la muestra de 10 datos y se puede concluir, que en promedio los peces evaluados miden 31,54 cm de longitud estándar.

Promedio Ponderado. En algunas situaciones, los valores de la variable, están afectados por una frecuencia particular o peso relativo (*w*). En estos casos, no es factible calcular una Media Aritmética, por cuanto, su valor resultaría distorsionado, siendo necesario tener en cuenta la variable de ponderación para su estimación; en consecuencia, para calcular la Media, se deben multiplicar estos valores de ponderación por los correspondientes de la variable y luego dividir este resultado entre la sumatoria de dichos pesos. La Media obtenida de este modo, se denomina ***Media*** o ***Promedio Aritmético Ponderado*** y se le llama corrientemente, ***Promedio Ponderado*** ($\bar{y}_w$). La fórmula de cálculo, es la siguiente:

$$\bar{y}_w = \frac{\sum y_i w_i}{\sum w_i} \qquad (2.3)$$

Donde: $\bar{y}_w$ es el Promedio Ponderado; y_i corresponde a cada uno de los valores de la variable; w_i es el peso relativo de cada valor de la variable.

Ejemplo 2.3: En una planta procesadora de leche, se quiere estimar el costo promedio mensual de los salarios devengado por los operarios, pero debe considerarse, que no todos los obreros laboran en las mismas condiciones, puesto que en esta factoría, por la asignación de tareas específicas y especializadas, 14 operarios laboran jornada completa con 8 horas extras en el mes, 5 operarios laboran jornada completa con 4 horas extras en el mes, 10 operarios trabajan la jornada completa sin ninguna hora extra en el mes y 6 operarios trabajan medio tiempo, sin horas extras.

La asignación mensual es la siguiente: con 8 horas extras $ 1.900.000, con 4 horas extras $ 1.600.000, sin horas extras $ 1.200.000 y $ 600.000 por medio tiempo sin horas extras. Observe que el costo de los salarios, tiene un peso relativo representado por el número de trabajadores con diferente número de horas extras, sin horas extras y de medio tiempo, así que el valor promedio del costo del salario debe ponderarse así:

Jornada	**Costo (y_i)**	**No. de operarios (w_i)**	$y_i\, w_i$
Con ocho horas extras	1.900.000	20	38.000.000
Con 4 horas extras	1.600.000	8	12.800.000
Sin horas extras	1.200.000	10	12.000.000
Medio tiempo	600.000	2	1.200.000
Sumatoria		**40**	**64.000.000**

Con estos valores, el Promedio Ponderado, se calcula según la fórmula (2.3.3).

$$\bar{y}_w = \frac{\sum y_i w_i}{\sum w_i} = \frac{64.000.000}{40} = \$\ 1.600.000$$

Este resultado lleva a concluir que, el promedio del costo salarial en esta factoría, ponderado según el número de operarios, es de $ 1.600.000.

A este miso resultado se puede llegar, tomando como criterio de ponderación (w_i) el porcentaje de operarios en cada categoría, en lugar del número de operarios en cada categoría, como se indica a continuación:

Jornada	**Costo (y_i)**	**Participación porcentual operarios (w_i)**	$y_i\, w_i$
Con ocho horas extras	1.900.000	0,5 (50%)	950.000
Con 4 horas extras	1.600.000	0,2 (20%)	320.000
Sin horas extras	1.200.000	0,25 (25%)	300.000
Medio tiempo	600.000	0,05 (5%)	30.000
Sumatoria		**1 (100%)**	**1.600.000**

$$\bar{y}_w = \frac{\sum y_i w_i}{\sum w_i} = \frac{1.600.000}{1} = \$\ 1.600.000$$

Media Geométrica. En otros casos se presentan ***relaciones de cambio*** que afectan a la variable estudiada, como sucede en el crecimiento poblacional de algunos seres vivos. En estas circunstancias lo más aconsejable, es calcular la ***Media Geométrica*** (*G*) que permite asignar ponderaciones iguales a dichas relaciones.

Por definición, la Media Geométrica es la raíz n-ésima del producto de los valores de la variable, siempre y cuando estos valores sean positivos. Para su cálculo, basta con aplicar la siguiente fórmula:

$$G = \sqrt[n]{(Y_1)(Y_2)\ldots..(Y_n)} \qquad (2.4)$$

Ejemplo 2.4: Se quiere medir el promedio de crecimiento, a través del número de colonias de la bacteria *E. coli* en un medio de cultivo, durante un período de tres horas. En este caso, lo más recomendable es calcular *G*, dado que existe una relación entre el tiempo, medido en horas y el número de colonias bacterianas que crecen en el medio, así:

Horas de cultivo	Número de colonias
1	2
2	4
3	5

Aplicando (2.3.4) se tiene: $G = \sqrt[3]{(2)(4)(5)} = 3{,}42$ colonias/hora

Media Armónica. Existen otros tipos de medias, entre las cuales se cuenta la ***Media Armónica* (*H*).** La primera tiene algunas aplicaciones prácticas en genética cuantitativa y en ciertas comparaciones de medias en el Diseño Experimental; la segunda, principalmente en Física.

La Media Armónica, se calcula así:

$$H = \frac{1}{\frac{1}{n}\sum\frac{1}{y_i}} = \frac{n}{\sum\frac{1}{y_i}} \qquad (2.5)$$

También se puede expresar la fórmula de cálculo, mediante los inversos, así:

$$\frac{1}{H} = \frac{\sum\frac{1}{y_i}}{n} = \frac{1}{n}\sum\frac{1}{y_i} \qquad (2.6)$$

Ejemplo 2.5: Para el conjunto de observaciones 5, 7, 9, 6, 8, la Media Armónica, se obtiene de la siguiente manera:

$$H = \frac{5}{\frac{1}{5}+\frac{1}{7}+\frac{1}{9}+\frac{1}{6}+\frac{1}{8}} = \frac{5}{0{,}745634} = 6{,}70$$

Con los inversos:

$\frac{1}{H} = \frac{1}{5}$ $0{,}04534 = 0{,}149126$, así que $H = \frac{1}{0{,}149126} = 6{,}70$

De manera similar a los estadígrafos anteriores, se concluye, que la Media Armónica de este grupo de datos, es de 6,70. Si el grupo de datos corresponde a una variable específica, se debe adicionar la unidad de medida.

Media Cuadrática. Para el caso de la *MQ*, se calcula utilizando la siguiente fórmula:

$$MQ = \sqrt{\frac{\sum {Y_i}^2}{N}} \qquad (2.7)$$

Ejemplo 2.6: Con el mismo conjunto de datos utilizado para el cálculo de la Media Armónica del Ejemplo 2.5, utilizando la fórmula (2.3.7) sería:

$$MQ = \sqrt{\frac{5^2 + 7^2 + 9^2 + 6^2 + 8^2}{5}} = \sqrt{51} = 7{,}14$$

La Media Cuadrática en el grupo de datos analizado, alcanza un valor de 7,14.

Cálculo de la Media Aritmética con datos agrupados. Hasta ahora se han indicado los procedimientos de cálculo, cuando los datos se encuentran ***sin agrupar;*** sin embargo, si éstos se presentan ***agrupados en Tablas de Distribución de Frecuencias***, es decir, organizados en categorías y no se dispone de la información sin agrupar, es posible tener ***una aproximación*** de las Medidas de Tendencia Central, con procedimientos particulares, que se indican a continuación.

Bajo estas circunstancias, simplemente se aplica la siguiente fórmula:

$$\bar{y} = \frac{\sum f_i m_i}{n} \qquad (2.8)$$

Donde:

f_i = Frecuencia absoluta de la i-ésima categoría.

m_i = Marca de clase o valor central de cada intervalo, que se obtiene, calculando el Promedio aritmético de los Límites superior e inferior de cada clase.

Ejemplo 2.7. Para ilustrar la forma de cálculo, se utilizarán los datos de la Tabla 2.1, con los cuales se ha construido la Tabla 2.2, con el fin de facilitar los cálculos posteriores.

Tabla 2.1

Peso al destete de cuyes machos (*Cavia porcellus*) en la Granja Botana de la Universidad de Nariño, Pasto, Colombia (Solarte y García, 2001). Cálculo con datos agrupados.

Clase	Frecuencia absoluta (f_i)	Marca de clase (m_i)	$f_i m_i$
1	6	143,5	861,0
2	5	174,5	872,5
3	8	205,5	1.644,0
4	4	236,5	946,0
5	11	267,5	294,5
6	16	298,5	4.776,0
Suma	**50**		**12.042,0**

Al aplicar la fórmula (2.3.8) se tiene: $\bar{y} = \frac{12.042}{50} = 240{,}84$ g.

Similar al caso, en el cual, la Media Aritmética se calcula con datos sin agrupar; la conclusión podría expresarse así: el peso promedio de los cuyes al momento del destete, en la Granja Botana, alcanza un valor de 240,84 g.

La Mediana (M_e). Es otra Medida de Tendencia Central, existente dentro del conjunto de datos analizado, que tiene como propiedad, dividir los valores de un conjunto ordenado de datos en dos fracciones, de tal manera que el 50% de las observaciones se sitúan por encima y el 50% restante por debajo de ella. Es decir, la Mediana se define, como el número situado en la mitad de un conjunto de datos *ordenados de menor a mayor.*

Cuando ***n*** es impar, la Mediana es el número situado en la mitad del conjunto, pero cuando ***n*** es par, la Mediana se obtiene, calculando la Media Aritmética, con las dos observaciones situadas en la mitad. Si ***n*** es pequeño, es fácil detectar el centro del conjunto de observaciones, pero a medida que ***n*** aumenta se hace más difícil. Para evitar confusiones, se aplica una fórmula que permite detectar dicho valor central, en un conjunto, con un número de elementos par o impar y ubicar la ***Situación Mediana (SM),*** que indica el sitio en que se encuentra la Mediana, la cual expresa de esta forma:

$$SM = \frac{n}{2} + \frac{1}{2} \qquad (2.9)$$

Ejemplo 2.8: El siguiente conjunto impar de datos, corresponde a valores de la Habilidad de Transmisión Predecible (PTA) para la proteína de la leche (en lb) en reproductores bovinos de la raza Holstein. La PTA es un estimativo del rendimiento medio de la progenie por efecto del potencial genético heredado del padre. Las cifras en mención, se obtuvieron de un catálogo de la Cooperative

Resources International (CRI, 2000): 78, 81, 77, 56, 69, 41, 54.

Para calcular la Mediana, primero deben ordenarse los datos de menor a mayor, de esta manera: 41, 54, 56, 69**,** 77, 78, 81.

A simple vista se puede detectar que, M_e = 69, ya que por debajo de 69 están los valores 41, 54 y 56, mientras que por encima están 77, 78 y 81. En otras palabras, el 50% de las cifras son menores que 69 y el otro 50% corresponde a valores mayores a esa cantidad.

Al aplicar (2.9) se tiene: $SM = \frac{7}{2} + \frac{1}{2} = 4$

Es decir que la M_e, es la cuarta observación ordenada, o sea 69 lb.

Ejemplo 2.9: Con un conjunto par de datos, de la misma fuente y para la misma variable, se tiene: 85, 78, 55, 69, 82, 70, 63, 84, que ordenado de menor a mayor queda: 55, 63, 69, 70, 78, 82, 84, 85. La M_e será el promedio de los dos valores centrales, por lo tanto:

Al aplicar la fórmula (2.9) se tiene: $SM = \frac{8}{2} + \frac{1}{2} = 4{,}5$

Al tener un valor de SM con decimal de 0,5, indica que corresponde a un conjunto par de datos y se debe calcular la Media de los dos datos centrales, en este caso los que se encuentran en las posiciones 4 y 5.

Por lo tanto: $M_e = \frac{70+78}{2} = 74$

La cuarta observación es 70 y la quinta es 78; el promedio de estos dos valores es 74, por lo tanto, el 50% de los reproductores Holstein, tienen un valor de PTA inferior a 74 libras y el 50% restante es superior a la misma cantidad.

Como se observa, a diferencia del Promedio, la Mediana no se afecta por valores extremos de la variable, es decir, por datos que están muy distantes del resto de observaciones, conocidos con el nombre de *"Outliers"* o *"Datos Aberrantes"* (Milton, 1994). Los "Outlieres" se presentan en un conjunto de datos, cuando realmente se presentan valores inusualmente grandes o pequeños o porque existen errores de medición o registro en la base de datos. Si la situación que se presenta es la primera, el investigador puede optar por efectuar los cálculos con o sin los datos aberrantes y hacer la aclaración respectiva.

Cálculo de la Mediana con datos agrupados. Cuando únicamente se dispone de datos organizados en Tablas de Frecuencias, para calcular la M_e, se requiere aplicar la siguiente fórmula de trabajo:

$$M_e = L_i + \frac{\frac{n}{2} - F_{i-1}}{f_i}(IC) \qquad (2.10)$$

Donde:

L_i = Límite inferior de la *Clase Mediana (CM).* Se entiende por Clase Mediana, aquella cuya frecuencia absoluta acumulada, contiene al 50% de los datos.

$\frac{n}{2}$ = Valor que define la Clase Mediana.

$F_{i\text{-}1}$ = Frecuencia absoluta acumulada de la clase anterior a la Clase Mediana.

f_i = Frecuencia absoluta de la Clase Mediana.

IC = Intervalo de clase.

La Mediana tiene las siguientes propiedades:

- La Mediana es un valor existente dentro del conjunto de datos y para su cálculo no se tienen en cuenta los extremos, por lo tanto, no está influenciada por éstos.

- La Mediana se utiliza con datos cuya escala es de razón, intervalo u ordinal.

- Es aconsejable utilizarla en distribuciones notablemente asimétricas, cuyos extremos no están bien definidos

Ejemplo 2:10: Con los mismos datos de la Tabla 2.1, que se vuelve a presentar por comodidad para el lector, al desarrollar la fórmula de cálculo.

Clase	**Límites de clase: Peso al destete (g)**	**Frecuencia absoluta** f_i	**Frecuencia absoluta acumulada** f_a	**Frecuencia relativa** F_i	**Frecuencia relativa acumulada** F_a
1	128-159	6	6	12	12
2	159-190	5	11	10	22
3	190-221	8	19	16	38
4	221-252	4	23	8	46
5	252-283	11	34	22	68
6	283-314	16	50	32	100
Total		**50**		**100**	

Con base en esta información, se obtiene las siguientes cifras:

$$n = 50 \qquad \frac{n}{2} = \frac{50}{2} = 25$$

Se puede observar que, 25 está contenido en la frecuencia absoluta acumulada de la clase 5, por ende, ésta es la Clase Mediana, cuyos límites de clase son 252-283, así que, el límite inferior de la Clase Mediana (L_i) es 252.

La frecuencia absoluta (f_i) de la Clase Mediana es 11.

La frecuencia absoluta acumulada ($F_{i\text{-}1} = F_{5\text{-}1} = F_4$) de la clase 4 (anterior a la clase 5 o Clase Mediana) es 23.

El tamaño del Intervalo de Clase (IC) es 31.

Al aplicar la fórmula (2.3.10) se tiene:

$$M_e = 252 + \frac{\frac{50}{2} - 23}{11}(31) = 257{,}63$$

Del anterior resultado se puede concluir, que el 50% de los animales tienen un peso inferior a 257,63 g y el otro 20% está por encima de este valor.

La Moda o Modo (M_o). Es el dato que más se repite en un conjunto de observaciones. En algunos casos, ningún valor se repite más que otro y se concluye, que la distribución no tiene Moda; cuando hay un solo valor que se repite más que los otros, la distribución es ***unimodal***; si dos datos se repiten igual número de veces, es ***bimodal*** y, en general, cuando existen más de dos valores que se repiten el mismo número de veces, la distribución es ***plurimodal***.

Ejemplo 2.11: A continuación, se desarrollan algunos ejemplos sobre el cálculo de la Moda.

a) Los siguientes datos son una muestra aleatoria del peso vivo individual a las 12 semanas, medido en gramos, de 14 cuyes (*Cavia porcellus*), recolectados por García y Solarte, (2001) en la Granja Botana de la Universidad de Nariño: 984, 967, 768, 872, 849, 1022, 756, 842, 942, 780, 647, 765, 770, 810, ya que en esta muestra, ningún dato se repite más que otro, se concluye que el conjunto de datos no tiene moda.

b) En otra muestra aleatoria, recolectada por los mismos autores, se registraron los siguientes datos: 984, 900, 768, 872, 849, 1022, 756, 842, 900, 780, 647, 765, 900, 810. El dato que corresponde a 900 g se repite tres veces, por lo tanto, la Moda es 900.

c) De esa misma población, se obtuvo una muestra para analizar el tamaño de camada al nacimiento, en la cual se encontraron los siguientes resultados: 2, 2, 3, 3, 3, 5, 4, 4, 4, 4, 5, 5, 3, 2, 2, 6, 4, 6, 3, 3, 6, 5, 4, 5. En esta muestra, los tamaños de camada 4 y 3 se repiten seis veces, así que hay dos Modas: $M_{o1} = 4$; $M_{o2} = 3$.

d) Una muestra de mayor tamaño para la misma variable, condujo a los siguientes resultados: 2, 1, 2, 3, 7, 3, 3, 5, 4, 5, 4, 4, 4, 5, 5, 3, 2, 7, 2, 6, 4, 6, 3, 3, 6, 5, 4, 1, 5. En este caso, la distribución es plurimodal (trimodal) y las modas son 3, 4 y 6, ya que estos tamaños de camada aparecen seis veces cada uno.

Cálculo de la Moda con datos agrupados. Para calcular la Moda con datos agrupados, se debe utilizar la siguiente fórmula:

$$M_o = L_i + \left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right)(IC) \qquad (2.11)$$

Donde:

L_i = Límite inferior de la Clase Modal. Se entiende por Clase Modal la de mayor frecuencia absoluta.

Δ_1 = Valor resultante de la diferencia entre la frecuencia absoluta de la Clase Modal menos la frecuencia absoluta de la clase inmediatamente anterior.

Δ_2 = Valor resultante de la diferencia entre la frecuencia absoluta de la Clase Modal menos la frecuencia absoluta de la clase inmediatamente siguiente.

IC = Intervalo de Clase.

Ejemplo 2.12: Para los datos de la Tabla 2.1, según (2.11), la Moda se calcula de la siguiente manera:

L_i = 283 es el límite inferior de la clase 6, definida como la Clase Modal, por tener la mayor frecuencia absoluta.

Δ_1 = 16 - 11 = 5

Δ_2 = 16 - 0 = 16

IC = 31

Al calcular con (2.3.11) se tiene:

$$M_o = 283 + \left(\frac{5}{5+16}\right)(31) = 290{,}38$$

> De acuerdo con el anterior resultado, el dato que más se repite es 290,38. Este es un valor aproximado de la Moda, que podría interpretarse, como el valor más frecuente en la población de donde se extrajo esta muestra.

La Moda se puede utilizar con datos cuya escala es de razón, intervalo, ordinal o nominal. Es aconsejable utilizarla con distribuciones marcadamente asimétricas.

Medidas de Posición

Dentro de estas medidas se incluyen los Cuartiles, los Deciles y los Percentiles.

Cuartiles. Como se mencionó antes, la M_e es una medida que indica el punto donde el conjunto de datos se divide en dos fracciones, cada una de las cuales contiene el 50% de los datos. De la misma manera, es posible encontrar otras medidas, denominadas ***Cuartiles***, que indican la división en fracciones del conjunto que representan el 25%, el 50%, o el 75% de los datos.

El primer cuartil (Q_1), o cuartil inferior, deja el 25% de los datos por debajo de su valor; el segundo (Q_2) o cuartil medio es igual a la Mediana, ya que deja el 50% de los datos por debajo de su valor; el tercero (Q_3), o cuartil superior, deja el 75% de los datos por debajo de su valor.

De forma análoga que, en la Mediana, se puede encontrar la *Situación de los Cuartiles*, mediante la fórmula que se indica en (2.12):

$$SQ_1 = \frac{1}{4}(n+1) \qquad (2.12)$$

Si al aplicar la fórmula (2.12) el resultado no es un número entero, SQ_1 se aproxima al entero más cercano. Cuando SQ_1 cae entre dos enteros, es decir, en media unidad, entonces se redondea hacia arriba.

Para SQ_3 se aplica la siguiente fórmula:

$$SQ_3 = \frac{3}{4}(n+1) \qquad (2.13)$$

De igual manera, si SQ_3 no es un número entero, se redondea al entero más cercano, pero si cae entre dos enteros, se redondea hacia abajo.

Ejemplo 2.13: Los siguientes datos corresponden a la producción de leche en kg por lactancia (260 días) de 60 ovejas de la raza Milchschaf, basados en la información de Larrosa y Borean (1990).

380, 410, 410, 410, 430, 440, 445, 450, 453, 455, 458, 460, 460, 460, 464, 466, 470, 475, 480, 484, 485, 487, 490, 490, 490, 493, 493, 495, 495, 497, 497, 499, 499, 502, 502, 502, 505, 505, 505, 510, 510, 515, 520, 520, 530, 540, 540, 545, 548, 548, 550, 550, 550, 550, 555, 555, 560, 562, 563, 565.

Como se puede observar, el conjunto fue ordenado en forma ascendente para determinar los Cuartiles Q_1 y Q_3 mediante el uso de las fórmulas (2.12) y (2.13), respectivamente; para el primer Cuartil, de la siguiente manera:

$$SQ_1 = \frac{1}{4}(60+1) = \frac{1}{4}(61) = 15{,}25 \cong 15$$

Puesto que SQ_1 no es un número entero, se aproxima al entero más cercano, en este caso a 15, lo que indica que Q_1 es 464, cifra de

producción ubicada en la posición 15, por lo tanto, se concluye, que el 25% de los datos de producción lechera de las ovejas, está por debajo de 464 kg.; también se podría concluir, que el 75% de las ovejas producen una cantidad de leche por encima de ese valor.

Para el tercer Cuartil, se calcula como sigue:

$$SQ_3 = \frac{3}{4}(60+1) = \frac{3}{4}(61) = 45{,}75 \cong 46$$

Teniendo en cuenta que SQ_3 no es un número entero, se redondea al entero más cercano, en este caso a 46, así que Q_3 corresponde a 540, es decir que el 75% de los animales tienen producciones por debajo de este valor, o el 25% por encima de esta cantidad.

Ejemplo 2.14: Según reportes de los mismos autores del Ejemplo 2.13, la producción de leche en kg por lactancia (150 días) de un rebaño conformado por ovejas de la raza Manchega, fue la siguiente:

98, 98, 100, 103, 103, 103, 105, 105, 108, 108, 108, 110, 110, 110, 110, 110, 111, 112, 113, 113, 113, 114, 114, 115, 116, 116, 118, 118, 120, 123, 123, 123, 125, 125, 126, 128, 130, 130, 130, 131, 131, 132, 132, 133, 134, 135, 135, 135, 135, 135, 136, 137, 137, 138, 139, 139, 141, 141, 142, 142, 144, 146, 146, 146, 148, 149, 149, 149, 152.

Igual que en el ejemplo inmediatamente anterior, los datos fueron ordenados en forma ascendente, con el fin de establecer los cuartiles Q_1 y Q_3, al aplicar las fórmulas (2.12) y (2.13) se tiene:

$$SQ_1 = \frac{1}{4}(69+1) = \frac{1}{4}(70) = 17{,}5 \cong 18$$

Puesto que SQ_1 está entre dos números enteros (17 y 18), se aproxima al entero mayor, en este caso a 18, lo que indica que Q_1 es 112 kg., de tal modo que el 25% de los ovinos tienen producciones inferiores a esta cantidad.

$$SQ_3 = \frac{3}{4}(69+1) = \frac{3}{4}(70) = 52{,}5 \cong 52$$

Como se ve, SQ_3 se encuentra entre dos números enteros (52 y 53), así que debe aproximarse al entero menor, en este caso a 52, para concluir que Q_3 es igual a 137 kg, lo que significa que, el 75% de los animales tienen producciones por debajo de este valor.

Cálculo de Cuartiles con datos agrupados. Cuando los datos están agrupados, los Cuartiles pueden ser calculados mediante la fórmula:

$$Q_k = L_i + \frac{\frac{kn}{4} - f_{a-1}}{f_i}(IC); k = 1,2,3 \qquad (2.14)$$

Donde:

L_i = Límite inferior de la clase cuartílica, es decir, donde está el Cuartil.
n = Número de datos incluidos en la muestra.
$f_{a\text{-}1}$ = Frecuencia absoluta acumulada de la clase anterior a la cuartílica.
f_i = Frecuencia absoluta de la clase cuartílica.
IC = Intervalo de Clase.

Ejemplo 2.15: A continuación, se presentan los datos del Ejemplo 2.14, agrupados en una Tabla de Frecuencias.

Tabla 2.2

Producción de leche en kg por lactancia (150 días) de un rebaño conformado por ovejas de la raza Manchega (Larrosa y Borean, 1990).

Clase	Límites de clase		Marca de clase	f_i	F_i	f_a	F_a
1	98,0	107,0	102,5	8	0,1159	8	0,1159
2	107,0	116,0	111,5	18	0,2609	26	0,3768
3	116,0	125,0	120,5	8	0,1159	34	0,4928
4	125,0	134,0	129,5	11	0,1594	45	0,6522
5	134,0	143,0	138,5	15	0,2174	60	0,8696
6	143,0	152,0	147,5	9	0,1304	69	1,0000

En caso de disponer únicamente de la información consignada en una Tabla de Frecuencias, los Cuartiles se calculan con (2.14). Antes de aplicar este procedimiento, es necesario determinar la *Clase Cuartílica,* que corresponde a la categoría, cuya frecuencia acumulada contiene el Cuartil que se desea calcular. Para esto basta con desarrollar el término $\frac{kn}{4}$ contenido en la fórmula (2.14), donde *k* se remplaza por el Cuartil buscado.

Para determinar el primer Cuartil, se establece primero la Clase Cuartílica 1, así:

$$\frac{n}{4} = \frac{1(69)}{4} = 17{,}25$$

Puesto que el valor 17,25 está contenido en la frecuencia acumulada de la segunda categoría, contenido en la frecuencia absoluta acumulada (f_a = 26), ésta es la Clase Cuartílica Uno y

con base en esta información se continúa con el desarrollo de (2.3.14):

$$Q_1 = 107 + \frac{\frac{1(69)}{4} - 8}{18}(9) = 107 + (0{,}5139 \times 9) = 111{,}625$$

Q_1 se interpreta del mismo modo, indicado en el Ejemplo 2.14, obsérvese que el valor obtenido con la fórmula, no es exactamente igual al encontrado con los datos sin agrupar, debido a que corresponde a una aproximación; sin embargo, la discrepancia es mínima.

Del mismo modo, se procede para encontrar el tercer cuartil.

$$\frac{kn}{4} = \frac{3\,(69)}{4} = 51{,}75$$

Este valor está contenido en la frecuencia absoluta acumulada de la Clase 5, por lo tanto, ésta es la Clase Cuartílica 3. Con esta información se calcula de la siguiente manera:

$$Q_3 = 134 + \frac{\frac{3(69)}{4} - 45}{15}(9) = 134 + 4{,}05 = 138{,}05$$

En este caso, también existe una ligera discrepancia con respecto al Q_3 calculado en el Ejemplo 2.14, hecho que confirma que, las medidas estadísticas obtenidas con datos agrupados, son una aproximación de los valores calculados, cuando se dispone de la información sin agrupar.

Deciles y Percentiles. Los ***Deciles*** dividen la distribución en 10 partes iguales; los ***Percentiles*** los dividen en 100 partes iguales. De manera similar que en los Cuartiles, los datos deben estar en orden de magnitud creciente.

En forma análoga al procedimiento desarrollado para la Mediana y para los Cuartiles, la *situación de los* ***Percentiles*** *(SP_k)* viene dada por la fórmula general:

$$SP_k = \frac{k\,(n)}{100} + 1; k = 1, 2, 3, \dots, 15, \dots, 99 \qquad (2.15)$$

Si al aplicar (2.15) SP_k no es entero, se redondea al entero más cercano.

Ejemplo 2.16: Para hallar el Percentil 75 con los datos del Ejemplo 2.14, en primer lugar, se encuentra la posición del Percentil según (2.15):

$$SP_{75} = \frac{75\,(69)}{100} + 1 = 52{,}75 \cong 53$$

Por lo tanto, el Percentil 75, es el dato 53 y su valor: P_{75} = 137 kg.

Significa que, el 75% de las producciones de leche de las ovejas Manchegas, están por debajo de 137 kg, valor que también corresponde al Q_3.

Para estimar el Percentil 15 con los mismos datos:

$$SP_{15} = \frac{15\,(69)}{100} + 1 = 11{,}35 \cong 11$$

Por lo tanto, el Percentil 15 es el dato número 11, así P_{15} = 108 kg, y significa que el 15% de las producciones de leche de las ovejas Manchegas están por debajo de 108 kg.

Cálculo de Percentiles con datos agrupados. La fórmula para calcular los Percentiles, cuando los datos están agrupados es la siguiente:

$$P_k = L_i + \frac{\frac{kn}{100} - f_{a-1}}{f_i}(a); k = 1, 2, 3, \ldots, 99 \qquad (2.16)$$

Todos los términos de la fórmula (2.16) deben definirse como en (2.14).

Ejemplo 2.17: Con los datos utilizados en la tabla del Ejemplo 2.14, se procede a determinar los Percentiles 9, 25, 95.

De manera similar al proceso desarrollado en el Ejemplo 2.14, tales Percentiles se estiman de la siguiente manera:

$$P_9 = 98 + \frac{\frac{9\,(69)}{100} - 0}{8}(9) = 98 + (0{,}77625 \times 9) = 104{,}986$$

$$P_{25} = 107 + \frac{\frac{25\,(69)}{100} - 8}{18}(9) = 107 + (0{,}514 \times 9) = 111{,}625$$

$$P_{95} = 143 + \frac{\frac{95\,(69)}{100} - 60}{9}(9) = 143 + (0{,}617 \times 9) = 148{,}55$$

La utilidad de las medidas calculadas como Cuartiles, Deciles y Percentiles, es fundamentalmente descriptiva y para la clasificación de los individuos en forma comparativa. Desde el punto de vista práctico, ofrecen una orientación general de lo que sucede en la población analizada. El profesional pecuario, puede de manera ágil, establecer comparaciones con los estándares de la especie que esté bajo su cuidado o administración.

Medidas de Dispersión

Como se indicó antes, las Medidas de Tendencia Central dan una idea del lugar hacia donde tienden a concentrarse la mayoría de observaciones, sin indicar el grado de dispersión de los valores de la variable con relación a ese "valor central". Para describir en forma más completa el comportamiento de cualquier variable de interés, también es necesario conocer las medidas de variabilidad, las cuales indican, la forma como se distribuyen o alejan los datos con respecto a la Media.

Las Medidas de Dispersión más utilizadas son: la ***Varianza,*** la ***Desviación Estándar***, el ***Rango***, el ***Coeficiente de Variación*** y el ***Recorrido Intercuartílico.***

La Varianza y la Desviación Estándar. Son las Medidas de Dispersión más utilizadas. La Varianza o ***Cuadrado Medio*** se define, matemáticamente, como el promedio aritmético del cuadrado de las desviaciones de cada observación con respecto a la Media.

La ***Varianza*** surge de la necesidad de tener una medida, que incluya rodas las diferencias de cada uno de los datos con respecto a la Media, teniendo en cuenta que $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$, de tal manera que, al elevar al cuadrado estos valores, se podrá tener un resultado diferente de cero; sin embargo, al tener un número, cuya unidad de medida está elevada al cuadrado, en algunas ocasiones se dificulta la interpretación real.

El parámetro de la Varianza se representa con la letra griega sigma al cuadrado (σ^2) y el estadígrafo con la letra ***S*** al cuadrado (S^2). Para la Desviación Estándar, el parámetro se representa con σ y el estadígrafo con ***S***. Simbólicamente, σ^2 se define mediante la siguiente fórmula:

$$\sigma^2 = \frac{(y_1-\mu)^2+(y_2-\mu)^2+(y_3-\mu)^2+\cdots+(y_N-\mu)^2}{N} = \frac{\sum_{i=1}^{N}(y_i-\mu)^2}{N} \qquad (2.17)$$

Sin embargo, es importante aclarar, que en prácticamente todos los casos de interés en Producción, Salud Animal y casi todas las áreas de interés, el parámetro no se mide, sino que se estima a partir del estadígrafo, el cual se obtiene en la muestra, por lo cual, (2.17) es inaplicable y se constituye en una expresión teórica, de manera que la fórmula de aplicación, corresponde a la Varianza de la muestra S^2 que se muestra en (2.18), la cual se divide entre n-1 para tener un estimador insesgado de σ^2.

$$S^2 = \frac{(y_1-\bar{y})^2+(y_2-\bar{y})^2+(y_3-\bar{y})^2+\cdots+(y_n-\bar{y})^2}{n-1} = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1} \qquad (2.18)$$

La Varianza se expresa en las mismas unidades de la variable original elevadas al cuadrado; la Desviación Estándar se expresa, en las mismas unidades de los datos, procurando tener por lo menos un decimal más que éstos, por ello, en la presentación y análisis de resultados, se prefiere S en lugar de S^2.

Para facilitar las labores de cálculo de S^2, se puede utilizar la siguiente fórmula de trabajo:

$$S^2 = \frac{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}{n-1} \qquad (2.19)$$

El término $\frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$ se denomina ***Factor de Corrección*** y se identifica como ***C***. La expresión $\sum_{i=1}^{n} y_i^2$ se conoce con el nombre de ***Suma de Cuadrados sin ajustar***.

Ejemplo 2.18: Los datos consignados en la Tabla 2.3 corresponden a la producción de leche en kilogramos por vaca/día, de las campeonas de los torneos y exposiciones de Gyr lechero, organizadas por la Asociación Brasileña de Criadores de Cebú (ABCZ, 2001).

Tabla 2.3

Producción de leche en kg/vaca/día de las campeonas Gyr lechero en Brasil (ABZC, 2001).

Vaca	**Producción (y_i)**	**$(y_i)^2$**	**Vaca**	**Producción (y_i)**	**$(y_i)^2$**
Hamada	19,62	384,944	Vitoria	24,85	617,523
Lagosta	17,06	291,044	Realeza	19,15	366,723
Doncella	19,17	367,489	Brasilia	22,70	515,290
Ervilha	18,02	324,720	Amizade	22,70	515,290
Jarra	18,85	355,323	California	24,38	594,384
Mentira	19,10	364,810	Entrancia	21,65	468,723
Gaviota	18,91	357,588	Fagacia	12,25	150,063
Faianca	17,93	321,485	Fada	34,60	1.197,160
Lucrecia	19,81	392,436	Delicia	27,20	739,840
Maravilha	21,25	451,563	Indígena	28,30	800,890
Gabarra	25,94	672,884	Enamorada	27,50	756,250
Amalia	18,75	351,563	Fieira	23,90	571,210
Tala	21,72	471,758	Cabedella	20,30	412,090
Rebarba	19,58	383,376	Frangancia	25,30	640,090
Zema	20,43	417,385	FB Gambia	26,40	696,960
Valentia	21,70	470,890	Cabina	17,60	309,760
Omaga	22,36	499,970	Gaivota	30,20	912,040
Varanda	20,91	437,228	**Sumatoria**	**770,09**	**17.580,739**

En la Tabla 2.3 se agregó una columna, con la producción al cuadrado, para disponer de todos los datos requeridos en (2.3.19):

$(\sum_{i=1}^{n} y_i)^2$ = (770,09)^2 = 593.038,608

$\sum_{i=1}^{n} y_i^2$ = 17.580,739

n = 35

Al reemplazar en (2.3.19) se tiene:

$$S^2 = \frac{17.580,739 - \frac{(770,09)^2}{35}}{35-1} = 18,7288 \text{ kg}^2$$

$$S = \sqrt{S^2} = \sqrt{18,7288} = 4,3277 \text{ kg}$$

Como se ha dicho, la Varianza no tiene una interpretación práctica, por lo tanto, sólo se podría decir que, la Varianza de la producción de leche de las campeonas Gyr lechero en Brasil, es de 18,7288 kg^2.

En el caso de la Desviación Estándar, es posible tener una conclusión comprensible; en este ejemplo, los valores correspondientes a la producción de leche de las vacas campeonas Gyr lechero en Brasil, en promedio, se desvían con respecto a la Media, en 4,3277 kg. Si bien, con esta medida, se puede tener una visión más real de la distribución de los datos, aún no es objetiva, dado que no es comparable con otro grupo de datos, en donde la escala o las unidades de medida sean diferentes.

Cálculo de la Varianza con datos agrupados. Cuando los datos están agrupados en una Tabla de Frecuencias, la Varianza se calcula de la siguiente manera:

$$S^2 = \frac{\sum_{i=1}^{n} f_i m_i^2 - \frac{\left(\sum_{i=1}^{n} f_i m_i\right)^2}{n}}{n-1} \qquad (2.20)$$

Donde:

fi = frecuencia absoluta de la i-ésima clase.
m_i = marca de clase de la i-ésima categoría.
n = tamaño de la muestra.

Ejemplo 2.19: Los datos de la Tabla 2.3 fueron agrupados y, además, para facilitar los cálculos correspondientes, se complementaron de la siguiente manera:

Clase	Límites de clase		m_i	f_i	$f_i m_i$	m_i^2	$f_i m_i^2$
1	12,25	15,98	14,115	1	14,115	199,233	199,233
2	15,98	19,71	17,845	12	214,140	318,444	3.821,328
3	19,71	23,44	21,575	11	237,325	465,480	5.120,286
4	23,44	27,17	25,305	6	151,830	640,343	3.842,058
5	27,17	30,90	29,035	4	116,140	843,031	3.372,124
6	30,90	34,63	32,765	1	32,765	1.073,545	1.073,545
∑				**35**	**766,315**		**17.428,574**

$\sum f_i m_i^2 = 17.428,574$

$(\sum f_i m_i)^2 = (766,315)^2 = 587.238,679$

$$S^2 = \frac{17.428,574 - \frac{587.238,679}{35}}{34} = 19,127 \text{ kg}^2$$

La Desviación Estándar para esta muestra, es la siguiente:

$$S = \sqrt{S^2} = \sqrt{19,127} = 4,373 \text{ kg}$$

> Se concluye que la Varianza es de 19,127 kg^2, mientras que, en promedio, los valores de producción de leche, se distancian de la media en 4,373 kg.

Usualmente, tanto la Varianza, como la Desviación Estándar, bien sea con datos agrupados como sin agrupar, se calcula con un dígito decimal más que la Media.

El Coeficiente de Variación o Variabilidad (*CV*). Es un valor, cuyo uso más importante es evaluar los resultados experimentales, donde se estudia la misma característica, generalmente por personas distintas y bajo condiciones diferentes.

El *CV* se define, como la Desviación Estándar expresada como Porcentaje de la Media, o también, el Porcentaje de Desviación de los Datos con respecto a la Media. Al ser una medida relativa, presenta una visión más comprensible y comparable por parte del profesional que analiza la información.

El *CV* se calcula con la siguiente fórmula:

$$CV = \frac{S}{\bar{y}}(100) \qquad (2.21)$$

Para establecer si un *CV* es grande o pequeño, es preciso disponer de datos similares para la variable analizada. Por ejemplo, los materiales vegetales tienen cada uno, su propio *CV*, lo mismo sucede con las especies animales.

A diferencia de la *S*, el *CV* es una medida relativa de variación. La primera se expresa en las mismas unidades de la variable y la segunda en porcentaje.

Ejemplo 2.20: Con la información contenida en la Tabla 2.3se calcula el Coeficiente de Variación, de la siguiente manera:

$\bar{y}$ = 22,00 kg S = 4,3277 kg.

$$CV = \frac{4,3277}{22,00}(100) = 19,67\%$$

No es posible establecer si este *CV* es alto o bajo, puesto que no se dispone de otra información que reporte el *CV* para el ganado Gyr en la característica analizada, que permita realizar la comparación.

El Rango (R). Es otra Medida de Dispersión que se define y calcula de la manera indicada en la construcción de Tablas de Frecuencias, es decir, la diferencia entre el dato mayor y el menor. Tiene escaso uso en el proceso de inferencia, ya que no es un estadígrafo satisfactorio, aunque con muestras menores de 10, es frecuentemente útil.

El Rango Intercuartílico (*RIQ*). Es una medida de variabilidad resistente a los "outliers". Si el *RIQ* es grande, debe interpretarse que, en un conjunto de datos ordenados, existe una gran dispersión de éstos con respecto al valor central; si por el contrario el *RIQ* es pequeño, gran parte de las observaciones se encontrará cerca del centro. El *RIQ* se calcula restando el Q_1 del Q_3.

Ejemplo 2.21: En el Ejemplo 2.16 se calculó el Q_1 = 18 y el Q_3 = 52, entonces:

RIQ = 52 – 18 = 34

En este caso es posible afirmar que, no existe una gran dispersión de los datos y que, la tendencia es el agrupamiento en torno al valor central representado por la Mediana.

Asimetría y Curtosis. Las distribuciones, también pueden diferir entre sí, en asimetría y grado de apuntamiento. Una distribución es ***Simétrica,*** cuando la Media, la Mediana y la Moda coinciden en el mismo punto, como se indica en la Figura 2.3.1.

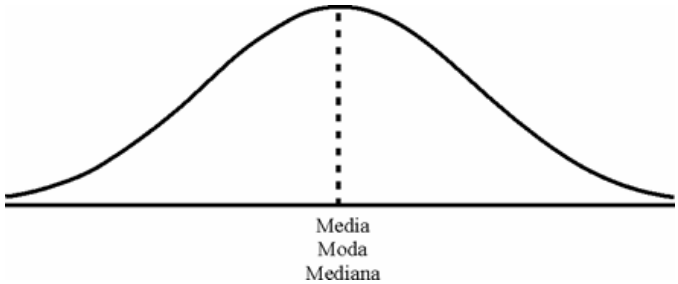


Figura 2.6 Representación de una Distribución Simétrica.

Pero si la distribución es ***Asimétrica***, entonces la Media, la Mediana y la Moda se ubican en puntos diferentes. Si Moda < Mediana < Media, la curva es asimétrica con sesgo positivo. Si Media > Mediana > Moda, la curva es asimétrica con sesgo negativo (Figura 2.7).

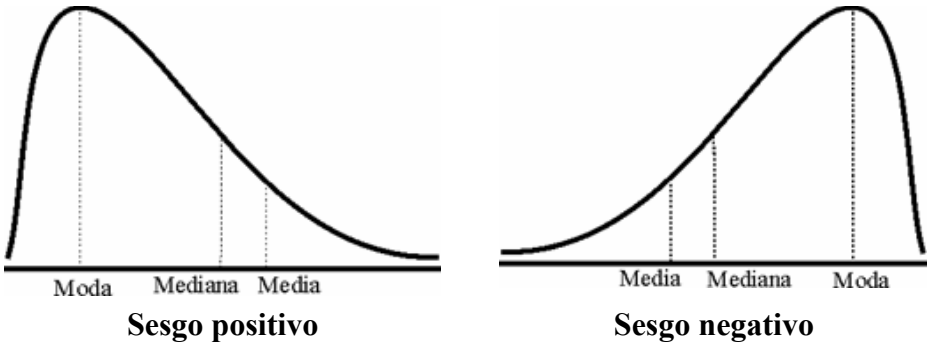


Figura 2.7 Distribuciones Asimétricas con sesgo positivo y negativo.

Coeficiente de Asimetría (C_A). Una medida de asimetría es el Coeficiente de asimetría (C_A), que permite medir el grado de sesgo positivo o negativo, definido por la siguiente fórmula:

$$C_A = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^3}{nS^3} \qquad (2.22)$$

Si: $C_A = 0$, la curva es simétrica.
$C_A > 0$, la curva es asimétrica con sesgo positivo.
$C_A < 0$, la curva es asimétrica con sesgo negativo.

Ejemplo 2.22: En un ensayo con peces ángel o escalares (*Pterophyllum scalare*), se recopiló información sobre longitud estándar medida en centímetros, cuyos valores están consignados a continuación.

15	13	12	14	14	10
14	12	11	13	7	8
11	10	10	9	7	8
8	9	7	7	12	13
15	15	11	14	10	5
13	14	14	6	5	4
15	12	12	7	6	10
12	13	15	5	7	8
8	9	10	9	11	12
12	12	13	12	12	13

Aplicando (2.22) se calculó el Coeficiente de Asimetría, de la siguiente manera:

$$C_A = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^3}{nS^3} = \frac{-578{,}877}{60 \times 27{,}1208} = -0{,}3557$$

> Este resultado lleva a concluir, que la longitud estándar de los peces escalares, indica que la muestra tomada proviene de una población con distribución aproximada a la normal, con un sesgo ligeramente negativo, para esta variable.

Coeficiente de Curtosis (C_{AP}). La Curtosis es una medida de la agudeza o grado de apuntamiento de una distribución, determinada por el Coeficiente de Curtosis o apuntamiento (Figura 2.8).

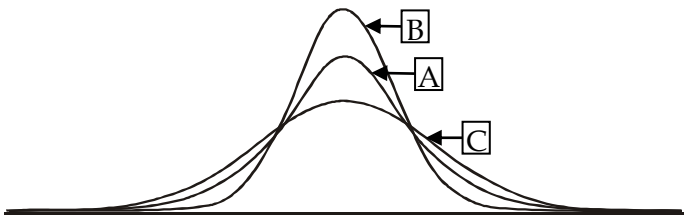


Figura 2.8: Tres tipos de Curtosis (A: Mesocúrtica, B: Leptocúrtica, C: Platicúrica).

El C_{Ap} se estima mediante la siguiente fórmula:

$$C_{Ap} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^4}{nS^4} \qquad (2.23)$$

Si: $C_{Ap} = 3$, la curva es Mesocúrtica o normal.
$C_{Ap} > 3$, la curva es Leptocúrtica.
$C_{Ap} < 3$, la curva es Platicúrtica.

En Estadística Inferencial es importante el concepto de Normalidad de las Muestras. Los conceptos de Asimetría y Curtosis son una herramienta útil para establecer si la muestra proviene de una población normal. Si la muestra es bastante Asimétrica y el Coeficiente de Curtosis está bastante alejado de

tres, se descarta normalidad, en caso contrario, es decir, si los datos son Simétricos y el Coeficiente de Curtosis cercano a tres, los datos provienen de una población con una distribución bastante aproximada a la normal y, en este caso, lo más razonable es utilizar una Prueba Estadística de Normalidad.

Ejemplo 2.23: Tomando los datos del Ejemplo 2.22, se aplica la fórmula (2.23) para calcular el Coeficiente de Curtosis, de la siguiente manera:

$$C_{Ap} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^4}{nS^4} = \frac{9.917{,}46}{60 \times 81{,}4838} = 2{,}0285$$

Con base en el valor resultante se podría decir que, los datos de la muestra de peces escalares generan una Curva Platicúrtica.

Es importante mencionar que, en trabajos experimentales, el tamaño de la muestra debe ser suficientemente grande, como para establecer de manera confiable la forma de la curva.

Diagramas de Caja. El Diagrama de Caja es una herramienta útil en el análisis de un conjunto de datos univariado, debido a que facilita el análisis de la simetría, detecta valores atípicos y vislumbra un ajuste de los datos, a una distribución de frecuencias determinada.



Figura 2.9 Diagrama de Caja

La mitad central de los datos que va desde el primero hasta el tercer Cuartil, se los representa mediante un rectángulo; la Mediana se la identifica con un segmento de recta vertical que va dentro del rectángulo. Un segmento de recta se extiende desde el tercer Cuartil hasta el valor máximo y otro se extiende desde el primer Cuartil hasta el valor mínimo, como se muestra en la Figura 2.9.

Ejemplo 2.24: De un experimento reportado por Milton (1994), donde se administró CO_2 a diferentes presiones atmosféricas, para establecer la influencia sobre el crecimiento del *Pseudomonas fragi*, medido como el cambio porcentual en la masa celular después de una hora, se extractó los resultados que se encuentras consignados en la Tabla 2.4.

Tabla 2.4

Masa celular de *Pseudomona frag*i bajo diferentes presiones atmosféricas (Milton, 1994)

Presión en atmósferas de CO_2			
a	**b**	**c**	**d**
0,0	**0,083**	**0,29**	**0,50**
62,6	50,9	45,5	29,5
59,6	44,3	41,1	22,8
64,5	47,5	29,8	19,2
59,3	49,5	38,3	20,6
58,6	48,5	40,2	29,2
64,6	50,4	38,5	24,1
		30,2	32,7
		27,0	24,4

En la Figura 2.10 se indica los Diagramas de Cajas para los cuatro tratamientos. Allí se puede apreciar que, en el tratamiento **a** es donde se presenta el mayor crecimiento de la masa celular, luego en el tratamiento **b**; el tratamiento de menor crecimiento es el **d**. La mayor variabilidad se observa en los tratamientos **c** y **d**, respectivamente, puesto que sus cajas son las más largas. En cuanto a la simetría, los tratamientos **a** y **b** exhiben las distribuciones más simétricas, puesto que sus respectivas Medianas están situadas cerca de la mitad de las cajas; los tratamientos **c** y **d** son asimétricos, el primero con sesgo a la izquierda (negativo) y el segundo con sesgo a la derecha (positivo).

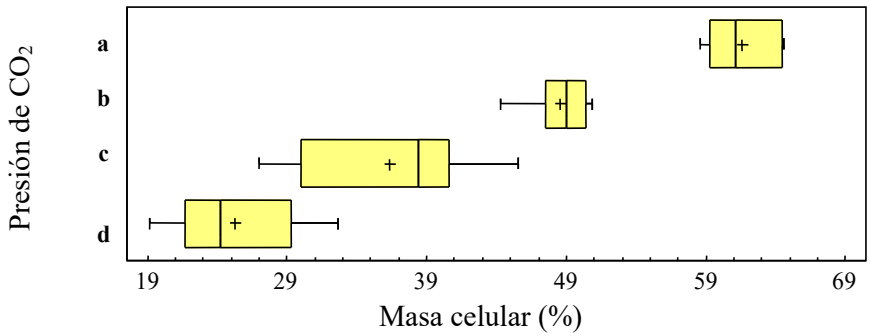


Figura 2.10 Diagramas de Caja para el crecimiento de la masa celular del *Pseudomonas fragi,* según la presión de CO_2 (Milton, 1994).

CAPÍTULO 3

Principios básicos de probabilidad

INTRODUCCIÓN

En el Capítulo 1 se mencionó que, la Teoría de las Probabilidades, se inició a principios del siglo XVII, aplicada a los juegos de azar, motivados por la posibilidad de acierto o de fracaso. En la actualidad, se considera una rama de la Estadística, que juega un papel preponderante en la inferencia estadística, al constituirse en un mecanismo para estudiar los sucesos aleatorios, es decir, aquellos que no pueden ser predichos de antemano.

En este texto se insiste, en la inconveniencia de formular definiciones de los aspectos estadísticos de mayor importancia, incluyendo los conceptos de probabilidades. Al abordar este tema se pretende, básicamente, manejar en forma adecuada, ciertos conceptos y procedimientos que permitan tener una idea clara de las Probabilidades y su importancia en los temas de inferencia estadística que se estudiarán posteriormente.

La Teoría de las Probabilidades, se utiliza con más frecuencia de lo que se cree, no sólo en las ciencias animales sino en muchas disciplinas. Por ejemplo, en Genética es posible predecir el desempeño de la progenie, con base en la información de los padres y otros ascendientes; también es factible, determinar la frecuencia con que aparecerán ciertos fenotipos en una determinada población, tales como, el color del pelo o una enfermedad de origen hereditario, entre otros;

en el control de calidad de productos lácteos o cárnicos de una factoría, es posible determinar la proporción de unidades procesadas y empacadas que no cumplirán con las normas sanitarias o de calidad admitidas para su comercialización. En los procesos reproductivos de los seres vivos, es necesario establecer la posibilidad, de que algunos de ellos puedan producir gametos con diferentes características; también se utilizan los conceptos de la Teoría de las Probabilidades en la planificación y seguimiento de programas para el control y/o erradicación de enfermedades de los animales.

Se advierte al lector que, la Teoría de las Probabilidades es un campo bastante amplio, que exige niveles de abstracción y conocimientos matemáticos avanzados, cuando se estudia con profundidad; sin embargo, en esta obra no se intentará abordar estudios de esa naturaleza. Para los interesados en profundizar en este tema, se recomiendan, entre otras, las obras de Meyer (1973) y Parzen (1992).

CONCEPTOS BÁSICOS PARA EL ESTUDIO DE LAS PROBABILIDADES

Con el propósito de iniciar el estudio de las Probabilidades, es necesario recalcar que, en las ciencias animales, la identificación y observación de ***Fenómenos Naturales***, bajo cierto conjunto de condiciones, constituye la materia prima para la labor profesional y de investigación. Así, por ejemplo, en una población bovina, se observan vacas con o sin cuernos; en una piara, cerdos con o sin síntomas de alguna enfermedad, en un rebaño, animales con diferente peso al nacimiento, animales sometidos al mismo tratamiento que tardan tiempos diferentes en su recuperación; en un lote de tilapia roja, se observan animales con manchas de diferentes características; en un medio de cultivo, bacterias resistentes a determinados antibióticos y muchos otros fenómenos naturales.

Al observar repetidamente el mismo fenómeno, no siempre se llega a resultados idénticos y, por el contrario, la variación de éstos, es una característica notoria; en un gran número de ***Repeticiones,*** un resultado aparecerá varias veces con cierta frecuencia, que se denomina ***Regularidad Estadística*** y los fenómenos donde se aprecia dicha regularidad reciben el nombre de ***Fenómenos Aleatorios.***

La representación numérica de la Regularidad Estadística se hace con números comprendidos entre cero y uno, como indicación de la ***frecuencia relativa*** con que se presentan los diferentes resultados en una serie de repeticiones ***independientes entre sí***.

Ejemplo 3.1 En una feria ganadera se encuentran novillos de tres razas diferentes: Cebú, Normando o Shorthorn. Un profesional agropecuario, se encarga de realizar una inspección general a cada animal y registra la raza a la que pertenece cada uno. En un momento dado, se le pregunta ¿De qué raza será el próximo novillo que será inspeccionado? Aparentemente esta pregunta no tiene respuesta, ya que el novillo puede ser de cualquiera de las tres razas. Sin embargo, si se realiza un "experimento" consistente en determinar y anotar la raza de cada novillo, aparecerán resultados particulares, como los presentados a continuación:

Cebú:	120
Normando:	28
Shorthorn:	2
Total:	**150**

Con estos resultados, se establecen las siguientes frecuencias relativas:

Para el ganado Cebú: $\frac{120}{150} = 0{,}8$

Para el Normando: $\frac{28}{150} = 0{,}19$

Para el Shorthorn: $\frac{2}{150} = 0{,}01$

Si se continúa registrando la información, seguramente obtendrá valores de frecuencia relativa diferentes a los anteriores, pero a medida que repita este procedimiento, hasta alcanzar un número "grande" de registros, las fluctuaciones de los valores de dichas frecuencias en las sucesivas repeticiones, se volverán más pequeñas, hasta estabilizarse. A esta estabilización de las frecuencias relativas, cuando el número de repeticiones de un experimento tiende al infinito, se le llama Regularidad Estadística, característica típica de los ***Experimentos Aleatorios,*** que se pueden definir, como los procedimientos que se llevan a cabo para registrar una observación.

Cada registro, corresponde a un ***Evento o Suceso*** aleatorio, definido como un resultado posible en un experimento. Para el Ejemplo 3.1, los sucesos son: que el próximo novillo sea Cebú, Normando o Shorthorn; es decir, hay tres resultados posibles, cuya frecuencia relativa en una sucesión grande de observaciones llevadas a cabo aleatoriamente, se acerca a un ***Valor Límite Estable*** a medida que el número de observaciones tiende al infinito. Por lo tanto, ese valor límite estable, puede considerarse como la ***Probabilidad*** de ocurrencia del evento aleatorio. El ***Espacio Muestral,*** denotado con la letra ***S***, contiene las descripciones de ***todos*** los posibles resultados de un experimento.

En el Ejemplo 3.1 se puede observar, como en todos los experimentos aleatorios, que los valores de las frecuencias relativas están comprendidos entre 0 y 1, los cuales reflejan, las expectativas de ocurrencia de un suceso determinado, de tal modo que, probabilidades próximas a uno, indican, que ese tipo de evento generalmente se produce, no que vaya a producirse con

certeza; probabilidades cercanas a cero, conducen a pensar, que el suceso difícilmente se producirá, mas no indican que no vaya a producirse; probabilidades cercanas a 0,5 indican que es tan verosímil que el suceso se produzca, como que no se produzca.

Después de analizados los anteriores conceptos, al formular la pregunta ¿De qué raza será el próximo novillo a ser inspeccionado?. El profesional podrá decir, que es una pregunta sin respuesta, pero si la pregunta fuese *¿Cuál es la Probabilidad, de que el próximo novillo a ser inspeccionado sea de una raza específica, de las tres que se han registrado?* tendrá elementos para responderla correctamente, en términos *probabilísticos*. Si la frecuencia relativa estable es de 0,8 para la Raza Cebú, podría contestar, que de los próximos 10 novillos que entran a la feria, ocho serán Cebú y seguramente, el dato predicho estará muy cercano al real y, a medida que se incremente el número de observaciones, su predicción coincidirá en mayor medida con lo que ocurrirá realmente.

De lo expresado hasta este punto, se puede concluir que, la Estadística no se ocupa de experimentos donde es posible anticipar el resultado, sino de aquellos donde los resultados están sujetos al ***azar***, es decir, los ***experimentos aleatorios.***

ESPACIO MUESTRAL

Con el fin de ampliar el concepto de Espacio Muestral (*S*), se presentan algunos ejemplos:

Ejemplo 3.2: En un experimento de campo, conducente a detectar mastitis con la prueba CMT, en un hato de 10 vacas, el espacio muestral se describe de la siguiente manera: $S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Esto significa, que en *S* están contenidos todos los

resultados posibles del experimento, 0 indica que ninguna vaca resulta positiva a la prueba; 1 que se detecta una vaca enferma y así sucesivamente hasta 10.

Ejemplo 3.3: De un recipiente que contiene seis bolas idénticas, excepto por su color, donde cuatro son blancas y dos son rojas, se extrae una bola, se observa y anota su color e inmediatamente se devuelve al recipiente. En este caso el espacio muestral es: $S = \{roja, blanca\}$.

Ejemplo 3.4: De un recipiente que contiene seis bolas idénticas, numeradas del 1 al 6, se extraen dos, primero una y luego otra, sin devolver al recipiente la primera antes de sacar la segunda. En este caso, el espacio muestral se describe del siguiente modo:

S = {(1-2), (1-3), (1-4), (1-5), (1-6), (2-1), (2-3), (2-4), (2-5), (2-6), (3-1), (3-2), (3-4), (3-5), (3-6), (4-1), (4-2), (4-3), (4-5), (4-6), (5-1), (5-2), (5-3), (5-4), (5-6), (6-1), (6-2), (6-3), (6-4), (6-5)}.

Es decir, bajo las condiciones especificadas para el experimento del Ejemplo 3.4, hay 30 resultados posibles, que conforman el espacio muestral.

Ejemplo 3.5: En el control de apareamientos, realizado rutinariamente en una piara, cuando nace una cría, se anota, en su correspondiente registro, el tamaño de camada de donde proviene el padre y el tamaño de camada, del cual proviene la madre.

Suponiendo que el padre haya nacido en una camada de tamaño siete y la madre en una de tamaño cinco, este apareamiento se

> registrará como el par ordenado (5, 7); la observación (7, 12) corresponderá al tamaño de camada 7 para el padre y 12 para la madre. Como los tamaños de camada pueden tomar cualquier valor entero entre 1 y 16, tanto para los machos como para las hembras, existe una gran cantidad de resultados difíciles de describir individualmente, por lo tanto, el espacio muestral, se describe de mejor manera así: se denomina ***X*** el tamaño de camada de los padres, ***Y*** el tamaño de camada de las madres, en consecuencia, el espacio muestral se describe así:
>
> $S = \{$*los pares ordenados* (X, Y) *tal que X es un entero del* 1 *al* 16*, Y es un entero del* 1 *al* 16$\}$.

Del análisis de los tres ejemplos, se desprenden varios hechos:

- El espacio muestral (*S*) puede tener un alto grado de dificultad para su descripción.
- *S* cambia según las condiciones en que se lleve a cabo el experimento, por eso es preferible hablar de ***un*** espacio muestral referido a un *experimento concreto*.
- *S* puede ser de ***tamaño finito***, pero existen otros espacios de ***tamaño infinito*** que se describen adoptando alguna notación específica, que se explicará más adelante.
- Un espacio muestral, no necesariamente está constituido por números.
- En todas las descripciones de *S,* convencionalmente se utilizan llaves {} con el fin de indicar, que realmente se está trabajando con ***Conjuntos***, dentro de las cuales se hace una lista de todos sus ***Elementos***. Por Conjunto se entiende "cualquier colección de conceptos o de objetos perfectamente especificada" (Meyer, 1973).

El ***Tamaño de un Conjunto*** está definido, por el número de elementos que pertenecen a él. Un Conjunto es Finito, si su tamaño es cualquiera de los números finitos; así por ejemplo, el conjunto de granjas que posee la Universidad de Nariño es 5, S = {*Botana, Chimangual, La Quinta, Mar Agrícola, Guamués*}; si el conjunto de vacas en producción láctea en una de las granjas es 10:

S = {01, 02, 03, 04, 05, 06, 07, 08, 09, 10}.

En este caso, los Elementos del Conjunto, se representan con el código que identifica a cada animal.

Un Conjunto Infinito puede ser numerable o no numerable; por ejemplo, el conjunto de los números impares es infinito numerable, en donde S = {1, 3, 5, 7, 9, ..., ∞}; en cambio, el número de puntos en la recta real en el intervalo 0 a 1 es infinito no numerable y, por lo tanto, $S = \{X \mid 0 \leq X \leq 1\}$.

En muchos textos, el Espacio Muestral se representa con la letra ***U***, para resaltar que se trata del conjunto universal que, como se sabe, contiene todas las descripciones.

EVENTOS

Para definir un Evento es necesario relacionarlo con el espacio muestral S, ya que un Evento ocurre si y sólo si el resultado observado corresponde a una de las descripciones contenidas en S; así, en el Ejemplo 3.3 se pueden presentar varios Eventos: extraer una bola blanca, extraer una bola roja, sacar primero una bola roja y luego una blanca, sacar primero una bola blanca y luego otra blanca. Se observa, que todos los casos descritos aparecen en S. Hecha esta precisión, un Evento será considerado como un ***Subconjunto*** de S.

Para describir un Conjunto se utilizan convencionalmente, letras mayúsculas y algunos símbolos; así, para indicar que un Elemento pertenece a un Conjunto, se utiliza el símbolo $\in$; para especificar que un Elemento no pertenece a un Conjunto se utiliza el símbolo $\notin$. El símbolo | debe leerse como "tal que", que especifica una condición particular de los elementos del Conjunto. El símbolo $\subset$ indica que todos los Elementos de un Conjunto son también Elementos de otro, de este modo la expresión A$\subset$ B, deberá entenderse como A es un subconjunto de B.

Dado que *S* contiene todas las descripciones posibles, también es un subconjunto de sí mismo, por lo que constituye un evento llamado ***Evento Seguro***.

Cuando se estudian Fenómenos Aleatorios, realmente el interés se centra en los eventos que pueden ocurrir y, concretamente, en la Probabilidad de su ocurrencia y no en el Espacio Muestral, sino en sus subconjuntos, que son los Eventos.

OPERACIONES CON CONJUNTOS

Con los Eventos y el Espacio Muestral, considerados en términos de conjuntos, es posible efectuar varias operaciones algebraicas; las de mayor utilidad son la Adición y la Multiplicación, cuyas reglas se resumen a continuación.

La Suma o Unión de Conjuntos ($A \cup B$)

Se describe como el conjunto de las descripciones que están contenidas en A, en B o en ambos; entonces ($A \cup B$) ocurre si y sólo si ocurre A o bien ocurre B o si ocurren ambos, es decir, cuando el resultado observado tiene como descripción muestral, un elemento de A, de B o de ambos.

Ejemplo 3.6: Si se define $S = \{Y|Y$ *es un Bos taurus puro de las razas Angus, Hereford, Holstein o Pardo Suizo*$\}$

Considerando los siguientes subconjuntos: $A = \{Y|Y$ *es una vaca Holstein*$\}$, $B = \{Y/Y$ *es una vaca Pardo Suizo*$\}$, el conjunto definido como las vacas productoras de leche $A \cup B = \{Y|Y$ *es una vaca de las razas Holstein o Pardo Suizo*$\}$.

El Producto o Intersección de Conjuntos ($A \cap B$)

Corresponde al conjunto de elementos presentes, tanto en A, como en B, o sea que $(A \cap B)$ ocurre si y sólo si A y B ocurren. Si A y B son eventos mutuamente excluyentes, no pueden ocurrir juntos, por lo tanto $(A \cap B) = \phi$.

Ejemplo 3.7: Definidos los conjuntos:

$A = \{Y|Y$ *es una vaca Pardo Suizo lactante*$\}$
$B = \{Y|Y$ *es una vaca Pardo Suizo en gestación*$\}$
el conjunto resultante de la intersección será:
$A \cap B = \{Y|Y$ *es una vaca Pardo Suizo lactante y gestante a la vez*$\}$

Complemento de A ($\bar{A}$)

Si A es un Evento, puede calcularse la Probabilidad de que A ocurra o no. Si A es el Conjunto, $\bar{A}$ es su complemento, es decir, representa todas las descripciones presentes en S que no están en A, por lo tanto, $\bar{A}$ ocurre si y sólo si A no ocurre.

Ejemplo 3.8 Siendo el Espacio Muestral $S = \{Y|Y$ *es un perro*$\}$, en el cual se define el Conjunto $A = \{Y|Y$ *es un perro vacunado contra*

el parvovirus} $\subset S$, entonces, $\bar{\boldsymbol{A}} = \{Y/Y$ *es un perro no vacunado contra el parvovirus*$\}$.

Las operaciones de unión, intersección y el complemento se pueden representar con los Diagramas de Venn de la Figura 3.1.

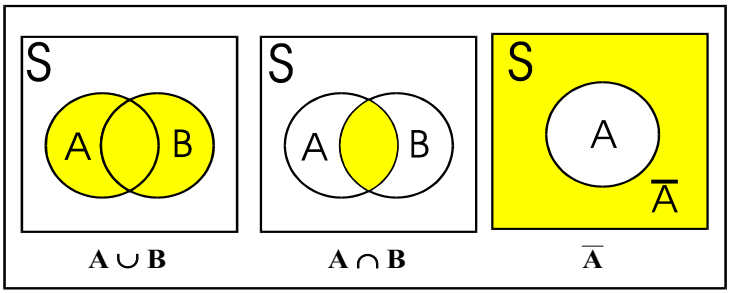


Figura 3.1 Diagramas de Venn para las operaciones con conjuntos.

Ejemplo 3.9: Una bolsa contiene 12 bolas idénticas, numeradas del 1 al 12.

$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

Si se toman dos subconjuntos de S:
$A = \{1, 2, 3, 4, 5, 6\}$
$B = \{4, 5, 6, 7, 8, 9\}$

Los resultados de las tres operaciones son los siguientes:

$A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$
$A \cap B = \{4, 5, 6\}$
$\bar{\boldsymbol{A}} = \{7, 8, 9, 10, 11, 12\}$

Conviene también indicar, que dos conjuntos son iguales ($A = B$), si y sólo si cada una de las descripciones en A pertenece también

a B y viceversa.

Las operaciones de unión e intersección definidas anteriormente, se pueden extender a más de dos conjuntos, así:

$A \cup B \cup C = A \cup (B \cup C) = (A \cup B) \cup C$
$A \cap B \cap C = A \cap (B \cap C) = (A \cap B) \cap C$

Las Propiedades Conmutativas de los Conjuntos indican que:

a. $A \cup B = B \cup A$
b. $A \cap B = B \cap A$

Y las Propiedades Asociativas:

a. $A \cup (B \cup C) = (A \cup B) \cup C$
b. $A \cap (B \cap C) = (A \cap B) \cap C$

Existen otros conjuntos que contienen unión, intersección y complementación. Los más importantes se indican enseguida:

a. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
b. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
c. $A \cap \phi = \phi$
d. $A \cup \phi = A$
e. $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$
f. $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$

Como ya se señaló, ϕ es el Conjunto Vacío que, en las operaciones de unión e intersección, se comporta como el cero (0) entre los números reales.

PROBABILIDAD DE UN EVENTO

Con los elementos señalados hasta este punto, se podría deducir que, dada una situación aleatoria, descrita por un Espacio Muestral *S*, la Probabilidad es una *función que asigna a cada evento un número real no negativo*, denotado por $P[A]$, llamada ***Probabilidad del Evento A***, que debe satisfacer los siguientes principios:

a. $P[A] \geq 0$ para todos los eventos A de S.

b. $P[S] = 1$

c. Si A y B son eventos que se excluyen mutuamente en S, entonces:
$P[A \cup B] = P[A] + P[B]$

d. Si $A_1, A_2,\ldots$ es una serie de eventos mutuamente excluyentes, entonces:

$P[A_1 \cup A_2 \cup ...] = P[A_1] + P[A_2] + ...$

De estos principios fundamentales, se derivan las siguientes propiedades:

a. Si $A_1, A_2, A_3 ..., A_n$ son eventos mutuamente excluyentes en un Espacio Muestral S, entonces:

$P[A_1 \cup A_2 \cup A_3 \cup ..... \cup A_n] = P[A_1] + P[A_2] + P[A_3] + ... + P[A_n]$

b. $P[\phi] = 0$, donde ϕ es el Evento Vacío, también llamado Evento Imposible.

c. $P[\bar{A}] = 1 - P[A]$

d. $P[A \cup B] = P[A] + P[B] - P[A \cap B]$, para cualquier par de eventos A y B

CÁLCULO DE PROBABILIDADES CON ESPACIOS MUESTRALES FINITOS

Con el fin de caracterizar $P[A]$ en modelos con Espacio Muestral Finito, se considerará, en primer lugar, el suceso que está constituido por ***un solo resultado***, denominado ***Suceso Elemental,*** que se denota por $A = \{a_i\}$. Bajo estas condiciones, a cada Suceso Elemental $\{a_i\}$ se le asigna un número p_i, llamado la ***Probabilidad de*** $\{a_i\}$, que satisface las siguientes condiciones:

a. $p_i \geq 0, i = 1, 2, ..., k$

b. $p_1 + p_2 + ... + p_k = 1$

La Probabilidad $P[A]$ de un evento A se define, entonces, como la Suma de las Probabilidades de los Sucesos Elementales.

Ejemplo 3.10: Si se lanza un dado ***una vez***, el Espacio Muestral asociado a este experimento, será: $S = \{1, 2, 3, 4, 5, 6\}$, el evento denotado $A = \{2\}$, consistente en que el resultado del experimento sea el número 2, es un Evento Elemental. En cambio, si el resultado del experimento consistiese en obtener un número par, entonces $A = \{2, 4, 6\}$ no es Suceso Elemental. Debe distinguirse, además, que 2 es un elemento del Espacio Muestral S, mientras que $\{2\}$ es un Evento Elemental.

Ejemplo 3.11: En una bolsa se colocan seis bolas idénticas, numeradas del 1 al 6; el experimento consiste en extraer una bola.

El Espacio Muestral estará dado por $S = \{1, 2, 3, 4, 5, 6\}$

Considerando un suceso particular, descrito de la siguiente manera:

A = {X|X corresponde a la primera bola extraída marcada con el número 6}

En este caso *A* es un Evento Elemental.

Sucesos Igualmente Probables

Por lo general, se asume que, cuando los espacios muestrales son finitos, los resultados son igualmente probables o ***Equiprobables***; sin embargo, ésta es una apreciación equivocada, puesto que hay muchos experimentos para los cuales no se garantiza esta suposición. Por ejemplo, no es tan probable que entren a la feria ganadera de Pasto, novillos blanco azul belga como Holstein mestizos, es decir, los eventos no son equiprobables pese a que el Espacio Muestral es finito.

El procedimiento para calcular *P*[*A*], cuando los eventos son igualmente probables, está dado por:

$$P[A] = \frac{Número\ de\ sucesos\ favorables\ para\ A}{Número\ total\ de\ sucesos} \qquad (3.1)$$

Se recalca que (3.1) es una consecuencia de suponer resultados equiprobables, sólo aplicable cuando se satisface esta suposición y no constituye una definición general de Probabilidad.

Ejemplo 3.12: Si se lanza un dado normal y se pregunta ¿Cuál es la Probabilidad de que el resultado sea el número 4?

El evento *A* = {4}; *S* = {1, 2, 3, 4, 5, 6}. En *S* hay seis resultados posibles, de los cuales, uno es favorable para *A*; por lo tanto, de acuerdo con (3.1), la Probabilidad de obtener un 4 al lanzar un dado normal es $P[A] = \frac{1}{6}$

Ejemplo 3.13: En el mismo experimento del Ejemplo 3.12, se pregunta ¿Cuál es la Probabilidad de obtener el número 3?

La respuesta se expresa como la probabilidad de obtener un 3: $P[A] = \frac{1}{6}$

Ejemplo 3.14: Considerando el mismo experimento del Ejemplo 3.12, ¿Cuál es la Probabilidad de obtener un número par? Esto es $A = \{2, 4, 6\}$.

De acuerdo con el mismo procedimiento, la Probabilidad de obtener un número par, se expresa: $P[A] = \frac{3}{6} = \frac{1}{2}$

Ejemplo 3.15: Se lanza una vez una moneda, ¿Cuál es la Probabilidad de que el resultado sea cara?

$S = \{cara, sello\}, \quad A = \{cara\}, \quad P[A] = \frac{1}{2} = 0{,}5$

En este ejemplo, la Probabilidad de que se obtenga una cara al lanzar una moneda, es 0,5.

Ejemplo 3.16: ¿Cuál es la Probabilidad de obtener tres caras, si se lanzan al mismo tiempo tres monedas?

El problema debe abordarse así:

a. Definir *S*. En cada moneda hay dos (2) resultados posibles que son cara (*c*) y sello (*s*); como se lanza tres monedas, se puede obtener ocho (8) resultados:

   $S = \{ccc, css, csc, sss, ssc, scs, ccs, scc\}$

b. ¿Cuántos de esos resultados son favorables para A? Hay un solo resultado favorable que es el evento pedido: $A = \{ccc\}$

c. Aplicar (3.1): $P[A] = \frac{1}{8} = 0{,}125$

Finalmente, es posible concluir que, la Probabilidad de obtener tres caras, en este caso, es de 0,15.

Ejemplo 3.17: Se lanza una moneda dos veces y se pregunta: ¿Cuál es la Probabilidad de obtener una cara?

En este caso $S = \{cc, cs, sc, ss\}$ y $A = \{aparece\ una\ cara\}$

Como se puede observar, el número de casos favorables de A son dos, *cs* y *sc*, luego la Probabilidad solicitada es $P[A] = \frac{2}{4} = \frac{1}{2} = 0{,}5$

Ejemplo 3.18: En el mismo experimento del Ejemplo 3.16, se pregunta: ¿Cuál es la probabilidad de obtener dos caras?

La respuesta es $P[A] = \frac{1}{4}$

Ejemplo 3.19: En el mismo experimento, si la pregunta fuese: ¿Cuál es la Probabilidad de obtener una cara y un sello?

La respuesta es $P[A] = \frac{1}{2}$

En los Ejemplos 3.12 a 3.19, es claro que el interés se centra, en la elección al azar de uno o más elementos de un conjunto dado, por lo tanto, es necesario formular algunas indicaciones generales de utilidad, en los siguientes casos:

- **Caso 1: Escoger al azar un objeto.** Si se dispone de ***n*** objetos o elementos, convencionalmente designados como $a_1, a_2, ..., a_n$, escoger al azar un objeto significa, que cada uno de ellos tiene la misma Probabilidad de ser elegido. En este caso:

$$P[elegir\ a_i] = \frac{1}{n} \qquad (3.2)$$

- **Caso 2: Escoger al azar dos objetos entre *n* objetos.** Significa que ***cada uno de los pares*** de objetos, sin considerar el orden en que sean tomados, tienen la misma Probabilidad de ser elegidos.

  Si se desea elegir dos objetos al azar de entre $\{a_1, a_2, a_3, a_4\}$, obtener a_1 y a_2 es tan probable como obtener a_2 y a_3 y así para cualquier par. Este hecho conduce a preguntar ¿Cuántos ***pares diferentes*** hay en esa colección de objetos? Si hay ***K*** de tales pares, la Probabilidad de escoger cada par, se define como:

$$P[elegir\ un\ par\ a_{ij}] = \frac{1}{K} \qquad (3.3)$$

  La manera de calcular *K,* se explicará más adelante.

- **Caso 3: Escoger al azar *n* objetos de entre *N* objetos donde $n \leq N$.** Significa que, cada *n*-tuplo (en algunos textos llamadas *n*-adas o *r*-plas), tiene tanta probabilidad de ser escogido como cualquier otro n-tuplo. Se entiende por *n*-tuplo una colección de 3, 4, 5 o más objetos.

  Si se desea elegir tres objetos al azar de entre $\{a_1, a_2, a_3, a_4, a_5, a_6\}$, obtener a_1, a_2 y a_3 es tan probable como obtener a_2, a_3 y a_4 y así para cualquier triada y en general n-tuplo. Este hecho conduce a preguntar

¿Cuántos n-tuplos diferentes hay en esa colección de objetos? Si hay ***K*** de tales n-tuplos, la Probabilidad de escoger cada uno, se define como:

$$P[\textit{elegir un n-tuplo } a_{ijk}] = \frac{1}{K} \qquad (3.4)$$

Métodos de Enumeración

En los ejemplos previos, para el cálculo de $P[A]$, con Eventos Equiprobables, ha sido fácil calcular el número de casos favorables y el número total de casos posibles, para luego aplicar la correspondiente fórmula y calcular cualquier Probabilidad requerida; sin embargo, existen situaciones donde estos cálculos se complican, por lo que surge la necesidad de tener un procedimiento de enumeración que facilite encontrar una determinada Probabilidad.

Ejemplo 3.20: En un refrigerador se encuentran almacenados 50 vasos con yogur, de los cuales, 10 no son aptos para su comercialización, porque han sobrepasado la fecha de vencimiento y 40 están en perfecto estado, sin que se pueda distinguir, a simple vista, unos de otros. Se eligen seis vasos al azar, sin devolver al refrigerador el vaso escogido antes de elegir el próximo y se pregunta ¿Cuál es la Probabilidad de que, la mitad de los vasos seleccionados, no sean aptos para su comercialización?

Para resolver este problema, se debe seguir varias etapas:

a. Se requiere obtener seis vasos de entre 50.
b. Establecer de cuántas maneras se pueden escoger seis de entre 50 vasos.

c. Entre todos esos resultados posibles, cuáles tienen la característica de corresponder con el 50% de vasos no aptos para comercializarse.

Como ésta, existen situaciones similares, donde se dificulta resolver los literales b y c como condición previa para calcular la Probabilidad. En consecuencia, se requiere estudiar algunos procedimientos sistemáticos de enumeración, introduciendo así, el estudio del ***Análisis Combinatorio*** que se trata a continuación.

Principio de la Multiplicación o del Conteo. El análisis combinatorio, parte de un principio básico, que indica que *"si un suceso cualquiera puede ocurrir de n_1 maneras y si cuando éste ha ocurrido otro suceso puede ocurrir de n_2 maneras, entonces el número de maneras en que los dos pueden ocurrir en el* ***orden especificado*** *es $n_1 \times n_2$"* (Spiegel, 1991). De este principio, resulta fundamental considerar, que las palabras claves en la definición son "*el orden especificado"*.

Ejemplo 3.21: En una finca, los animales se identifican individualmente, con una placa metálica, cuya numeración está constituida por cuatro dígitos que van del 0 al 9, sin permitir repeticiones. Bajo ese sistema ¿Cuántas placas de marcación pueden elaborarse?

El primer número puede ser cualquiera del 0 al 9, el segundo cualquiera excepto el primero, el tercero cualquiera diferente al primero y segundo y el cuarto cualquiera diferente de los tres primeros, puesto que ningún dígito puede estar repetido.

Con base en lo anterior, es posible brindar una respuesta multiplicando de la siguiente manera: $10 \times 9 \times 8 \times 7 = 5.040$ placas de marcación diferentes.

Ejemplo 3.22: Si en el mismo sistema del Ejemplo 3.21, no se permite que las placas empiecen por 0, ¿Cuántas marcas pueden formarse?

La respuesta es: $9 \times 8 \times 7 \times 6 = 3.024$ placas.

Ejemplo 3.23: La clave para acceder al menú principal de un computador está compuesta por tres letras. ¿Cuántos códigos de acceso pueden formarse, si se permiten repeticiones de letras?

Dado que en el alfabeto español tiene 28 letras y éstas se pueden repetir, entonces:

Respuesta: $28 \times 28 \times 28 = 21.952$

Ejemplo 3.24: Con la información del Ejemplo 3.23, sin permitir repeticiones de letra, ¿Cuántas claves se pueden formar?

Respuesta: $28 \times 27 \times 26 = 19.656$ códigos

Ejemplo 3.25: Si el sistema de marcación de los animales en una finca tiene siete dígitos del 0 al 9, no se permiten repeticiones y los dos primeros números corresponden a las dos últimas cifras del año de nacimiento, ¿Cuántas marcas pueden formarse, para los animales nacidos en 1996?

El primer número sólo puede ser el 9, así que únicamente hay un resultado posible, al igual que en el segundo sólo puede ser 6, el tercer dígito puede ser cualquiera entre el 0 y el 9, así que existen 10

resultados posibles, el cuarto puede ser cualquiera menos el tercero, y así sucesivamente hasta el séptimo, de tal modo que la respuesta es:

Respuesta: $1 \times 1 \times 10 \times 9 \times 8 \times 7 \times 6 = 30.240$ marcas

Ejemplo 3.26: Si se lanza el mismo dado tres veces, ¿Cuántos son los resultados posibles?

Respuesta: $6 \times 6 \times 6 = 216$, es decir, hay 216 posibilidades de composición.

Ejemplo 3.27: Si se lanza un dado tres veces, imponiendo la condición de que todos los números sean diferentes en cada lanzamiento, ¿Cuántos son los resultados posibles?

Respuesta: $6 \times 5 \times 4 = 120$, en cuyo caso hay 120 opciones diferentes.

Ejemplo 3.28: Si se lanza un dado tres veces ¿Cuál es la Probabilidad de que todos los números sean diferentes?

El número posible de casos totales está dado por $6 \times 6 \times 6 = 216$

El número de casos favorables para A (todos los números son diferentes) se calcula de la siguiente manera: $6 \times 5 \times 4 = 120$ posibilidades.

De esta manera, la probabilidad de que todos los números sean diferentes al lanzar un dado tres veces, es $P[A] = \frac{120}{216} = \frac{5}{9}$

Ejemplo 3.29: Si se lanza un dado tres veces ¿Cuál es la Probabilidad de que el primer número sea seis y los demás cualquier número?

La solución se puede abordar como sigue:

Evento *A*: En el primer lanzamiento se obtiene 6 y en los demás cualquier número

Número total de casos: 6 × 6 × 6 = 216

Número de resultados favorables para *A*: 1 × 6 × 6 = 36

En consecuencia, la Probabilidad de que, al lanzar un dado tres veces, en el primer lanzamiento se obtenga un 6 y en los otros cualquier número, es $P[A] = \frac{36}{216} = \frac{1}{6}$

Ejemplo 3.30: El tema de la síntesis de proteína, es complejo, pero por ahora, es suficiente con indicar brevemente algunos aspectos de importancia. Para una revisión profunda del tema, pueden consultarse, entre muchas obras, las de Suzuki et al (1993), Darnell et al (1986) y Wallace et al (1991).

En la síntesis de proteínas, los aminoácidos se encadenan para formar péptidos y luego cadenas polipetídicas, es decir, una proteína. Para que este mecanismo funcione, se hace necesario la intervención del ADN y el ARN.

El ADN o Ácido Desoxirribonucleico, es una macromolécula de doble hélice compuesta de *un Glúcido* con cinco átomos de Carbono, una pentosa llamada *Desoxirribosa, un grupo fosfato* y *cuatro bases nitrogenadas* que son *Adenina (A*), *Guanina (G), Timina (T)* y *Citosina (C).* La unión del azúcar, la base nitrogenada y el fosfato, recibe el nombre de *Nucleótido*. El ARN o Ácido Ribonucleico se sintetiza a partir de una de las dos

cadenas de ADN y en su estructura la Desoxirribosa es reemplazada por la *Ribosa* y la Timina por el *Uracilo (U)*.

En el código molecular ADN-ARN, el *orden* en que aparezcan las bases nitrogenadas es de mucha importancia, puesto que, según la secuencia en que se ubiquen los *Nucleótidos*, se sintetizará una u otra proteína; de hecho, las mutaciones puntuales se originan por alteración en el orden de las bases.

Cada segmento de ARN compuesto por tres nucleótidos, no necesariamente distintos, forman una "palabra" denominada técnicamente *Codón*, que es específica para transportar o codificar a un determinado aminoácido, así por ejemplo la palabra *UUU* codifica para el aminoácido Fenilalanina, *AUG* para la metionina, *UUA* para la Leucina, *AUU* para la Isoleucina, *ACU* para la Treonina, *GGG* para la Glicina y así existen secuencias específicas para transportar los 20 aminoácidos, incluyendo, además, señales al inicio y terminación de la síntesis.

Con base en las anteriores explicaciones ¿Cuántas palabras de tres letras para codificar aminoácidos, se pueden formar en el código genético?

Respuesta: $4 \times 4 \times 4 = 64$

Efectivamente, 64 palabras codifican para los 20 aminoácidos e incluyen las señales de inicio y parada de la síntesis de proteína y constituyen el denominado ***Código Genético.***

Es importante aclarar que, a pesar de la existencia, como ya se dijo, de 20 aminoácidos que son: Alanina, Arginina, Asparagina,

Aspartato, Cisteína, Fenilalanina, Glicina, Glutamato, Glutamina, Histidina, Isoleucina, Lisina, Metionina, Prolina, Serina, Tirosina, Treonina, Triptófano y Valina, existen 64 codones, hecho que indica, que hay más de un codón para codificar el mismo aminoácido. Así por ejemplo, para la Leucina los codones son CUU, CUC, CUA, CUG.

¿Cuántas palabras que empiecen por U se pueden formar?

Respuesta: $1 \times 4 \times 4 = 16$.

¿Cuál es la Probabilidad de encontrar Nucleótidos que tengan alguna base repetida?

El procedimiento para responder la pregunta es el siguiente:

a. En primer lugar se define el evento:
$A = \{X \mid X \text{ es un nucleótido con alguna base repetida}\}$

b. Determinar el número de casos en que ninguna base se repite:
$4 \times 3 \times 2 = 24$

c. Haciendo uso de la Propiedad del Evento Complementario, se puede establecer el número de Nucleótidos con alguna base repetida, así:

$64 - 24 = 40$

e. Calcular la probabilidad con (3.7.1):

$P[A] = \frac{40}{64} = 0{,}625$

Los ejemplos, hasta ahora presentados, son casos sencillos, donde se permiten o no repeticiones. En otras situaciones, puede existir un mayor grado de complejidad, que hace necesario la aplicación de otros métodos que se explican a continuación.

Permutaciones. Una Permutación de ***n*** objetos tomados de ***r*** en ***r***, es una ***elección donde importa el orden*** en que *r* objetos se toman de entre *n*. El número de Permutaciones se denota como ${}_nP_r$ o $P_{(n,r)}$ o también $P_{n,r}$, que se lee "número de permutaciones de *n* elementos tomados de *r* en *r*" y está dado por la fórmula general:

$$ {}_nP_r = \frac{n!}{(n-r)!} \qquad (3.5) $$

Ejemplo 3.31: ¿Cuántas Permutaciones son posibles con las letras a, b, c tomadas de dos en dos?

De acuerdo con (3.7.5) ${}_3P_2 = \frac{3!}{(3-2)!} = \frac{3 \times 2 \times 1}{1} = 6$ Permutaciones son posibles

Ejemplo 3.32: ¿Cuántas Permutaciones son posibles con las letras a, b, c tomadas de tres en tres?

${}_3P_3 = \frac{3!}{(3-3)!} = 3! = 6$ es el número de Permutaciones posibles

A partir del Ejemplo 3.32, es posible generalizar una regla de la siguiente manera: ***¡El número de Permutaciones de n objetos tomados de n en n es igual a n!***

Ejemplo 3.33: ¿Cuántas Permutaciones se pueden formar con las letras de la palabra microorganismos?

En este ejemplo se presenta una situación diferente a las estudiadas en los Ejemplos 3.31 y 3.32, puesto que ahora hay ***n*** objetos, pero n_1 son de un tipo, n_2 de otro y así hasta n_k. La palabra tiene 15 letras (objetos), de las cuales tres son ***o***, dos son ***i***, dos son ***m***, dos son ***r***, dos son ***s***, y sólo una letra para ***c***, ***g***, ***n, a***.

Así, el número de Permutaciones será:

$$P = \frac{15!}{3! \times 2! \times 2! \times 2! \times 1! \times 1! \times 1! \times 1!} = 13.621.608.000$$

Ejemplo 3.34: ¿Cuántas Permutaciones son posibles con las letras de la palabra "casas"?

Este caso es idéntico al del ejemplo inmediatamente anterior y se calcula el número de Permutaciones de la siguiente manera:

$$P = \frac{5!}{2! \times 2! \times 1!} = 30$$

De los Ejemplos 3.33 y 3.34, se puede formular la siguiente generalización: ***El número de permutaciones de n objetos, donde n_1 son de un tipo, n_2 de otro tipo y n_k de otro diferente, y además $n_1 + n_2 + \ldots + n_k = n$ está dado por la siguiente expresión:***

$$P = \frac{n!}{n_1! \times n_2! \times \ldots \times n_k!} \qquad (3.6)$$

Combinaciones. Una Combinación de ***n*** objetos tomados de ***r*** en ***r,*** es una selección de ***r*** de ellos, ***sin importar el orden*** en que se toman. La notación general para las Combinaciones, es la siguiente:

$${}_nC_r = \binom{n}{r} = \frac{n!}{r!\,(n-r)!} \qquad (3.7)$$

Ejemplo 3.35: ¿Cuántas Combinaciones son posibles con las letras a, b, c tomadas de dos en dos?

La respuesta sobre el número de Combinaciones se plantea con la aplicación de la fórmula (3.7):

$$_3C_2 = \binom{3}{2} = \frac{3!}{2!\,(3-2)!} = \frac{3 \times 2 \times 1}{2 \times 1\,(1)} = \frac{6}{2} = 3$$

Ejemplo 3.36: En un corral hay 20 novillos Brahaman, se escogen siete para ser comercializados. ¿De cuántas maneras se pueden escoger los siete novillos?

Aplicando (3.7)

$$_{20}C_7 = \binom{20}{7} = \frac{20!}{7!\,(20-7)!} = \frac{20 \times 19 \times 18 \times 17 \times 16 \times 15 \times 14}{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1} = 77.520$$

maneras de seleccionar

Ejemplo 3.37: Se dispone de cinco novillas Shorthorn y de siete Angus; se escogen al azar dos novillas Shorthorn y tres Angus para un experimento. ¿De cuántas formas se puede hacer esta elección?

Las novillas Shorthorn se pueden escoger de $_5C_2$ maneras y las Angus de $_7C_3$.

El número total de maneras está dado por $_5C_2 \times {}_7C_3$

$$\binom{5}{2}\binom{7}{3} = \frac{5!}{2!\,(5-2)!} \times \frac{7!}{3!\,(7-3)!}\ 10 \times 35 = 350$$

Existen 350 formas, como se pueden seleccionar las novillas de las dos razas.

Ahora ya es posible resolver el **Ejemplo 3.20**, que decía: En un refrigerador se encuentran almacenados 50 vasos con yogur, de los cuales 10 no son aptos para su comercialización, porque han sobrepasado la fecha de vencimiento y 40 están en perfecto estado, sin que se pueda distinguir, a simple vista, unos de otros. Se eligen seis vasos al azar, sin devolver al refrigerador el vaso escogido antes de elegir el próximo y se pregunta ¿Cuál es la Probabilidad, de que la mitad de los vasos seleccionados no sean aptos para su comercialización? La solución contempla los siguientes pasos:

- En primer lugar, se establece el total de formas para escoger seis de entre 50 vasos. Evidentemente, se trata de una *Combinación*, ya que no interesa el orden en que los vasos son elegidos. En este caso, aplicando la fórmula (3.7) se tiene:

$$_{50}C_6 = \binom{50}{6} = \frac{50!}{6!\,(50-6)!} = 15.890.700$$

- En segundo lugar, se debe conocer el número de resultados, que cumplen con la condición de tener el 50% de los vasos no aptos para consumo. Se sabe que entre los 50 vasos hay 10 no aptos para consumo, de los cuales se escoge en total seis, lo que implica que, en esa muestra, la mitad serían no aptos, así que hay $_{10}C_3$ maneras de seleccionarlos:

$$_{10}C_3 = \binom{10}{3} = \frac{10!}{3!\,(10-3)!} = 120$$

- Finalmente, se debe establecer el número de resultados que cumplen con la condición de tener el 50% de los vasos aptos

para consumo. Esto implica, que hay ${}_{40}C_3$ maneras de escoger vasos con esta condición:

$${}_{40}C_3 = \binom{40}{3} = \frac{40!}{3!\,(40-3)!} = 9.880$$

Es importante anotar, que el número total de maneras de escoger los seis vasos de entre 50, al igual que el número de maneras de escoger los tres vasos aptos de entre 40 y los tres vasos no aptos de entre 10 son, en los tres casos, eventos ***igualmente probables***. En consecuencia, definido $A = \{3 \text{ vasos aptos, } 3 \text{ vasos no aptos}\}$, la Probabilidad de la ocurrencia de A está dada por (3.1):

$$P[A] = \frac{\binom{10}{3}\binom{40}{3}}{\binom{50}{6}} = \frac{120 \times 9.880}{15.890.700} = \frac{1.185.600}{15.890.700} = 0{,}0746$$

Del Ejemplo 3.37 y del desarrollo del Ejemplo 3.20, se pueden extraer las siguientes ***generalizaciones:***

- *Cuando se dispone de N objetos y se eligen* ***al azar, sin reemplazo, n*** *de esos objetos habrán* $\binom{N}{n}$ *maneras de escogerlos,* ***todas con la misma Probabilidad de ser elegidas.***

- *Si N objetos están constituidos por* $\boldsymbol{r_1}$ *y* $\boldsymbol{r_2}$ *clases de objetos que satisfacen la condición* $\boldsymbol{r_1 + r_2 = N}$*, de donde se eligen al azar* ***n*** *de ellos, la Probabilidad de que esos* ***n*** *objetos escogidos al azar sin reemplazo* ***contengan un evento particular*** *A, con elementos de* $\boldsymbol{r_1}$ *y* $\boldsymbol{r_2}$*, elegidos de* $\boldsymbol{L_1}$ *maneras de entre* $\boldsymbol{r_1}$ *y* $\boldsymbol{L_2}$ *maneras de entre* $\boldsymbol{r_2}$*, donde* $\boldsymbol{L_2 = n - L_1}$*, está dada por*:

$$P[A] = \frac{\binom{r_1}{L_1}\binom{r_2}{L_2}}{\binom{N}{n}} \qquad (3.8)$$

La fórmula (3.7.8) puede extenderse a los casos donde ***N*** esté constituida por $r_1, r_2, \ldots, r_k$ clases de objetos y calcular la Probabilidad de un evento particular *A* que contenga elementos de $r_1, r_2, \ldots, r_k$ escogidos de $L_1, L_2, \ldots, L_k$ maneras, donde deben satisfacerse las siguientes condiciones:

$r_1 + r_2 + \ldots + r_k = N$

$L_1 + L_2 + \ldots + L_k = n$

$$P[A] = \frac{\binom{r_1}{L_1}\binom{r_2}{L_2}\cdots\binom{r_k}{L_k}}{\binom{N}{n}} \qquad (3.9)$$

Ejemplo 3.38: En un refrigerador de un laboratorio están depositados 30 frascos idénticos, que contienen medios para cultivos celulares, de los cuales, 15 contienen MME (Medio Mínimo Esencial), 9 RPMI y 6 un medio de cultivo ideado en el laboratorio. Se escogen al azar 6 frascos y se pide calcular las siguientes Probabilidades:

- ¿Cuál es la Probabilidad de que se hayan extraído 4 MME, 1 RPMI y 1 del laboratorio?

$A = \{4\ MME, 1\ RPMI, 1\ laboratorio\}$ $\quad r_1 = 15 \quad L_1 = 4$

$N = 30 \quad r_2 = 9 \quad L_2 = 1$

$n = 6 \quad r_3 = 6 \quad L_3 = 1$

$$P[A] = \frac{\binom{15}{4}\binom{9}{1}\binom{6}{1}}{\binom{30}{6}} = \frac{(1.365)\ (9)\ (6)}{593.775} = 0{,}1241$$

De este ejemplo es posible indicar, la siguiente generalización: $\binom{n}{1} = n$

- ¿Cuál es la Probabilidad de que todos sean MME?

$$P[A] = \frac{\binom{15}{6}\binom{9}{0}\binom{6}{0}}{\binom{30}{6}} = \frac{(5.005)\ (1)\ (1)}{593.775} = 0{,}008429$$

El caso general que se puede deducir de este ejemplo, es el siguiente: $\binom{n}{0} = 1$

Cuando se selecciona objetos al aza, es fundamental especificar, si éstos se escogen con o sin reemplazo; en los casos en los cuales la selección es con reemplazo o con sustitución, se indica que un objeto seleccionado tendrá otra oportunidad de ser elegido en un muestreo siguiente y, por lo tanto, interesa el orden en que se hace la elección.

Cuando la selección se hace sin reemplazo, no existe interés en el orden en que se escogen los objetos, así que el número total de muestras (NM), de tamaño ***r***, que se pueden obtener de una población, bajo esta condición, está dado por $\binom{N}{r}$

Cuando el muestreo es con reemplazo, el número total de muestras (NM), de tamaño *r*, que se pueden obtener de una población, está dado por N^r.

Ejemplo 3.39: En un bioterio hay 20 ratones (*Mus musculus*). Un veterinario necesita escoger 4 individuos, para inocularlos con una cepa rugosa de *Brucella abortus* ¿Cuántas muestras

conformadas por 4 ratones son posibles, si el muestreo se hace con reemplazo?

$NM = N^r = 20^4 = 160.0000$

¿Cuántas muestras conformadas por 4 ratones son posibles, si el muestreo se hace sin reemplazo?

$$NM = \binom{20}{4} = \frac{20!}{4!(20-4)!} = 4.845$$

En la práctica, tanto en Producción como en Salud Animal, cuando se trata de escoger unidades experimentales para investigación, inspección o análisis de animales, insumos o productos, las muestras se obtienen sin reemplazo.

Cálculo de Probabilidades cuando los resultados no son equiprobables

En situaciones como éstas, lo más apropiado, es calcular la Probabilidad de un evento particular, en términos de ***frecuencia relativa***, tal como se indicó al inicio del capítulo.

Ejemplo 3.40: Un laboratorio lanza al mercado una nueva vacuna contra la fiebre aftosa. En total fueron vacunados 2.600 animales, de los cuales 1.680 no desarrollaron la enfermedad. Si se pregunta ¿Cuál es la Probabilidad de que un animal vacunado no desarrolle la enfermedad?

$$P[A] = \frac{1.680}{2.600} = 0{,}64$$

El Ejemplo 3.40 se puede generalizar, para cualquier situación en que un experimento es repetido varias veces y se registren los resultados observados.

Bajo esta condición, la Probabilidad aproximada de que un suceso A se produzca, está dado por:

$$P[A] \approx \frac{f}{n} = \frac{Número\ de\ veces\ que\ ocurre\ A}{Número\ de\ veces\ que\ se\ realiza\ el\ experimento}$$

(3.10)

Otras reglas para calcular Probabilidades

En el cálculo de Probabilidades, se requiere la aplicación de otras reglas que se resumen a continuación.

Regla General de la Adición. Si A y B son dos sucesos cualesquiera, la Probabilidad de que ocurra el uno o el otro está dada por:

$$P[A \text{ o } B] = P[A] + P[B] - P[A \text{ y } B] \qquad (3.11)$$

Donde $P[A \text{ y } B] = P[A \cap B]$ y significa que A y B pueden ocurrir juntos. En esta clase de problemas, resulta clave, para saber si se aplica o no (3.7.11), la utilización de la letra **"o"** cuando se pide una Probabilidad específica.

Ejemplo 3.41: Se lanza un dado una vez. ¿Cuál es la probabilidad de obtener un 2 o un 3? Si se define A = {Obtener el número 2} y B = {Obtener el número 3}, la respuesta está dada por:

$$P[A \text{ o } B] = P[A] + P[B] - P[A \cap B] = \frac{1}{6} + \frac{1}{6} - 0 = \frac{2}{6} = \frac{1}{3}$$

En este ejemplo, la intersección, es el conjunto vacío, ya que al lanzar un dado no es posible obtener un resultado 2 y 3 al mismo tiempo.

En general*, la Probabilidad de ocurrencia de uno o de otro suceso mutuamente excluyente, es igual a la suma de las Probabilidades individuales de cada suceso*.

Ejemplo 3.42: Se estima que, en el departamento de Nariño, el 30% de las vacas en producción padecen de hipocalcemia (*A*), el 3% sufren de cetosis (*B*) y el 2% son hipocalcémicas y sufren cetosis (*A* y *B*). Si se escoge una vaca al azar, ¿Cuál es la Probabilidad de que, el animal seleccionado, sea hipocalcémico o tenga cetosis?

$P[A \text{ o } B] = P[A \cup B] - P[A \cap B] = 0{,}3 + 0{,}03 - 0{,}02 = 0{,}31$

En este ejemplo los sucesos no son mutuamente excluyentes, ya que los dos pueden ocurrir juntos, es decir, existe la Probabilidad de que una vaca sea hipocalcémica y al mismo tiempo tenga cetosis; por lo tanto, hay intersección y se debe restar esta Probabilidad.

Regla de la Multiplicación. La Probabilidad de que dos sucesos independientes ocurran simultáneamente, es igual al producto de sus respectivas Probabilidades.

Dos sucesos *A* y *B* son independientes si y sólo si, cumplen con la condición:

$$P[A \cap B] = P[A] \times P[B] \qquad (3.12)$$

Ejemplo 3.43: Se lanza un dado dos veces, ¿Cuál es la probabilidad de obtener el número 4 en el primer lanzamiento y el 5 en el segundo? Definido $A = \{4 \text{ en el primer lanzamiento}\}$, $B = \{5 \text{ en el segundo lanzamiento}\}$:

$P[A \text{ y } B] = P[A \cap B] = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

Aquí la clave para saber si se aplica o no (3.7.12), es la utilización de la letra **"y"** cuando se pide una Probabilidad específica.

En general, dos sucesos se definen como independientes, cuando la ocurrencia de uno no afecta la ocurrencia de otro.

Ejemplo 3.44: Se escogen individuos al azar y se les pregunta sobre su tipo de sangre y edad. Como se puede observar, los dos eventos, tipo de sangre y edad, son independientes. Conociendo que la Probabilidad de encontrar personas con sangre tipo A es 0,39 y con más de 30 años de edad es 0,42. ¿Cuál es la probabilidad de encontrar un individuo con sangre tipo A y más de 30 años de edad?

Definidos: A = {Es el evento Tiene sangre tipo A}. $P[A] = 0{,}39$

B = {Es el evento Tiene más de 30 años}. $P[B] = 0{,}42$

$$P[A \cap B] = P[A] \times P[B] = 0{,}39 \times 0{,}42 = 0{,}1638$$

Probabilidad Condicional

Muchas veces es necesario encontrar la Probabilidad de un evento A, sabiendo que ha ocurrido antes un evento B que afecta la ocurrencia de A. En este caso, el evento B es un Espacio Muestral nuevo. Esta podría denominarse *Probabilidad del evento A dado que el evento B ha ocurrido* $P[A|B]$, o simplemente *Probabilidad de A dado B*, que se define como:

$$P[A|B] = \frac{Número \quad de \quad casos \quad favorables \quad a \quad [A \text{ Ç } B]}{Número \quad de \quad casos \quad favorables \quad a \quad B}$$

Sean A y B dos eventos de un Espacio Muestral S, donde $P[B] > 0$, la Probabilidad Condicional de que ocurra el evento A dado que el evento B ya ocurrió, se simboliza por $P[A|B]$ y se define como:

$$P[A|B] = \frac{P[A \cap B]}{P[B]}, P[B] > 0 \qquad (3.13)$$

Ejemplo 3.45: Estudios demostraron que, los ejemplares de una cierta raza de liebres de alta montaña mueren antes de lo normal, aún en ausencia de depredadores o de enfermedad conocida alguna. Dos de las causas de muerte identificadas son: baja cantidad de azúcar en la sangre y convulsiones. Se estima que el 7% de los animales presenta ambos síntomas, el 40% tienen bajo nivel de azúcar en la sangre y el 25% sufre convulsiones. ¿Cuál es la Probabilidad de que un animal elegido al azar, tenga bajo nivel de azúcar en la sangre, dado que sufre de convulsiones?

El interés se centra en los siguientes eventos:

C = {el animal escogido al azar tiene convulsiones}
S = {el animal escogido al azar tiene bajo nivel de azúcar en la sangre}

Del enunciado de este ejemplo se tiene que:

$P[C] = 0{,}25$
$P[S] = 0{,}40$
$P[C \cap S] = 0{,}07$

$$P[S|C] = \frac{P[S \cap C]}{P[C]} = \frac{0{,}07}{0{,}25} = 0{,}28$$

La Probabilidad de que un animal elegido al azar sufra convulsiones, dado que tenga bajo nivel de azúcar en la sangre, es:

$$P[C|S] = \frac{P[C \cap S]}{P[Z]} = \frac{0{,}07}{0{,}40} = 0{,}175$$

En el anterior ejemplo se puede apreciar que: $P[A|B] \neq P[B|A]$

Por lo general, esto se cumple si A y B son eventos de un Espacio Muestral S, con:

$P[A] > 0$ y $P[B] > 0$

En muchos casos, el interés puede reducirse solamente a un subconjunto de la población inicial. Si se define $n[A \cap B]$ como el número de casos favorables a $[A \cap B]$, $n[B]$ como el número de casos favorables a $[B]$ y N como el número total de casos, (3.13) se puede expresar así:

$$P[A|B] = \frac{n[AÇB]}{n[B]} = \frac{\frac{n[AÇB]}{N}}{\frac{n[B]}{N}} = \frac{P[AÇB]}{P[B]}, \quad P[B] > 0 \qquad (3.14)$$

Donde $P[A \cap B]$ y $P[B]$ se determinan a partir del espacio original S.

Algunas veces hay interés en determinar la Probabilidad de que ocurran conjuntamente dos eventos A y B, es decir, la Probabilidad de que ocurra el evento A y el evento B.

Si en un experimento pueden ocurrir los eventos A y B y éstos son sucesos de un Espacio Muestral S, entonces:

$$P[A \cap B] = P[A] \times P[B|A] \qquad (3.15)$$

Ejemplo 3.46: En un recipiente hay tres bolas blancas y dos negras. Se extraen dos bolas al azar, primero una y luego otra, sin reemplazo, ¿Cuál es la Probabilidad de que las dos bolas sean negras?

Se define: $A = \{La\ primera\ bola\ extraída\ es\ negra\}$
$B = \{La\ segunda\ bola\ extraída\ es\ negra\}$

- $A \cap B$ es el evento de que ocurra A y luego B.
- La Probabilidad de extraer en primer lugar una bola negra es $\frac{2}{5}$.
- La Probabilidad de extraer una segunda bola negra de las que restan es $\frac{1}{4}$, ya que el Espacio Muestral original se ha modificado, puesto que en el recipiente quedan cuatro bolas, de las cuales una es negra.

Entonces:

$$P[A \text{ Ç } B] = P[A] \times P[B|A] = \frac{2}{5} \times \frac{1}{4} = \frac{2}{20} = \frac{1}{10}$$

Ejemplo 3.47: En la replicación del ADN, a veces se presentan errores que pueden dar lugar a mutaciones observables en el organismo. En ocasiones, tales errores están inducidos químicamente. En un experimento se expuso un cultivo de bacterias a la presencia de un producto químico, que tiene un 40% de Probabilidad de producir error; sin embargo, el 65% de los errores es silencioso, en el sentido, de no dar lugar a una mutación observable. ¿Cuál es la Probabilidad de que se observe una colonia mutada?

Sean los eventos:

A: inducir a error.
B: dar lugar a una mutación observable.

La Probabilidad de que se observe una colonia mutada, es equivalente a la Probabilidad de que se induzca al error y que se produzca o dé lugar a una mutación observable, es decir:

P[observar una colonia mutada] = $P[A \cap B]$

$P[A \cap B] = P[A] \cdot P[B|A] = 0{,}40 \times (1 - 0{,}65)$

$= 0{,}40 \times 0{,}35 = 0{,}14$

A partir de este resultado se puede concluir que, la Probabilidad de observar una colonia mutada, es de 0,14; también se puede expresar que, de 100 colonias observadas, en 14 se puede observar mutaciones.

Ejemplo 3.48: En un experimento con cabras (*Capra hircus*), se evalúa la variable incremento de peso en función del sexo y de dos dietas, una con la adición de metionina y otra sin este aminoácido. El Espacio Muestral puede expresarse así:

	Hembras	**Machos**	**Total**
Dieta 1	10	15	25
Dieta 2	12	13	25
Total	22	28	50

Se elige una cabra al azar; ¿Cuál es la Probabilidad de que haya sido sometida a la dieta 2, dado que es macho? También es posible

formular la misma pregunta en estos términos: si se elige al azar un macho ¿Cuál es la Probabilidad de que haya sido sometida a la dieta 2?

Definidos los eventos:

A: la cabra elegida es sometida a la dieta 2.
B: el animal escogido es macho

$$P[A|B] = \frac{n[AÇB]}{n[B]} = \frac{13}{28} = 0{,}46$$

Como conclusión, se podrá decir que, la Probabilidad de que la cabra seleccionada haya sido sometida a la dieta 2, dado que es un macho, es de 0,46 o también que, de 100 animales de la especie *Capra hircus* que hayan sido sometidos a este experimento, 46 fueron machos sometidos a la dieta 2.

Ejemplo 3.49: Los bovinos de la raza Normando, poseen lunetas, es decir, una aureola de pelos pigmentados alrededor de los ojos, condición conocida también como "anteojo". Algunos animales no poseen tal condición y se los considera más susceptibles a padecer oftalmias. En una población de Normandos se ha calculado que, la Probabilidad de encontrar animales sin anteojos es 0,23 y, entre éstos, la Probabilidad de elegir un animal con oftalmia, es de 0,18. Si se escoge al azar un novillo sin anteojos de esa población, ¿Cuál es la Probabilidad de que tenga la enfermedad?

Los eventos son:

A: el novillo carece de anteojos.

B: el novillo tiene la enfermedad.

La Probabilidad de que el novillo elegido al azar tenga la enfermedad, dado que carece de anteojos, se la expresa como $P[B|A]$.

$$P[B|A] = \frac{P[BÇA]}{P[A]}$$

Puesto que $P[B \cap A] = 0,18$ y $P[A] = 0,23$, entonces:

$$P[B|A] = \frac{0,18}{0,23} = 0,7826$$

El resultado de este cálculo, lleva a concluir, que la Probabilidad de que un novillo de la raza Normando, elegido de la población en estudio, que padezca oftalmia, dado que no presenta lunetas, es 0,7826. También se podría decir, que el 78,26% de esa población de animales que no presenta lunetas, también padece de oftalmia.

Ejemplo 3.50: En un hato dedicado a la producción de leche, una vaca determinada, ha tenido tres partos, en uno de los cuales tuvo un hijo hembra. ¿Cuál es la Probabilidad de que el segundo hijo sea macho? Si se considera únicamente partos de un solo hijo, se tiene:

M: Es el evento de que el hijo sea macho.
H: Es el evento de que sea una hembra.

El Espacio Muestral *S* estará formado por los siguientes eventos:

S = {MMM, MMH, MHM, HMM, MHH, HMH, HHM, HHH}

Sea:

A: Evento de que la vaca señalada tenga por lo menos una cría hembra

Este evento es equivalente, a que la vaca tenga una ~~a~~ o más crías hembras y se expresa como:

$$A = \{\text{MMH, MHM, HMM, MHH, HMH, HHM, HHH}\}$$

Por otra parte, sea:
B: Evento de que el segundo hijo sea macho

Entonces:

$$B = \{\text{MMM, MMH, HMM, HMH}\}$$

Entonces la Probabilidad de que el segundo hijo sea macho, dado que esta vaca tiene por lo menos una cría hembra, se la expresa como:

$$P[B|A] = \frac{n[BÇA]}{n[A]}$$

$n[A] = 7$
$B \cap A = \{\text{MMH, HMM, HMH}\}$
$n[B \cap A] = 3$
Por lo tanto: $P[B|A] = \frac{3}{7} = 0{,}4286$

En conclusión, se podría expresar, como la Probabilidad de que una vaca tuvo tres partos, en uno de los cuales tuvo una hembra, tenga un macho en su tercer parto, es de 0,4286.

Eventos Independientes

Cuando se presentan dos eventos, la ocurrencia del primero puede, de alguna manera, alterar o no la probabilidad de ocurrencia del segundo; esto es, los eventos pueden ser dependientes o independientes. El concepto de Independencia Estadística es muy importante en las áreas de la Estadística Aplicada.

Sean A y B dos eventos de un Espacio Muestral S, se dice que los eventos A y B son independientes si y solo si:

$$P[A|B] = P[A] \text{ y } P[B|A] = P[B]$$

Teorema de Bayes

Este teorema se utiliza, para hallar la $P[A|B]$ cuando $P[A \cap B]$ y $P[B]$ no se conocen de inmediato. Es muy empleado en Teoría de la Decisión, para evaluar nueva información y revisar estimaciones anteriores de la Probabilidad, de que las cosas se encuentran en un estado o en otro.

Teorema de Bayes: Si $A_1, A_2, \ldots, A_n$ son n eventos mutuamente excluyentes, cuya unión es el conjunto universal, B es un evento arbitrario tal que $P[B] > 0$, $P[B|A_k]$ y $P[A_k]$ son conocidos para $k = 1, 2, \ldots, n$, entonces:

$$P[Ai|B] = \frac{P[Ai]\ P[B|Ai]}{P[A_1]P[B|A_1]+P[A_2]P[B|A_2]+\ldots+P[A_n]P[B|A_n]} \qquad (3.16)$$

Ejemplo 3.51: En un centro médico veterinario, se pudo establecer, que, en una población constituida por perros de la raza Labrador, el 90% de los animales de color negro son portadores del gen ***e***, el cual es hipostático y permite la dominancia del color amarillo, mientras que el 5% de los animales negros no son portadores de dicho alelo. Si la proporción de animales negros es de 0,45, ¿Cuál

es la Probabilidad de que un Labrador de color negro, seleccionado al azar, sea portador del alelo ***e***?.

Sean los eventos:

A_1: el perro Labrador es negro.
A_2: el perro labrador no porta el alelo ***e***.
B: el perro porta el alelo ***e***.

Según el enunciado del ejemplo, se tiene que:

$P[B|A_1] = 0{,}90$ $\quad\quad$ $P[A_1] = 0{,}45$
$P[B|A_2] = 0{,}05$ $\quad\quad$ $P[A_2] = 0{,}55$

El evento, el perro es negro dado que es portador del alelo ***e***, se simboliza como $P[A_1|B]$; según el Teorema de Bayes:

$$P[A_1|B] = \frac{P[A_1] \times P[B|A_1]}{P[A_1]P[B|A_1] + P[A_2]P[B|A_2]}$$
$$= \frac{(0{,}45) \times (0{,}90)}{(0{,}45)(0{,}90) + (0{,}55)(0{,}05)} = 0{,}94$$

De lo anterior se deduce que, la probabilidad de que el perro seleccionad sea negro, dado que es portador del alelo *e*, es de 0,94.

Ejemplo 3.52: Se estima que, la Probabilidad de que una vaca tenga parto distócico es de 0,00022. Si no es de raza Nellore, la Probabilidad de distocia es de 0,00017, mientras que si lo es ésta aumenta a 0,00064. Se supone que el 10% de los partos corresponden a vacas Nellore. Se pide hallar la Probabilidad de que una vaca que tenga parto distócico sea de la raza Nellore.

Sean los eventos:

A_1: la vaca no es de raza Nellore.
A_2: la vaca es de raza Nellore.
B: la vaca tiene parto distócico.

Según el enunciado del problema:

$P[B|A_1] = 0{,}00064$
$P[B|A_2] = 0{,}00017$
$P[A_1] = 0{,}10$
$P[A_2] = 0{,}90$

La Probabilidad de que una vaca que tenga parto distócico sea de raza Nellore se simboliza mediante $P[A_1|B]$ y según el Teorema de Bayes:

$$P[A_1|B] = \frac{P[A_1]\ P[B|A_1]}{P[A_1]P[B|A_1] + P[A_2]P[B|A_2]}$$

$$= \frac{(0{,}10)(0{,}00064)}{(0{,}10)(0{,}00064) + (0{,}9)(0{,}00017)} = 0{,}295$$

Con base en esto, se concluye que, la Probabilidad condicional de que una vaca que tenga parto distócico sea de la raza Nellore, es de 0,295.

La fibrosis quística es una enfermedad genética que afecta a los perros. La enfermedad se debe a una mutación en el gen CFTR (Cystic Fibrosis Transmembrane Conductance Regulator). Se pide determinar la

probabilidad de que un perro tenga la mutación del gen CFTR dado que su padre es portador de la mutación:

La información complementaria es la siguiente:

- La frecuencia del alelo normal (*N*) es del 96%, por lo tanto, $p = 0{,}96$.
- La frecuencia del alelo mutado (*n*) es del 4%, por tanto, $q = 0{,}04$.

Por el tipo de acción génica, un estudio genético determinó la probabilidad de que un perro tenga la enfermedad, si es homocigoto para la mutación *nn*, es del 100% ($P[E|mm] = 1$).

- El perro objeto de estudio, tiene un padre que es portador de la mutación, por lo tanto, su genotipo será *Nn*.

Se pide, calcular la Probabilidad de que el perro sea portador de la mutación, asumiendo que la madre es normal y no portadora, cuyo genotipo sería *NN*, caso en el cual, el genotipo del perro estudiado sería *Nn*, o que tenga la enfermedad, por lo que su genotipo sería *nn*.

Para calcular esa probabilidad, aplicando el teorema de Bayes, se deben hacer los siguientes cálculos, así, las probabilidades del perro en estudio serían:

- 50% de probabilidad de ser *Nn* (portador).
- 50% de probabilidad de ser *NN* (sano); pero, si la madre fuera también portadora, su genotipo sería *Nn* y, las probabilidades genotípicas del perro en estudio, serían:
    - ✓ 25% de probabilidad de ser *NN* (sano)
    - ✓ 50% de probabilidad de ser *Nn* (portador)
    - ✓ 25% de probabilidad de ser *nn* (enfermo)

Con el fin de brinda una mejor comprensión, se puede hacer un cuadrado de Punnet, para constatar las cifras indicadas.

Ahora se aplica el Teorema de Bayes (3.16):

$$P[Ai|B] = \frac{P[Ai]\ P[B|Ai]}{P[A_1]P[B|A_1] + P[A_2]P[B|A_2] + \ldots + P[A_n]P[B|A_n]}$$

$$P[Nn|Padre\ Nn] = \frac{P[Nn] \times P[Padre\ Nn|Nn]}{P[Padre\ Nn]}$$

Dado que el padre es *Nn*, la probabilidad de que el perro sea *Nn* es de 0,5 (por la Herencia Mendeliana),

Ahora, si se asume que el otro progenitor es *NN*, *P*[*Nn*|*Padre Nn*] sería de 0, si el otro progenitor es *NN*; pero, si el otro progenitor también es *Nn*, entonces *P*[nn|*Padre Nn*] sería de 0,25.

Así que, los resultados son estos:

- La Probabilidad de que el perro sea portador de la mutación (*Nn*), dado que su padre es portador y la madre es *NN*, es del 50%.
- La Probabilidad de que el perro tenga la enfermedad (*nn*), si ambos progenitores son *Nn*, es del 25%.

Teorema de Bayes

El Teorema de Bayes fue propuesto por el clérigo, filósofo y matemático inglés Thomas Bayes, para calcular la **Probabilidad Condicional** de un evento aleatorio *A*, disponiendo de información a priori sobre dicho suceso,

el cual se utiliza para calcular la $P[A|B]$, que es la probabilidad de A dado B, cuando no se conocen de inmediato $P[A \cap B]$ y $P[B]$.

Si A_1, A_2,…, A_n son n eventos mutuamente excluyentes, cuya unión es el conjunto universal, B es un evento arbitrario, tal que $P[B] > 0$, $P[B|A_k]$ y $P[A_k]$ son conocidos, para $k = 1, 2, \ldots, n$, entonces:

$$P[Ai|B] = \frac{P[Ai]\ P[B|Ai]}{P[A_1]P[B|A_1]+P[A_2]P[B|A_2]+\ldots+P[A_n]P[B|A_n]} \qquad (3.16)$$

El Teorema de Bayes es una poderosa herramienta utilizada intensamente en áreas como la medicina, la genética, la economía y muchas otras. A continuación. se presenta un ejemplo aplicado a la Ciencia Animal.

Ejemplo 3.52: Buitrago (2015) asegura que la Diarrea Viral Bovina (BDV) es una enfermedad de distribución mundial, causada por el BVDV, un virus RNA perteneciente al género Pestivirus y a la familia Flaviviridae.

El mismo autor, basando en varios estudios, reporta que, la prevalencia (BDV) en Colombia, oscila entre el 32 y el 59%, aproximadamente y que existen varios métodos de diagnóstico de esta enfermedad, siendo el ensayo inmuno enzimático (ELISA) uno de los más utilizados.

Para los fines relacionados con el desarrollo del ejemplo se asumirá que, la Prueba ELISA tiene una sensibilidad del 93% y una especificidad de 90%, siendo oportuno recordar que, la **sensibilidad se refiere a** la capacidad de la prueba, para identificar correctamente a los individuos **que tienen** la enfermedad y la **especificidad** a la capacidad de la prueba, para identificar correctamente a los individuos

que **no tienen la enfermedad**; por lo tanto, una prueba diagnóstica con alta sensibilidad **minimiza los falsos positivos**; una prueba con alta especificidad **minimiza los falsos positivos**.

Se pide calcular la probabilidad de que un animal tenga BDV, dado que resultó positivo al realizar el análisis con la prueba ELISA. Un procedimiento recomendable, para calcular probabilidades de este tipo, consiste en organizar la información así:

En primer lugar, se debe determinar los eventos que se encuentran en el problema, en este caso son dos, por lo tanto, la expresión indicada en (3.16), para este caso concreto adopta la siguiente forma:

$$P[A|B] = \frac{P[A]\ P[B|A]}{P[B]}$$

Donde, A y B son los eventos; $P[A|B]$ es la probabilidad de A dado B; $P[A]$ es la probabilidad de A; $P[B|A]$ es la probabilidad de B dado A.

Considerando una tasa de prevalencia de la enfermedad del 27%, la probabilidad que un animal esté enfermo de BDV, en la población en estudio, **será el evento *A*** y, su probabilidad de ocurrencia es 0,27, así que ***P*[*A*] = 0,27**, información *a priori* de mucha utilidad.

Para la prueba ELISA, se asumió una sensibilidad del 93% en la detección de BDV. Entonces, el **evento *B* es que un animal resulte positivo a la enfermedad** al realizar la prueba de laboratorio, así que ***P*[*B*]= 0,93**.

La **Especificidad** es la probabilidad de que la prueba sea negativa, si el animal no está enfermo, permitiendo calcular la tasa **de falsos positivos**. Puesto que la especificidad mide los negativos correctos, entonces los falsos positivos son simplemente su complemento, es decir 1 – 0,90=0,10.

Nótese que, es probable encontrar animales positivos a BDV, puesto que realmente están enfermos y son falsos positivos a la enfermedad, puesto que, la prueba no es perfecta en la detección. Como ya tenemos información de las probabilidades de A y B, mas el valor de la especificidad de la prueba, ahora es posible obtener todos los elementos de la fórmula para calcular la probabilidad pedida $P[B|A]$, es decir, la probabilidad de que un animal, que tenga la enfermedad, dado que resulto positivo en la prueba.

La probabilidad de que la prueba sea positiva, dado que el animal está enfermo, es la sensibilidad, que en la fórmula corresponde a $P[B|A]$, es decir, el animal es identificado como positivo, dado que resultó positivo en la prueba ELISA, faltando calcular la Probabilidad Ajustada de B, que incluye los verdaderos positivos y los falsos positivos.

La probabilidad de encontrar verdaderos positivos, se calcula multiplicado la sensibilidad de la prueba por la tasa de prevalencia, en este caso:

$(0{,}93 \times 0{,}27) = 0{,}2511$

Así, la probabilidad de encontrar falsos positivos se calcula multiplicando el valor porcentual de los falsos positivos, por el

valor porcentual de animales sanos en la población, es decir, uno menos la tasa de prevalencia, en este caso

$0,10 \times (1- 0,27) = 0,073$

Ahora ya se dispone de la probabilidad total de positivos, que es igual a:

$0,2511 + 0,073 = 0,3241$

Finalmente se aplica la fórmula (3.16):

$$P[A|B] = \frac{0,27 \times 0,93}{0,3241} = \frac{0,2511}{0,3241} = 0,7747$$

Es necesario observar que, la probabilidad de encontrar un animal enfermo, dado que resultó positivo a BDV con la prueba ELISA, es del 77,5%, aproximadamente, y no del 100%, como podría suponerse, debido a que, en la población, es posible encontrar animales realmente positivos y falsos positivos, puesto que la sensibilidad y la especificidad no son del 100%.

Es conveniente indicar que, en estos casos si, tanto la prevalencia, como la sensibilidad y la especificidad tiene valores bajos, la probabilidad de encontrar animales enfermos, dado que resultaron positivos en la prueba diagnóstica, será también baja y, por ende, se recomendaría realizar pruebas diagnósticas de mayor sensibilidad y especificidad. Por ejemplo, con una prevalencia de 1%, sensibilidad de la prueba diagnóstica del 80% y especificidad del 70%, la probabilidad de encontrar un animal enfermo dado que resultó positivo en la prueba es del 2,62%; es decir, es muy difícil declarar a un animal enfermo, solo porque resultó positivo en la prueba; lo correcto sería realizar una prueba diagnóstica con los más altos valores posibles de sensibilidad y especificidad.

CAPÍTULO 4

Funciones de probabilidad para variables aleatorias discretas y continuas

En los capítulos anteriores, se ha estudiado los principios básicos más importantes para el Cálculo de Probabilidades. En el presente capítulo, se buscará relacionar dichos conceptos, con el ***Análisis de Datos***, con la finalidad de extraer conclusiones aplicables a la población, a partir de la información obtenida de la muestra y, de este modo, concatenar la Teoría de Probabilidad con la Estadística Aplicada.

Para abordar con claridad esta temática, es necesario recordar el concepto de ***Variables Aleatorias***, también conocidas con el nombre de ***Variables Estocásticas*** y su clasificación en ***Continuas y Discretas,*** que en gran medida, determina la capacidad para ***predecir*** el valor que tomarán las variables bajo una determinada situación, sin perder de vista que, dada la condición del ***azar,*** las predicciones se hacen en términos de ***incertidumbre***, por lo que resulta más apropiado describir el comportamiento de las variables con las ***Funciones de Probabilidad y de Probabilidad Acumulada.***

VARIABLE ALEATORIA

Cuando se describe un Espacio Muestral de un experimento, un resultado particular, no es necesariamente un número. Por ejemplo, un animal en un

hato puede ser clasificado como "enfermo" o "sano"; en cambio, en otras situaciones experimentales, se realiza una observación y el resultado se registra como un número, es decir, se asigna un número real $\boldsymbol{x}$ a cada elemento de un Espacio Muestral $\boldsymbol{S}$.

De acuerdo con esto, una variable aleatoria $\boldsymbol{X}$ se define como una función que asocia un número real con cada elemento del Espacio Muestral. Es preciso aclarar que, en este libro se utiliza la letra mayúscula $\boldsymbol{X}$ para denotar una variable aleatoria y su correspondiente letra minúscula $\boldsymbol{x}$ para algunos de sus valores.

Rango de una Variable Aleatoria *X*

Se refiere al conjunto de los valores posibles, que puede tomar la variable, cuando se realiza un experimento particular.

Ejemplo 4.1: Un experimento consiste en lanzar dos monedas normales que tienen cara (*c*) y sello (*s*). Si *X* es la variable aleatoria *número de caras que aparecen*, el Espacio Muestral se lo puede expresar como: $S = \{cc, cs, sc, ss\}$.

La variable aleatoria *X* se asocia con el número 0 en el punto muestral *ss*, con el 1 en cada punto *cs* y *sc*, con el 2 en el punto muestral *cc*. Entonces, el rango de la variable aleatoria *X* es el conjunto {0, 1, 2}.

Ejemplo 4.2: Si el experimento consiste en lanzar dos dados corrientes, el Espacio Muestral viene dado $S = \{(1, 1), (1, 2), ..., (6, 6)\}$. Sea *X* la variable aleatoria que hace corresponder a cada punto (a, b) de *S* ***el máximo*** de sus números, es decir: $X_{(a, b)} = máximo_{(a, b)}$.

Así sería para los casos: $X_{(1, 5)} = 5$, $X_{(5, 6)} = 6$, $X_{(6, 6)} = 6$. Aquí el rango de *X* viene dado por el conjunto {1, 2, 3, 4, 5, 6}.

Variables Aleatorias Discretas

Una variable aleatoria X es discreta, cuando toma solamente un número finito o infinito numerable de valores, es decir, su rango es finito o infinito numerable.

Función de Probabilidad para Variables Aleatorias Discretas. Luego de definir una Variable Aleatoria X sobre el Espacio Muestral de un experimento, es necesario estimar la Probabilidad de que ésta tome un valor determinado, para lo cual, la función de Probabilidad hace corresponder a cada valor posible de X, el valor de Probabilidad de su ocurrencia dentro del experimento.

La Probabilidad de que la Variable Aleatoria X tome un valor determinado x, es igual a la suma de las Probabilidades de los puntos muestrales para los cuales $X = x$ y se simboliza por:

$p[x] = P[X = x]$, donde deben cumplirse las siguientes propiedades:

a. $p[x] \geq = 0$, para todo x de X.
b. $\sum p[x] = 1$

A la colección de pares $(x_i,\ p(x_i))$ se la denomina ***Distribución de Probabilidad de X*** y es posible representarla en forma de tabla.

Ejemplo 4.3: En el ejemplo 4.2 se quiere calcular $(x_i,\ p(x_i))$.

$S = \{(1, 1), (1, 2), ..., (6, 6)\}$
$X_{(a,\ b)} = max(a,\ b)$
$Rango = \{1, 2, 3, 4, 5, 6\}$

Se puede apreciar que la Probabilidad de cada punto muestral de S es $\frac{1}{36}$, detallada en la siguiente forma:

$p(1) = P[X = 1] = P[\{(1, 1)\}] = \frac{1}{36}$

$p(2) = P[X = 2] = P[\{(1, 2), (2, 2), (2,1)\}] = \frac{3}{36}$

$p(3) = P[X = 3] = P[\{(1, 3), (2, 3), (3, 3), (3, 2), (3, 1)\}] = \frac{5}{36}$

$p(4) = P[X = 4] = P[\{(1, 4), (2, 4), (3, 4), (4, 4), (4, 3), (4, 2), (4, 1)\}] = \frac{7}{36}$

En forma similar: $p(5) = P[X = 5] = \frac{9}{36}$ y $p(6) = P[X = 6] = \frac{11}{36}$

De estos cálculos podría interpretarse, para uno de los casos $p(3)$, que la Probabilidad de que la variable X tome un valor de 3 es de $\frac{5}{36}$; en el experimento, la Probabilidad de que los dos dados caigan en los números (1, 3), (2, 3), (3, 3), (3, 2) o (3, 1) es 0,1389.

Esta distribución de Probabilidad puede ser representada en el siguiente esquema:

x_i	1	2	3	4	5	6
$p(x_i)$	1/36	3/36	5/36	7/36	9/36	11/36

A su vez, estos datos pueden ser representados gráficamente, como se muestra en la Figura 4.1.

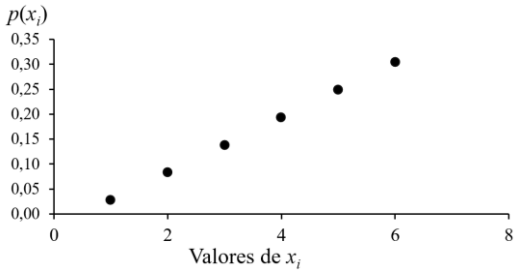


Figura 4.1 Representación Gráfica de la Distribución de Probabilidad.

Función de Distribución Acumulativa. La Función de Distribución Acumulativa, de la variable aleatoria *X,* es la Probabilidad de que *X* sea menor o igual a un valor específico *x* y está dado por:

$$F(x) = P[X \leq x] = \sum_{x_i \leq x} p(x_i) \qquad (4.1)$$

Ejemplo 4.4: Para el ejemplo 4.3:

$F(1) = P[X \leq 1] = p(1) = \frac{1}{36}$

$F(2) = P[X \leq 2] = p(1) + p(2) = \frac{1}{36} + \frac{3}{36} = \frac{4}{36}$

$F(3) = P[X \leq 3] = p(1) + \mathrm{p}(2) + \mathrm{p}(3) = \frac{9}{36}$

$F(4) = P[X \leq 4] = p(1) + p(2) + \mathrm{p}(3) + \mathrm{p}(4) = \frac{16}{36}$

$F(5) = P[X \leq 5] = p(1) + p(2) + p(3) + p(5) = \frac{25}{36}$

$F(6) = P[X \leq 6] = p(1) + p(2) + p(3) + p(4) + p(5) + p(6) = \frac{36}{36} = 1$

Como una de las conclusiones en este ejemplo, se puede decir que, la Probabilidad de que los dados caigan en un número menor o igual a 4 ($F(4)$), es $\frac{16}{36}$.

La Función de Distribución Acumulativa $F(x)$, de una variable aleatoria discreta *X*, cumple las siguientes propiedades:

a. $0 \leq F(x) \leq 1$, para todo x.

b. Si $x_i \geq x_j$, entonces $F(x_i) \geq F(x_j)$, lo que significa que es una función no decreciente.

Si se desea hallar la Probabilidad de que una variable aleatoria tome un valor mayor que x, se tiene:

c. $P[X > x] = 1 - F(x)$.

Para variables aleatorias de valor entero, se tiene además que:

d. $P[x_i \leq X \leq x_j] = F(x_j) - F(x_i - 1)$.
e. $P[X \geq x] = 1 - F(x - 1)$.

Ejemplos 4.5: Con los datos del ejemplo 4.3 se puede construir el siguiente esquema:

x_i	1	2	3	4	5	6
$p(x_i)$	1/36	3/36	5/36	7/36	9/36	11/36
$F(x)$	1/36	4/36	9/36	16/36	25/36	36/36

De allí se puede establecer que:

$$P[X > 3] = 1 - F(2) = 1 - \frac{4}{36} = \frac{32}{36}$$

La Probabilidad de que los dados caigan en números mayores que 3 es 0,8889.

$$P[3 \leq X \leq 5] = F(5) - F(2) = \frac{25}{36} - \frac{4}{36} = \frac{21}{36}$$

La Probabilidad de que los dados caigan en números entre 3 y 5 inclusive, es 0,5833.

Ejemplo 4.6: En una planta procesadora de carnes, se sabe por experimentación que, la demanda diaria de salami, en toneladas, es como se muestra enseguida:

Toneladas vendidas	3	4	5	6	7	8	9
Probabilidad	0,05	0,12	0,20	0,24	0,17	0,14	0,08

Con base en esta información, se pide calcular las siguientes Probabilidades:

a. ¿Cuál es la Probabilidad de que se vendan por lo menos siete toneladas?

b. ¿Cuál es la Probabilidad de que se vendan cuatro toneladas o más?

c. ¿Cuál es la Probabilidad de que se vendan como máximo cinco toneladas?

Si se define X como la variable aleatoria *número de toneladas* del producto que se vende, entonces:

a) $P[X \geq 7] = 1 - F(6) = 1 - [p(3) + p(4) + p(5) + p(6)]$

$= 1 - 0,61 = 0,39$

La Probabilidad de que se vendan, por lo menos siete toneladas al día, es 0,39.

b) $P[X \geq 4] = 1 - F(3) = 1 - 0,05 = 0,95$

La Probabilidad de que se vendan, cuatro toneladas de salami al día, es 0,95.

c) $P[X \leq 5] = F(5) = p(3) + p(4) + p(5) = 0{,}37$

La Probabilidad de que se vendan, como máximo cinco toneladas de salami, es 0,37.

Valor Esperado o Esperanza Matemática de una Variable Aleatoria Discreta

El Valor Esperado de una variable aleatoria se obtiene al calcular el valor medio de la variable, para todos sus valores posibles. El Valor Esperado es una media técnica o ideal.

El Valor Esperado de una variable aleatoria X se representa, generalmente, por $E[X]$ o la letra griega μ y se expresa como:

$$E[X] = \mu = \sum_{x \in X} Xp(x) \qquad (4.2)$$

El Valor Esperado posee propiedades análogas a las establecidas para la Media Aritmética:

a. Si c es una constante, entonces $E[c] = c$.
b. Si a y b son constantes, entonces $E[aX + b] = aE[X] + b$
c. Si $g(x)$ y $h(x)$ son funciones de una variable aleatoria X, entonces $E[g(x) \pm h(x)] = E[g(x)] \pm E[h(x)]$
d. $E[X - \mu] = 0$

Ejemplo 4.7: Hallar el Valor Esperado en el Ejemplo 4.6.

$$\mu = \mathrm{E}[X] = \sum_{i=3}^{9} Xi\ p(x_i) = (3 \times 0{,}05) + (4 \times 0{,}12) + (5 \times 0{,}2) + (6 \times 0{,}24) + (7 \times 0{,}17) + (8 \times 0{,}14) + (9 \times 0{,}08) =$$

$6{,}1 \approx 6$ toneladas de salami

Lo que significa, que en promedio se espera vender seis toneladas diarias de salami.

Varianza de una Variable Aleatoria

Si X es una variable aleatoria con valor esperado de $E[X]$, la Varianza de X se representa por $V[X]$ o σ^2 y se la puede expresar como:

$$V[X] = \sigma^2 = E[X^2] - (E[X])^2 = \sum X^2 \; p(x) - \mu^2$$

La raíz cuadrada positiva de la Varianza, es la desviación típica σ. La Varianza, como se indicó en el primer capítulo, mide la dispersión de los datos alrededor de la Media.

Propiedades análogas a las de la Varianza de observaciones se cumplen para la Varianza de una variable aleatoria:

a. Si a es una constante, entonces $V[a] = 0$.
b. Si a y b son constantes, entonces $V[aX + b] = a^2V[X]$.

Ejemplo 4.8: Con la información indicada en el Ejemplo 4.6, la Varianza se calcula de la siguiente manera:

$$\sigma^2 = \sum {X_i}^2 p(x) - \mu^2$$

$\mu = 6{,}1$, entonces $\mu^2 = 37{,}21$

$$\sum_{i=1}^{9} {X_i}^2 p(x_i) = (9\times 0{,}05) + (16 \times 0{,}12) + (25 \times 0{,}2) + (36 \times 0{,}24) + (49 \times 0{,}17) + (64 \times 0{,}14) + (81 \times 0{,}08) = 39{,}78$$

$\sigma^2 = 39{,}78 - 37{,}21 = 2{,}57$

DISTRIBUCIONES DE PROBABILIDAD ESPECIALES PARA VARIABLES DISCRETAS

La Distribución Binomial

La Variable Aleatoria Binomial es un tipo específico de variable discreta, que representa ***únicamente dos*** clases de respuesta: ***Éxito o Fracaso,*** obtenidos en un número ***n*** de ***pruebas independientes***. Los casos que se mencionan a continuación ilustran de mejor manera este enunciado:

- Si 1.000 partos no gemelares de vacas Holstein son sucesos independientes; sólo existen dos posibles resultados en cuanto al sexo de las crías: que sean machos o que sean hembras. Al obtener los resultados de todos esos partos, será posible contabilizar las crías que corresponden a cada sexo.

- Se vacunan 5.000 cerdos contra la PPA (Peste Porcina Africana). Hay dos posibles resultados: que los animales contraigan la enfermedad (no fueron protegidos por la vacuna) o que no contraigan la enfermedad (fueron protegidos por la vacuna); al final del ensayo, es posible contabilizar el número de animales sanos y enfermos. Se asume que la efectividad de la vacuna (Probabilidad de Protección) se mantiene constante de un ensayo a otro.

Generalmente, se define como "Éxito" la constatación de la característica de interés para el investigador; así, en el ejemplo de la vacuna, "Éxito" se refiere a la Probabilidad de encontrar un ***animal enfermo***.

Si las ***n*** pruebas son independientes, significa que el resultado de una no afecta el resultado de cualquier otra y, además, la probabilidad de Éxito

permanece constante a lo largo del experimento, es posible establecer la Función de Probabilidad de este tipo de variables, así:

$$P(x,n,p) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0,\ 1,\ 2,\ldots,\ n \qquad (4.3)$$

Donde:

p: Probabilidad de Éxito en una sola prueba.
$1 - p = q$: Probabilidad de Fracaso.
n: número de pruebas.
x: número de Éxitos en n pruebas.

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

El Valor Esperado de la variable aleatoria binomial es $E[X] = \mu = np$.

La Varianza es $V(X) = \sigma^2 = npq$.

La Función de Distribución Acumulativa de la Variable Aleatoria Binomial viene dada por:

$$F(x) = P[X \leq x] = p(0) + p(1) + \ldots + p(x) \qquad (4.4)$$

Ejemplo 4.9: Al lanzar una moneda tres veces, ¿Cuál es la Probabilidad de que aparezcan los siguientes resultados?:

a) 3 caras, b) 2 caras, c) 1 cara, d) ninguna cara, e) por lo menos 2 caras, f) como máximo 2 caras.

$n = 3$, ya que el experimento se repite 3 veces.

p = 0,5 es la Probabilidad de Éxito (obtener una cara) en un lanzamiento; la Probabilidad de Fracaso (q) se obtiene así: $q = 1 - p = 1 - 0{,}5 = 0{,}5$.

Los experimentos son independientes, ya que el resultado del primer lanzamiento no afecta el de los siguientes. En consecuencia, se cumplen los supuestos del modelo binomial; por lo tanto, es posible calcular las Probabilidades pedidas con la fórmula (4.2.1), donde, X es el número de caras que podrían aparecer, con lo cual es posible calcular las Probabilidades pedidas, de la siguiente manera:

a) $P(X = 3) = \binom{3}{3} 0,5^3 \times 0,5^{3-3} = \frac{3!}{3!(3-3)!} 0,5^3 \times 0,5^{3-3} = \frac{6}{6} \times 0{,}125 \times 1$

$= 0{,}125$

b) $P(X = 2) = \binom{3}{2} 0,5^2 \times 0,5^{3-2} = \frac{3!}{2!(3-2)!} 0,5^2 \times 0,5^{3-2}$

$= \frac{6}{2} \times 0{,}25 \times 0{,}5 = 0{,}375$

c) $P(X = 1) = \binom{3}{1} 0,5^1 \times 0,5^{3-1} = \frac{3!}{1!(3-1)!} 0,5^1 \times 0,5^{3-1}$

$= \frac{6}{2} \times 0{,}5 \times 0{,}25 = 0{,}375$

d) $P(X = 0) = \binom{3}{0} 0,5^0 \times 0,5^{3-0} = \frac{3!}{0!(3-0)!} 0,5^0 \times 0,5^3 = 0{,}125$

e) $P[X \geq 2] = p(2) + p(3) = 0{,}5$

Al aplicar las Propiedades de las Variables Aleatorias Discretas, que se pueden deducir a partir de la fórmula (4.4), se tiene:

$P[X \geq 2] = 1 - F(1) = 1 - [p(0) + p(1)] = 1\text{- } 0{,}5 = 0{,}5$

f) $P[X \leq 2] = F(2) = p(0) + p(1) + p(2) = 0{,}875$

Cuando el tamaño de muestra es grande, el cálculo de Probabilidades mediante la ecuación (4.4) es dispendioso, por lo tanto, se recurre al uso de valores que ya están tabulados para diferentes valores de *n, p, x*. En la Tabla 2 del Apéndice se presentan las cifras de la Función de Distribución Acumulativa Binomial.

Ejemplo 4.10: El cálculo de Probabilidad para el literal f) del Ejemplo 4.9, con el uso de la tabla allí mencionada, se lleva a cabo de la siguiente manera:

Se calcula $F(2)$, ya que $F(2) = P[X \leq 2]$.

En la Tabla 2 del Apéndice, en la primera columna se localiza $n = 3$ y en la segunda columna $x = 2$, luego $p = 0{,}5$ en la primera fila. En la intersección entre $x = 2$ y $p = 0{,}5$ y se encuentra $F(2) = 0{,}875$, como se puede observar en la siguiente muestra de un aparte del apéndice citado, para indicar su uso.

$$P(X \leq x) = F(x;\ n,\ p) = \sum_{i=0}^{x} \binom{n}{i} p^x (1-p)^{n-x}$$

		p										
n	***x***	**0,01**	**0,05**	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**0,35**	**0,40**	**0,45**	**0,50**
2	0	0,980	0,902	0,810	0,722	0,640	0,562	0,490	0,422	0,360	0,302	0,250
	1	0,999	0,997	0,990	0,977	0,960	0,937	0,910	0,877	0,840	0,797	0,750
	2	1,000	1,000	1,000	1,000	1,000	1,000	1	1,000	1,000	1,000	1,000
3	0	0,970	0,857	0,729	0,614	0,512	0,421	0,343	0,274	0,216	0,166	0,125
	1	0,999	0,992	0,972	0,939	0,896	0,843	0,784	0,718	0,648	0,574	0,500
	2	1,000	0,999	0,999	0,996	0,992	0,984	0,973	0,957	0,936	0,908	**0,87**
	3	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

En esta tabla también se puede hallar $P[X = x]$, con la siguiente fórmula:

$$P[X = x] = F(x) - F(x - 1) \qquad (4.5)$$

Para el literal b) del Ejemplo 4.9, aplicando la fórmula (4.5) se procede así:

$P[X = 2] = F(2) - F(1) = 0{,}875 - 0{,}5 = 0{,}3750$

Ejemplo 4.11: En una población de cerdos afectados de neumonía causada por Bortedella, enfermedad caracterizada por tos, disnea, debilitamiento, taquinea y alta mortalidad, el 40% se recupera después de recibir tratamiento a base de quinolonas. Si se extrae una muestra aleatoria de 20 cerdos, calcular las siguientes Probabilidades:

a) Que cinco o menos se recuperen.
b) Por lo menos seis se recuperen.
c) Entre 7 y 10 inclusive se recuperen.
d) ¿Cuál es el valor esperado de animales recuperados?

Se define X como la variable aleatoria "número de cerdos que se recuperan". Se tiene que $n = 20$ y $p = 0{,}4$. Consultando los valores consignados en la Tabla 2 del Apéndice, las respuestas serán:

a) $P[X \leq 5] = F(5) = 0{,}1256$.

La Probabilidad de que cinco cerdos o menos se recuperen es 0,1256.

b) $P[X \geq 6] = 1\text{-}\ F(5) = 1 - 0{,}1256 = 0{,}8744$.

La Probabilidad de que por lo menos seis cerdos se recuperen es 0,1256.

c) $P[7 \leq X \leq 10] = F(10) - F(6) = 0{,}8725 - 0{,}2500 = 0{,}6225$.

La Probabilidad de que se recuperen entre 7 y 10 cerdos es 0,6225.

d) $E[X] = np = 20 \times 0{,}4 = 8$, lo que significa que, en una muestra aleatoria de 20 cerdos, se espera que en promedio se recuperen ocho animales.

La Distribución Poisson

Es una familia de distribuciones discretas, llamada de esta forma en homenaje a Simeón Denis Poisson, quien la describió a finales del siglo XIX. Las Variables Aleatorias de Poisson, implican la observación de un conjunto ***discreto de sucesos*** en un ***intervalo continuo*** de tiempo, longitud o espacio; se advierte, que el concepto de intervalo no se trata, en este contexto, como en la forma matemática usual. Por ejemplo, si al analizar una muestra de 1,0 cc de semen bovino, se buscan espermatozoides con cola enrollada, el suceso

es el espermatozoide con la característica anotada y el "intervalo" es un centímetro cúbico de semen. En otro ejemplo se considera, un veterinario que contabiliza semanalmente el número de llamadas recibidas en su consultorio, solicitándole realizar cesáreas en equinos; en este caso, el suceso que toma valores discretos, es el número de llamadas y el intervalo es una semana.

En estos casos, se define ***X,*** como el número de sucesos de interés en un intervalo de unidades de tamaño ***S.*** Con el propósito de establecer la Función de Probabilidad de *X,* es necesario considerar los siguientes aspectos:

a. La unidad de medida básica del problema.
b. La media del número de sucesos ocurridos por unidad de medida
c. El tamaño del intervalo de observación.

Establecidos los anteriores criterios, la Función de Probabilidad de *X* estará dada por:

$$p(x) = \frac{e^{-\lambda s}(\lambda s)^x}{x!} \qquad (4.6)$$

$x = 1, 2, ..., k$

$\lambda > 0$

$s > 0$

Donde:

e: Base del logaritmo natural (2,718282)

$\lambda=\mu$: Media de la ocurrencia del suceso particular

s: Intervalo del experimento

σ^2: λ

Tal como se calcula la Función de Probabilidad Acumulada, la Distribución de Poisson está descrita por:

$$F(x) = P[X \leq x] = p(0) + p(1) + \ldots + p(x) \qquad (4.7)$$

Cuando una variable aleatoria discreta adopta esta forma, recibe el nombre de Variable Aleatoria Poisson y su Función de Probabilidad está expresada en (4.6).

Los anteriores conceptos se resumen, al indicar que una Distribución Poisson, describe la frecuencia de un evento que puede tomar valores discretos 0, 1, 2, 3, ..., *x,* correspondientes a un suceso, cuyo ***número promedio*** de ocurrencias ***es bajo***, en relación con el número ***total de veces*** que éste podría ocurrir, tal como se observa en los ejemplos antes anotados. Dado que la Probabilidad del suceso es pequeña, frecuentemente se menciona a la Distribución Poisson, como una Función de Distribución de Sucesos "raros".

Ejemplo 4.12: Luria y Delbruck, citados por Suzuki et al (1993), realizaron una prueba experimental, conocida con el nombre de ***Prueba de Fluctuación***, con células de *E. Coli,* para calcular la tasa de mutación. Un estudio específico, tenía como objetivo, determinar un tipo particular de mutación, consistente en detectar bacterias que se vuelven resistentes al Fago T1, virus que ataca y mata a la mayoría de las células de *E. coli*.

Cuando se siembra un número elevado de bacterias y de fagos, la mayoría de las bacterias muere, pero las resistentes al Fago T1 sobreviven y producen colonias que se pueden aislar. Este tipo de células mutantes, se conocen con el nombre de células resistentes a T1 o Ton^r. Con el método de Luria y Delbruck, se estableció que la media (μ) de los sucesos mutacionales por división celular era

de 3×10^{-8}. Si se observan 10^8 células, ¿Cuál es la Probabilidad de que ninguna sea mutante?

Para responder la pregunta, se aplica el siguiente procedimiento:

a. Calcular $\lambda = 3 \times 10^{-8}$, $s = 10^8$, $\lambda s = 3$

b. Aplicar la fórmula (4.6), teniendo en cuenta que el valor de $X = 0$.

$$P(X = 0) = \frac{e^{-3}(3)^0}{0!} = 0{,}04978$$

Como respuesta al problema planteado, se puede concluir que, la probabilidad de que en el espacio observado ninguna célula sea mutante, es 0,04978.

Ejemplo 4.13: Mediante análisis de cariotipo, se ha establecido que, una de las anomalías específicas en cromosomas sexuales, es el síndrome de Klinefelter, donde el individuo afectado presenta cariotipo anormal XXY, que en el fenotipo se expresa con disminución del tamaño de los testículos, hialinización de los túbulos seminíferos, escaso desarrollo de los caracteres sexuales secundarios, de tal suerte, que los pacientes tienen apariencia de varón, pero son positivos para la prueba de la cromatina sexual. Diversos estudios han estimado, una tasa promedio del síndrome en 1,2 por cada 1.000 nacimientos. Si se analizan 5.000 láminas con preparados celulares para establecer el cariotipo, ¿Cuál es la Probabilidad de encontrar la anormalidad en tres de ellas?

Se procede de la misma manera que en el ejemplo inmediatamente anterior:

a. $\lambda s = 0{,}0012 \times 5.000 = 6$
b. Aplicar la fórmula (4.6)

$$P(X = 3) = \frac{e^{-6}(6)^3}{3!} = 0{,}089$$

En este caso se concluye que, al observar 5.000 láminas con cariotipos, la Probabilidad de encontrar tres con el síndrome de Klinefelter es 0,089.

Existe una importante aplicación de la Distribución Poisson, cuando se considera una variable aleatoria binomial, donde *n* es grande y la Probabilidad de Éxito (*p*) es muy pequeña, de tal modo que la Probabilidad de Fracaso (q) se acerca a 1,0. En estos casos, se ha demostrado, que la Distribución Poisson con parámetro $\lambda s = np$ se aproxima a la Binomial; en la práctica, esta aproximación es buena si *n* es mayor o igual a 30 y *p* menor o igual a 0,05 y muy buena cuando *n* es mayor o igual a 100 y *np* es menor o igual a 10.

Ejemplo 4.14: La Probabilidad de que un perro muera como consecuencia de la aplicación de una vacuna pentavalente contra moquillo, hepatitis infecciosa, parvovirosis, parainfluenza canina y leptospirosis, es de 0,00002. Si se administra la vacuna a 100.000 caninos, asumiendo éstos como un conjunto independiente de ensayos, ¿Cuál es la Probabilidad en cada uno de los siguientes casos?:

a. Que mueran como máximo 2 perros.
b. Que mueran 3 o más animales.
c. Que mueran entre 5 y 8 inclusive.
d. Que mueran exactamente 3 perros.

Esencialmente, se trata de un experimento binomial, con n = 100.000 y p = 0,00002. Puesto que p está muy próxima a cero y n es muy grande, para calcular las Probabilidades pedidas, se aproximará a una Distribución de Poisson con $\lambda s = np$ = 100.000 × 0,00002 = 2.

Al igual que para la distribución binomial, para la Distribución de Poisson, también existen tablas. En el Tabla 3 del Apéndice, se dan los resultados correspondientes de la función de distribución $F(x, \lambda s)$, para diferentes valores de x y λs.

Si X es la variable aleatoria: muere un determinado número de perros, entonces:

a. $P[X \leq 2] = F(2) = 0{,}6767$.

   En la tabla de la Distribución de Poisson, se encuentra que para $\lambda s = 2$ y $x = 2$, por lo tanto $F(2, 2) = 0{,}6767$, como se puede observar en el siguiente aparte del apéndice citado.

	λs									
x	**1,1**	**1,2**	**1,3**	**1,4**	**1,5**	**1,6**	**1,7**	**1,8**	**1,9**	**2,0**
0	0,3329	0,3012	0,2725	0,2466	0,2231	0,2019	0,1827	0,1653	0,1496	0,1353
1	0,6990	0,6626	0,6268	0,5918	0,5578	0,5249	0,4932	0,4628	0,4338	0,4060
2	0,9004	0,8795	0,8571	0,8335	0,8088	0,7834	0,7572	0,7306	0,7037	**0,6767**
3	0,9743	0,9662	0,9569	0,9463	0,9344	0,9212	0,9068	0,8913	0,8747	0,8571

   La Probabilidad de que mueran como máximo dos perros, como consecuencia de la aplicación de la vacuna polivalente, es 0,6767.

b. $P[X \geq 3] = 1 - F(2) = 1 - 0{,}6767 = 0{,}3233$.

La Probabilidad de que mueran tres o más animales es 0,3233.

c. $P[5 \leq X \leq 8] = F(8) - F(4) = 0{,}9998 - 0{,}9473 = 0{,}0525$.

La Probabilidad de que mueran entre cinco y ocho perros es 0,0525.

d. $P[X = 3] = F(3) - F(2) = 0{,}8571 - 0{,}6767 = 0{,}1804$.

La Probabilidad de que mueran tres animales es 0,1804.

La Distribución Multinomial

Es una generalización de la distribución binomial, donde, si los sucesos a_1, a_2, ..., a_k, pueden ocurrir con frecuencias p_1, p_2, ..., p_k, respectivamente, entonces, la Probabilidad de que a_1, a_2, ..., a_k ocurran x_1, x_2, ..., x_k veces, respectivamente, se calcula de la siguiente forma:

$$f(X_k) = \frac{N!}{X_1!X_2!...X_k!} {p_1}^{x_1} \times {p_2}^{x_2} \times ... \times {p_k}^{x_k}$$

$$P[x_1, x_2, ..., x_k] = \frac{n!}{x_1!\ x_2!...x_k!} {p_1}^{x_1} \times {p_2}^{x_2} \times ... \times {p_k}^{x_k} \qquad (4.8)$$

Donde:

$x_1 + x_2 + ... + x_k = n$
$p_1 + p_2 + ... + p_k = 1$

Ejemplo 4.15: Si se lanza un dado 15 veces, ¿Cuál es la Probabilidad de obtener dos veces cada cara del dado?

De acuerdo con (4.8), el procedimiento para el cálculo es:

$$P[x_1, x_2, x_3, x_4, x_5, x_6] = \frac{15!}{2!2!2!2!2!2!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2$$

$$= \frac{1307674368000}{64}(9{,}8464)10^{-15} = 0{,}00020$$

El resultado indica que, la Probabilidad de que el dado caiga dos veces cada cara, es 0,00020.

Ejemplo 4.16: En el cruzamiento entre animales heterocigóticos para el color del manto en la raza Holstein, en 12 nacimientos, ¿Cuál es la Probabilidad de que tres crías sean negras homocigóticas, cinco negras heterocigóticas y cuatro rojas?

Con la aplicación del mismo procedimiento, se tiene:

$$P[x_1, x_2, x_3] = \frac{12!}{3!5!4!}\left(\frac{1}{4}\right)^3\left(\frac{2}{4}\right)^5\left(\frac{1}{4}\right)^4 = \frac{479001600}{17280}(1{,}9073)10^{-6} = 0{,}0528$$

La Probabilidad de que, de las 12 crías próximas a nacer, tres sean negras homocigotas, 5 negras heterocigotas y 4 rojas, es 0,0528.

En los experimentos multinomiales, se espera que el número de veces que ocurrirán $a_1, a_2, ..., a_k$ en n ensayos, se estima por $np_1, np_2, ..., np_k$. Así, en el Ejemplo 4.16 sobre las crías Holstein, se esperan 3 negras homocigóticas (12×0,25), 6 negras heterocigóticas (12×0,5) y 3 rojas (12×0,25).

VARIABLES ALEATORIAS CONTINUAS

Antes de iniciar este tema, es conveniente indicar que, por simple convención, las Variables Aleatorias Continuas se identificarán con la letra *Y*, en el presente texto.

Una variable aleatoria *Y* es continua, si los valores que toma, consisten en uno o más intervalos de la recta real, es decir, si su rango es infinito no numerable.

Para las Variables Aleatorias Discretas, se asignan probabilidades a todos los valores puntuales de la variable. Para el caso continuo, esto no es posible, puesto que el rango es no numerable, así que la Probabilidad de que *Y* tome un valor específico: $P[Y = y] = 0$.

Una Variable Aleatoria Continua, está caracterizada por una función *f(y)* que recibe el nombre de ***Función de Densidad de Probabilidad*** y ésta proporciona un medio para calcular la Probabilidad en un intervalo $a \leq Y \leq b$.

La función *f(y)* es una Función de Densidad de Probabilidad, si:

a) $f(y) \geq 0$ para $-\infty < y < \infty$.

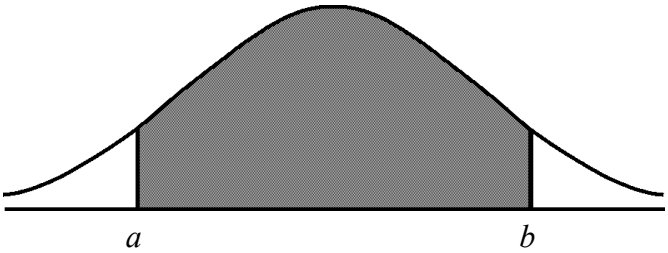
b) $\int_{-\infty}^{+\infty} f(y)\ dy = 1$.

c) $P[a \leq Y \leq b] = \int_{a}^{b} f(y)\ dy$

Debido a que, en una Variable Aleatoria Continua, la Probabilidad de que *Y* tome un valor específico es cero, se tiene que:

$$P[a \leq Y \leq b] = P[a \leq Y < b] = P[a < Y \leq b] = P[a < Y < b]$$

La primera propiedad dice que *f* es no negativa; la segunda afirma que *f* se constituye de manera que el área bajo su curva, limitada por el eje de las *Y*, es igual a uno, cuando se calcula en el rango de *Y* para el que está definido *f*(*y*); la tercera propiedad expresa cómo calcular la Probabilidad de que *Y* tome un valor desde $\boldsymbol{y = a}$ hasta $\boldsymbol{y = b}$ (Figura 4.2).

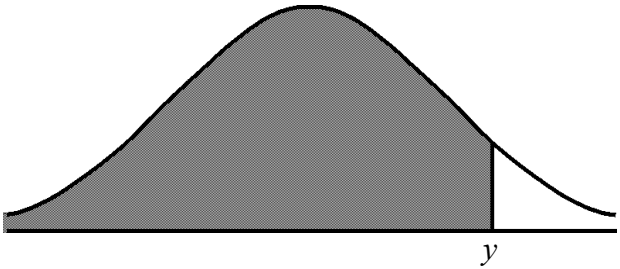


$$P(a < Y < b) = \int_a^b f(y)dy$$

Figura 4.2 Representación Gráfica de la Probabilidad de ($a < Y < b$)

La Función de Distribución Acumulada *F(y)* de una Variable Aleatoria Continua *Y*, con Función de Densidad de Probabilidad *f(y)* (Figura 4.3), está dada por:

$$F(y) = P[Y \leq y] = \int_{-\infty}^{x} f(t)\ dt, \text{ para } -\infty < y < \infty \qquad (4.9)$$



$$P[Y \leq y]$$

Figura 4.3 Representación gráfica de la Función de Distribución Acumulada.

De aquí se desprende que: $P[Y > y] = P[Y \geq y] = 1 - F(y)$

También se tiene que: $P[a < Y < b] = F_Y(b) - F_Y(a)$.

El valor esperado de una Variable Aleatoria Continua *Y* se define como:

$$E[Y] = \mu = \int_{-\infty}^{+\infty} Yf(y)\ dy \qquad (4.10)$$

La Varianza se define como:

$$V[Y] = \sigma^2 = \int_{-\infty}^{+\infty} (y - \mu)^2 f(y)\ dy \qquad (4.11)$$

La Distribución Normal

Con esta denominación se refiere a una "familia" de Variables Aleatorias Continuas, descrita por primera vez, por De Moivre en 1733, como el valor límite de la densidad binomial con un número de ensayos grande. Los resultados encontrados por de Moivre, pasaron prácticamente inadvertidos durante 50 años y fueron Gauss y Laplace quienes, al estudiar problemas de Astronomía, en forma independiente, presentaron la Distribución Normal como una descripción de los errores en las medidas astronómicas.

La Distribución Normal es de gran utilidad, ya que constituye una excelente aproximación, para una cantidad apreciable de distribuciones, que tienen mucha importancia práctica y posee varias propiedades matemáticas, que conducen a deducir resultados teóricos de interés.

La Distribución Normal ha tenido especial importancia en Producción y Salud Animal, para el correcto análisis e interpretación de datos, especialmente experimentales. El desarrollo de la Teoría de la Selección Animal, así como el estudio de una amplia gama de variables relacionadas con la fisiología de los animales domésticos, han tenido como base, a la Distribución Normal.

Una variable aleatoria *Y*, que toma todos los valores reales $-\infty < y < \infty$, tiene una Distribución Normal o Gaussiana, si su Función de Densidad de Probabilidad es de la siguiente forma:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{1}{2}\left[\frac{y-\mu}{\sigma}\right]^2\right)}, -\infty < y < \infty \qquad (4.12)$$

Donde σ^2 es la Varianza, μ es la Media, e la base de los logaritmos naturales, para lo cual se cumple la condición $-\infty < \mu < \infty$ y $\sigma > 0$

En adelante aparecerán referencias a la Distribución Normal, con mucha frecuencia, por lo que será necesario abreviar la anterior notación con una más breve, donde se indica que, *Y* tiene distribución $N(\mu, \sigma^2)$, si su ***Función Densidad*** está determinada por (4.12).

Como se puede apreciar en (4.12), pueden existir muchas Distribuciones Normales que se diferencian una de otra por su Media y su Varianza. La fórmula antes mencionada, no es fácil de desarrollar, ya que, para calcular Probabilidades será necesario establecer áreas bajo la curva, mediante el uso de integrales; sin embargo, el trabajo con Tablas de Probabilidad facilita enormemente esta labor, por lo tanto, no interesa la complejidad de la ecuación.

Propiedades de la Distribución Normal. La Distribución Normal tiene las siguientes propiedades:

a) La gráfica de densidad para cualquier variable aleatoria normal es una ***curva simétrica*** en forma de campana, conocida con los nombres de ***Curva de Gauss*** o ***Curva Laplaciana***, cuyo centro es μ, punto que indica donde está centrada la curva a lo largo del eje horizontal, tal como se muestra en la Figura 4.4.

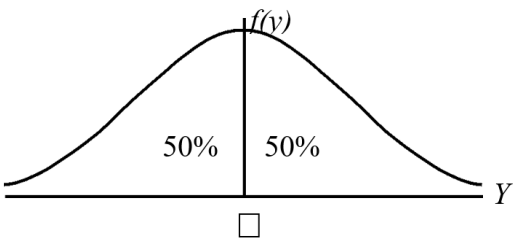


Figura 4.4 Curva Normal

b) La curva tiene unos puntos de inflexión, que indican los valores de Y iguales a una desviación estándar a cada lado de $\boldsymbol{\mu}$, entonces, cada punto de inflexión representa $Y = \mu \pm \sigma$. La situación de estos puntos determina la forma de la curva, de tal modo que, entre más grande es $\boldsymbol{\sigma}$, más lejos estarán localizados los puntos de inflexión y, en consecuencia, la curva será más plana, por lo tanto, es fácil establecer con claridad que, $\boldsymbol{\mu}$ es el parámetro de localización y $\boldsymbol{\sigma}$ es el parámetro de forma.

a) Todas las Variables Aleatorias Normales son continuas y su Probabilidad de ocurrencia, se calcula al encontrar las áreas correspondientes bajo la curva.

b) En una Distribución Normal, la Media, la Moda y la Mediana son iguales.

c) El Coeficiente de Asimetría, en una curva normal, es igual a cero y el de la curtosis, es igual a tres. Conviene aclarar que, la medida de curtosis se hace, por lo general, con referencia a la normal; una curva es ***Leptocúrtica***, si tiene un pico puntiagudo, ***Platicúrtica*** si es "aplastada" y ***Mesocúrtica*** si es intermedia, tal como sucede en las curvas normales.

d) La curva es simétrica, lo que indica que a cada lado de la media se localiza el 50% de los datos, lo que quiere decir que, el área total bajo la curva es uno, tal como se aprecia en la Figura 4.4.

g) En el rango $\mu \pm \sigma$ se localiza aproximadamente el 68% de los datos, entre $\mu \pm 2\sigma$ el 95% de los datos y entre $\mu \pm 3\sigma$ el 99% de los datos, de la manera indicada en la Figura 4.5.

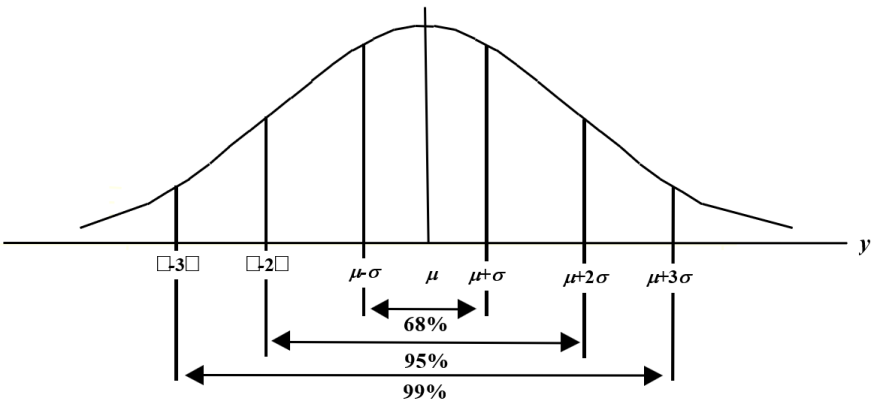


Figura 4.5 Áreas bajo la curva en una Distribución Normal.

Función de Distribución Acumulada de una Distribución Normal. Si Y es una variable aleatoria normal, con Media μ y Varianza σ^2, es decir $Y \sim N(\mu, \sigma^2)$, entonces, la Función de Distribución Acumulada $F(y)$ es:

$$F_Y(y) = P[Y \leq y] \qquad (4.13)$$

Esto corresponde al área bajo la Función de Densidad a la izquierda de y, como se ilustra en la Figura 4.6.

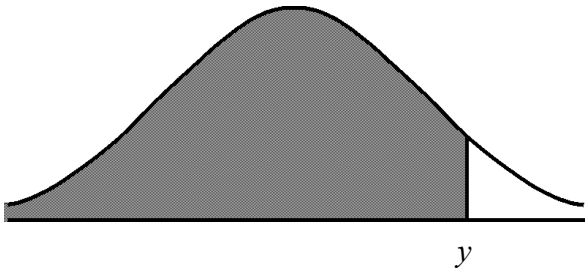


Figura 4.6 Representación gráfica de la función de distribución acumulada normal.

Cualquier Probabilidad, puede obtenerse a partir de la Función de Distribución Acumulada; por ejemplo, la Distribución de Probabilidad Acumulada en un intervalo (*a, b*), está dada por

$P[a < Y < b] = F_Y(b) - F_Y(a)$, como se indica en la Figura 4.7.

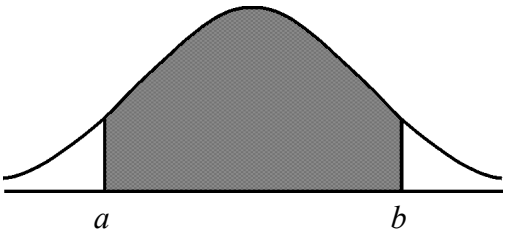


Figura 4.7 Representación gráfica de la Función de Distribución $P[a < Y < b]$.

Teniendo en cuenta que la Distribución Normal corresponde a una variable continua, la Probabilidad en un punto es igual a cero, por lo tanto:

$$P[a < Y < b] = P[a \leq Y \leq b] = P[a < Y \leq b] = P[a \leq Y \leq b]$$

También se tiene que: $P[Y \geq y] = P[Y > y] = 1 - F_Y(y)$ como se muestra en la Figura 4.8.

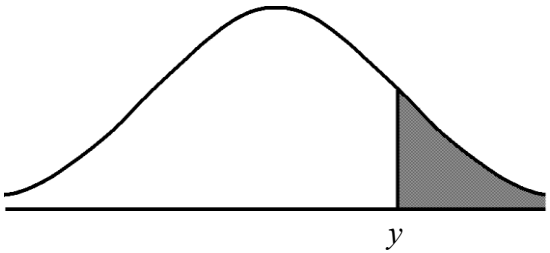


Figura 4.8 Representación gráfica de la Función de Distribución Acumulada $P[Y \geq y]$.

Distribución Normal Estándar. Como se mencionó anteriormente, para calcular las probabilidades de funciones de densidad normales, se tendría que resolver, mediante el cálculo de integrales, para diferentes valores de μ y σ^2. Afortunadamente, las probabilidades de cualquier Distribución Normal, pueden expresarse en términos de las Probabilidades de una normal determinada, denominada ***Normal Estandarizada o Tipificada Z*** que, tiene como Media cero y Varianza uno, para la cual, ya se encuentran calculados y tabulados los valores de probabilidad.

Para ello, es necesario realizar la transformación de los valores de la variable *Y* a valores de *Z*, por medio de la siguiente fórmula:

$$Z = \frac{y-\mu}{\sigma} \qquad (4.14)$$

Es decir, si *Y* se distribuye $N(\mu\, y\ \sigma^2)$, entonces $Z = \frac{y-\mu}{\sigma}$ se distribuye $N(0, 1)$.

En la Tabla 4 del Apéndice, se encuentra tabulada la función de distribución acumulada de una variable aleatoria normal estándar. En esta tabla se dan valores de la función $F_z(z) = P[Z \leq z]$

Con el fin de ilustrar el uso de esta tabla, para valores de *Z* desde –3,5 hasta 3,5, se halla la Probabilidad de que *Z* sea menor o igual que 1,96. En la columna de la izquierda de la Tabla 4 del Apéndice, se localiza el número 1,9 y luego en esta fila se desplaza horizontalmente hasta encontrar la columna bajo 0,06, donde se encuentra el valor 0,975, como se indica en el aparte correspondiente, del Apéndice antes citado.

z	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	**0,97500**	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520

Es decir, $P[Z \leq 1{,}96] = 0{,}975$ que geométricamente se puede expresar, como el área bajo la curva a la izquierda de 1,96 (Figura 4.9).

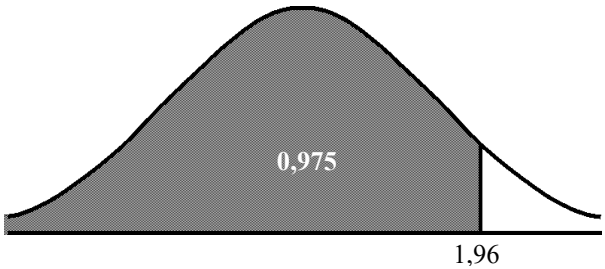
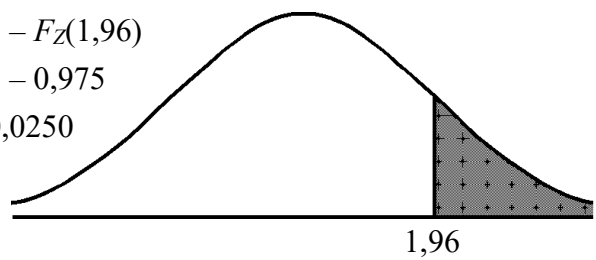


Figura 4.9 Representación gráfica de la Función de Distribución Acumulada *P*[Z ≥ 1,96].

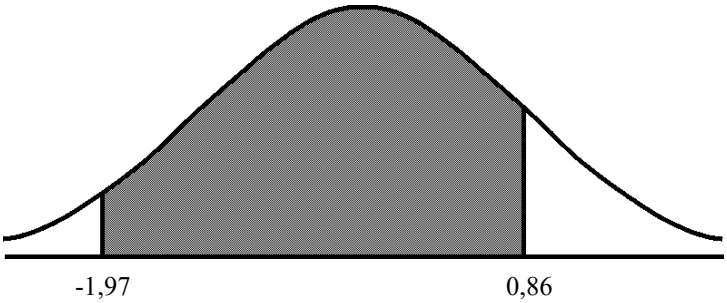
Ejemplo 4.17: Dada una Distribución Normal Estándar, calcular $P[Z > 1{,}96]$.

El valor de la Probabilidad pedida se calcula de la siguiente manera:

$$\begin{aligned} P[Z > 1{,}96] &= 1 - P[Z \leq 1{,}96] \\ &= 1 - F_Z(1{,}96) \\ &= 1 - 0{,}975 \\ &= 0{,}0250 \end{aligned}$$



Ejemplo 4.18: Dada una Distribución Normal Estándar, calcular $P[-1{,}97 \leq Z \leq 0{,}86]$. La Probabilidad pedida corresponde al área de la parte sombreada en la siguiente representación:



Esta probabilidad se calcula así:
$P[-1{,}97 \leq Z \leq 0{,}86] = F_Z(0{,}86) - F_Z(-1{,}97)$.

Al utilizar la Tabla 4 del Apéndice se obtiene $F_Z(0{,}86) = 0{,}8051$ y $F_Z(-1{,}97) = 0{,}0244$. Por lo tanto: $P[-1{,}97 \leq Z \leq 0{,}86] = 0{,}8051 - 0{,}0244 = 0{,}7807$.

Del Ejemplo 4.18 se deduce que, en ocasiones, es necesario calcular Probabilidades de una Variable Aleatoria Normal Y, con Media μ y Varianza σ^2, en un intervalo, por ejemplo $[a \leq y \leq b]$. En estos casos:

$$P[a \leq Y \leq b] = P\left[\frac{a-m}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right] = F_Z\left(\frac{b-\mu}{\sigma}\right) - F_Z\left(\frac{a-m}{\sigma}\right)$$

Donde Z es la Variable Aleatoria Normal Estándar y $F_Z(z)$ representa la Función de Distribución Acumulada.

Ejemplo 4.19: Si Y representa el peso al sacrificio de los cerdos Pietrain y se distribuye según una normal, con $\mu = 100$ y $\sigma = 10$, calcular las siguientes probabilidades:

a. Encontrar cerdos con pesos entre 95 y 120 kg = $P[95 \leq Y \leq 120]$.
b. Encontrar cerdos con pesos menores a 85 kg = $P[Y < 85]$.
c. Encontrar cerdos con pesos mayores o iguales a 108 kg = $P[Y \geq 108]$.

Las probabilidades pedidas se obtienen de la siguiente forma:

a) $P[95 \leq Y \leq 120] = P\left[\frac{95-\mu}{\sigma} \leq Z \leq \frac{120-\mu}{\sigma}\right]$

$= P\left[\frac{95-100}{10} \leq Z \leq \frac{120-100}{10}\right]$

$= P[-0,5 \leq Z \leq 2]$

$= F_Z(2) - F_Z(-0,5)$

$= 0,9772 - 0,3085 = 0,6687$

Como respuesta a la pregunta, se puede decir que, existe 0,6687 de Probabilidad de encontrar cerdos Pietrain entre 95 y 120 kg.

b) $P[Y < 85] = P\left[Z < \frac{85-\mu}{\sigma}\right]$

$= P\left[Z < Z < \frac{85-100}{10}\right]$

$= P[Z < -1,5] = F_Z(-1,5) = 0,0668$

En este caso, la Probabilidad de encontrar cerdos Pietrain con peso menor de 85 kg es 0,0668.

c) $P[Y \geq 108] = P\left[Z > \frac{108-\mu}{\sigma}\right] = P\left[> \frac{108-100}{10}\right]$

$= P[Z > 0{,}8] = 1 - P[Z \leq 0{,}8]$

$= 1 - F_Z(0{,}8) = 1 - 0{,}7881 = 0{,}2119$

La Probabilidad de encontrar cerdos de esta raza, con peso superior a 108 kg es 0,2119.

Ejemplo 4.20: Si la producción de vacas Holstein se aproxima a una Distribución Normal, con Media de 5.000 L y Varianza de 16.000 L^2. Hallar la Probabilidad de encontrar vacas con producción mayor o igual que 5.180 L.

La Probabilidad buscada se calcula como sigue:

$P[Y \geq 5.180] = P\left[Y \geq \frac{5.180-\mu}{\sigma}\right]$

Dado μ = 5.000 L y σ^2 = 16.000 L^2, entonces σ = 126,49 L, se tiene que:

$P[Y \geq 5.180] = P\left[Z \geq \frac{5.180-5.000}{126{,}49}\right] = P[z \geq 1{,}42]$

$= 1\text{- P}[Z \geq 1{,}42]$

$= 1 - F_Z(1{,}42) = 1 - 0{,}9222 = 0{,}0778$

Se concluye que, la Probabilidad de encontrar vacas con producciones por lactancia mayores o iguales a 5.180 L es 0,0778,

es decir, se esperaría que el 7,78% de los animales, tengan producciones iguales o superiores al valor indicado.

Ejemplo 4.21: En la misma población descrita en el Ejemplo 4.20, se pregunta: ¿Cuál es la Probabilidad de encontrar vacas con producciones superiores a 5.300 L y menores de 5.400 L?

La Probabilidad pedida se calcula así:

$$P[5.300 < Y < 5.400] = P\left[\frac{5.300-5.000}{126,49} < Z < \frac{5.400-5000}{126,49}\right]$$

$$= P[2,37 < Z < 3,16] = F_Z(3,16) - F_Z(2,37)$$

$$= 0,9992 - 0,9911 = 0,0081$$

El resultado se interpreta como la Probabilidad de encontrar vacas que produzcan entre 5.300 y 5.400 L es 0,0081. Esto significa que, el 0,81% de las vacas de la población analizada, tiene producciones en el rango indicado.

Si se conoce *N*, es posible estimar el mínimo de animales con la producción indicada. Por ejemplo: si en la población hay 600 vacas, se espera que $0,0081 \times 600 = 4,98 \approx 5$ sea el número de vacas que estén produciendo la cantidad de leche en el intervalo indicado.

Ejemplo 4.22 En una estación acuícola dedicada al cultivo de trucha, se calculó la Media μ = 17,8 cm de longitud estándar y σ^2 = 64 cm^2, como medida a la cosecha. En el mercado nacional, no se aceptan truchas con longitud menor de 15,3 cm. ¿Cuál es la Probabilidad de encontrar animales con esa condición?

Asumiendo que la variable se distribuye aproximadamente como una normal, la Probabilidad se calcula de la siguiente manera:

$$P[Y < 15{,}3] = P\left[Z < \frac{15{,}3-17{,}8}{8}\right] = P[Z < -0{,}3125]$$

$$= F_Z(-0{,}3125) = 0{,}3783$$

Ejemplo 4.23: Si en la empresa, descrita en el Ejemplo 4.22, hay 4.000 animales, se espera que de éstos 4.000 × 0,3783 = 1.513,2 ≈ 1.513 individuos, no cumplan con las condiciones de mercadeo.

Ejemplo 4.24: Suponiendo que las truchas con talla superior a 21 cm se pagan mejor en el mercado, ¿Cuál es la Probabilidad de encontrar animales con esa condición?

$$P[Y > 21] = P\left[Z < \frac{21-17{,}8}{8}\right] = P[Z > 0{,}4]$$

$$= 1 - F_Z(0{,}4) = 1 - 0{,}6554 = 0{,}3446$$

Esto quiere decir que, aproximadamente el 34,5% de los animales tienen la talla que garantiza mejor precio y en número serían 4.000 × 0,3446 ≈ 1.378 animales, aproximadamente.

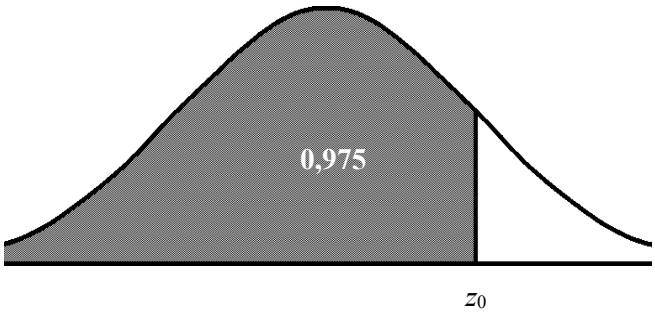
En el cálculo de las Probabilidades de los ejemplos anteriores, se conoce el valor de un *Z*o determinado; sin embargo, muchas veces es necesario hacer lo contrario, es decir, dada la Probabilidad hallar *Z*o. A continuación se ilustra estos casos.

Ejemplo 4.25: Dada una Distribución Normal Estándar, encuentre el valor de z_0, tal que:

a. $P[Z \leq z_0] = 0{,}9750$
b. $P[Z > z_0] = 0{,}6500$

a. La solución $P[Z \leq z_0] = 0{,}9750$ se ilustra en la siguiente representación gráfica:

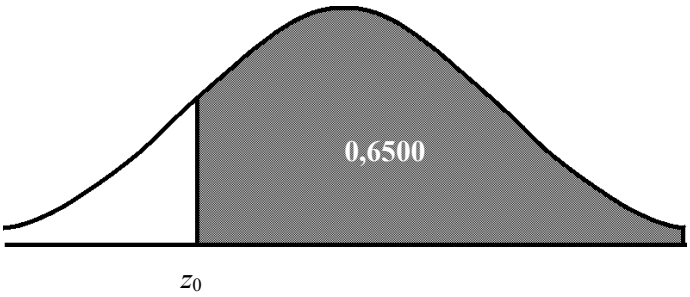
b.



Aquí se observa que, el área a la izquierda de z_0 es 0,975. En el interior de la Tabla 4 del Apéndice, se busca el número 0,975 o el que más se aproxime. Una vez hallado este número, se observa que a la izquierda, en la primera columna se encuentra 1,9 y hacia arriba en la primera fila, se encuentra el número 0,06, es decir, $z_0 = 1{,}96$, como se muestra en la siguiente sección del Apéndice citado, con los valores de la distribución acumulada normal estándar.

z	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169

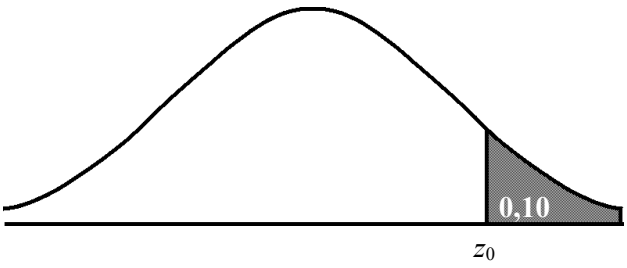
c. Para este punto del ejemplo: $P[Z > z_0] = 0{,}6500$, se muestra en la siguiente representación.



El área a la derecha de z_0 es 0,6500; por lo tanto, el área a la izquierda de este valor de z_0 es 0,3500 como resultado de (1 – 0,3500); en la Tabla 4 del Apéndice se busca el número 0,3500, cuyo número próximo es 0,3483 y en consecuencia $z_0 = -0{,}39$.

Ejemplo 4.26: Con los datos del Ejemplo 4.22, hallar la longitud mínima del 10% de las truchas de mayor tamaño.

Para la solución al problema, se parte del siguiente esquema.



Así, sea z_0 la longitud mínima necesaria para estar en el 10% de los animales de mayor tamaño.

$P[Y > z_0] = 0{,}10$, entonces,

$P[y \leq z_0] = 0{,}90$

$P\left[Z \text{ £ } \frac{z_0-\mu}{\sigma}\right] = 0{,}90$

Dado que $\mu = 17{,}8$ y $\sigma = 8$, $P\left[Z \text{ £ } \frac{z_0-17{,}8}{8}\right] = 0{,}90$

En la Tabla 4 del Apéndice se encuentra el valor de z que le corresponde a la Probabilidad 0,90, es decir:

$z_{0{,}90} = 1{,}28$

Por lo tanto: $\frac{z_0-17{,}8}{8} = 1{,}28$

$z_0 = 1{,}28$ y $8 + 17{,}8 = 28{,}04$

Es decir, la longitud mínima es de 28,04 cm para estar en el 10% superior

Teorema Central del Límite

Muchas de las variables aleatorias, que se encuentran en la práctica, son promedios de variables aleatorias independientes. La Media $\bar{Y}$ es un estadístico muy importante en problemas de toma de decisiones para medias poblacionales desconocidas.

El Teorema Central del Límite indica cómo se distribuye la Media Muestral $\bar{Y}$. Sean Y_1, Y_2, ..., Y_n variables aleatorias independientes, con idéntica distribución de Media μ y Varianza σ^2 finita, el Promedio Muestral $\bar{Y} = \frac{Y_1+Y_2+\ldots+Y_n}{n}$ tiene una distribución con Media μ y Varianza $\frac{\sigma^2}{n}$ que tiende hacia

una Distribución Normal, conforme ***n*** se hace grande, es decir $\frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}}$ tiende a ser normal estándar cuando ***n*** es grande.

La importancia de este teorema radica en que, para cualquier distribución común de un conjunto de variables aleatorias, el promedio de un número grande de ellas es aproximadamente normal, independientemente de la distribución común de las *N* aleatorias. Este teorema es válido no únicamente para variables aleatorias continuas, sino que se extiende a variables aleatorias discretas.

Aproximación Normal de la Distribución Binomial

Teniendo en cuenta que la Distribución Binomial resulta de una serie de ***n*** experimentos independientes, entonces, aplicando el Teorema Central del Límite, para un número grande de ***n***, se puede aproximar la Distribución Binomial a la normal.

Sea *X* una variable binomial con Media *np* y Desviación Estándar $\sqrt{np(1-p)}$, entonces, la variable aleatoria $\frac{x-np}{\sqrt{np(1-p)}}$ tiende a la normal estándar conforme el número de ensayos independientes ***n*** es lo suficientemente grande.

La aproximación es adecuada para $np > 5$, cuando $p \leq 0{,}5$ o para $n(1-p) > 5$ cuando $p > 0{,}5$.

Si se quiere calcular la Probabilidad en un intervalo, se tiene que:

$$P(a \leq X_B \leq b) \approx P\left(\frac{a-np}{\sqrt{np(1-p)}} \leq Z \leq \frac{b-np}{\sqrt{np(1-p)}}\right) \qquad (4.15)$$

Donde *Z* es normal estándar.

Teniendo en cuenta que la Variable Aleatoria Binomial X_B es discreta y la aproximación a la normal que es continua, se puede obtener una mejor aproximación, utilizando un ajuste que se denomina Corrección por Continuidad.

Utilizando esta corrección (4.15), se tiene que:

$$P(a \leq X_B \leq b) \approx P\left(\frac{a - np - 0{,}5}{\sqrt{np(1-p)}} \leq Z \leq \frac{b - np + 0{,}5}{\sqrt{np(1-p)}}\right)$$

Si se quiere calcular la Probabilidad en un punto x, se tiene:

$$P(X_B = x) \approx P\left(\frac{x - np - 0{,}5}{\sqrt{np(1-p)}} \leq Z \leq \frac{x - np + 0{,}5}{\sqrt{np(1-p)}}\right)$$

Ejemplo 4.27: En una finca dedicada a la producción de pie de cría de ganado Brahaman puro se confirmó, por palpación rectal, que 56 vacas estaban preñadas. De acuerdo con esta cifra, ¿Cuál es la Probabilidad de que nazcan:

a. Entre 30 y 35 machos?
b. 19 machos?

Se tiene que: $n = 56$
p = Probabilidad de que sea macho.
$p = 0{,}5$

A partir de estos datos, las soluciones serían las siguientes:

a) $P(30 \leq x_B \leq 35) \approx P\left(\frac{30-np-0{,}5}{\sqrt{np(1-p)}} \leq Z \leq \frac{35-np+0{,}5}{\sqrt{np(1-p)}}\right)$

$$= P\left(\frac{30-28-0{,}5}{\sqrt{14}} \leq Z \leq \frac{35-28-0{,}5}{\sqrt{14}}\right)$$

$= P[0{,}40 \leq Z \leq 2{,}00]$

$= F_Z(2{,}0) - F_Z(0{,}40)$

$= 0{,}9772 - 0{,}6554 = 0{,}3218$

La Probabilidad de que nazcan 30 y 35 macho es 0,3218.

b) $P[X_B = 19] \approx P\left[\frac{19-np-0{,}5}{\sqrt{np(1-p)}} \leq Z \leq \frac{19-np+0{,}5}{\sqrt{np(1-p)}}\right]$

$= P\left[\frac{19-28-0{,}5}{\sqrt{14}} \leq Z \leq \frac{19-28-0{,}5}{\sqrt{14}}\right]$

$= P[-2{,}54 \leq Z \leq -2{,}27] = F_Z(-2{,}27) - F_Z(-2{,}54)$

$= 0{,}0116 - 0{,}0055 = 0{,}0061$

La Probabilidad de que nazcan 19 macho es 0,006.

CAPÍTULO 5

Distribuciones muestrales y estimación

Una vez analizados los conceptos fundamentales de Probabilidad y de Distribución de Probabilidad, descritos en los capítulos anteriores, es posible abordar la Estadística Inferencial, que se encarga de generalizar las conclusiones que se obtienen en la muestra, a la población de donde se obtuvo ésta.

Para hacer afirmaciones que tengan validez sobre toda la población, con base en el comportamiento observado en la muestra, es necesario que sea adecuada, es decir, que conserve las características esenciales de la población, esto es, que sea representativa; además, para garantizar inferencias válidas acerca de la población, es importante que la muestra reúna ciertos requisitos, especialmente la de ser aleatoria.

MUESTRA ALEATORIA

Un conjunto de observaciones $y_1, y_2, ..., y_n$ extraídas de una población finita de tamaño $\boldsymbol{N}$, es una Muestra Aleatoria de tamaño $\boldsymbol{n}$ si se elige de tal manera que, cada subconjunto de $\boldsymbol{n}$ elementos de la población, tenga la misma Probabilidad de ser seleccionado.

Una manera de seleccionar una Muestra Aleatoria de una población finita es enumerar todos los elementos de la población con fichas y extraer $\boldsymbol{n}$ miembros de manera aleatoria; en este sentido, aleatorio significa total imparcialidad en

la selección de la muestra, para lo cual, se puede asignar un número aleatorio a cada elemento, utilizando una hoja de cálculo, para seleccionar la muestra requerida, ordenando los elementos, de menor a mayor, según los números aleatorios asignados.

Otra técnica consiste en emplear una tabla de números aleatorios, como la indicada en la Tabla 5 del Apéndice (Steel y Torrie, 1.985). Esta tabla está formada por los dígitos 0, 1, 2, 3, 4, 5, 6, 7, 8 y 9, distribuidos en 100 filas por 100 columnas, dando origen a 10.000 números aleatorios, de los cuales, 5.094 son pares y 4.906 impares. Si un número apareciese con más frecuencia que otro, sólo se debe al azar. En la citada Tabla existen 1.015 ceros, 1.026 unos, 1.013 doses, 975 treses, 976 cuatros, 932 cincos, 1.067 seises, 1.013 sietes, 1.023 ochos y 960 nueves.

Es bueno aclarar que, de acuerdo con los métodos descritos, es necesario que, los individuos de la población, estén plenamente identificados y numerados.

Ejemplo 5.1: Se define una población conformada por 40 cerdos Large White, de la que se necesita tomar una muestra al azar con 10 de ellos, a fin de analizar la variable peso a los 90 días. Para esto es necesario disponer de los registros, donde esté consignado el peso individual a esa edad y la correspondiente identificación de cada animal, tal como se aprecia en la Tabla 5.1.

Tabla 5.1

Registro del peso individual a los 90 días, de una población de cerdos Large White.

No. del animal	Peso (kg)	No. del animal	Peso (kg)	No. del animal	Peso (kg)
00	72	14	68	27	66
01	89	15	78	28	68
02	69	16	81	29	69
03	78	17	88	30	74
04	85	18	74	31	69
05	80	19	103	32	77
06	73	20	87	33	59
07	61	21	68	34	85
08	68	22	59	35	61
09	87	23	61	36	63
10	89	24	68	37	68
11	72	25	85	38	68
12	67	26	67	39	64
13	92				

Suponiendo que, para los fines de un estudio particular, una muestra aleatoria de 10 datos es suficiente para tener una idea clara y confiable del comportamiento productivo de toda la piara, habría que responder la pregunta: ¿Cuáles son los 10 cerdos que se deben elegir para desarrollar el estudio?

Para garantizar la elección al azar, se pueden seguir estos pasos:

1. Disponer del listado de animales, identificados con un código, en este caso con dos dígitos del 00 al 39. Cada animal tendrá registrado su peso.

2. En la Tabla 5 del Apéndice, se marca arbitrariamente la fila. En el ejemplo se marcó el 15 y con el mismo procedimiento se marcó la columna 20. Luego se empiezan a seleccionar los dígitos de dos en dos, a partir del punto de corte de la fila 15 con la columna 20 y se leen los números de izquierda a derecha, hasta completar diez parejas. Así se encontraron las parejas 6 3, 8 5, 2 4, 4 8, 4 0, 0 2, 5 9, 2 8, 8 5 y 7 2. Según las cifras encontradas deberían seleccionarse, como parte de la muestra, los cerdos 63, 85, 24, 48, 40, 02, 59, 28, 85, y 72.

   Como el tamaño de la población es $N = 40$, las cifras 63, 85, 48, 40, 59, 85 y 72 no existen en el listado, por lo tanto, se descartan. En consecuencia, hasta el momento, de los 10 cerdos que se necesitan, únicamente se han logrado seleccionar tres que son el 24, 02 y 28; aún faltan 7. Para seleccionar las 7 restantes, simplemente se continúa formando parejas de dígitos desde el punto donde se encontró la última cifra, es decir en 72, por lo tanto, los siguientes números son: 03, 10, 79, 01, y 69. Los dígitos 03, 10 y 01 sirven para la muestra y se seleccionan, así que ya se dispone de 6 animales seleccionadas. Los dígitos 79 y 69 se descartan. Todavía hacen falta cuatro animales, pero ya se terminó la fila de números, de manera que, se debe bajar una fila y se procede a formar nuevas parejas, pero de derecha a izquierda, descartando las cifras que no sirvan, hasta completar la muestra.

En el ejemplo, se tuvo esta situación: 20, 07, 56, 09 y 00, de estos dígitos, el 56 se descarta, el resto se aceptan y así se completa el tamaño requerido de la muestra. En conclusión, se han seleccionado al azar los cerdos identificados con los códigos 24, 02, 28, 03, 10, 01, 20, 07, 09 y 00.

3. Como resultado de lo anterior, los números que se relacionan a continuación corresponden a 10 animales escogidos al azar, a partir de la Tabla 5.1.

No. del animal	Peso a los 90 días (kg.)
24	68
02	69
28	68
03	78
10	89
01	89
20	87
07	61
09	87
00	72

El procedimiento antes descrito equivale a introducir 40 balotas marcadas con los dígitos 00 al 39 en una bolsa y extraer al azar 10 de ellas, mezclando bien las balotas luego de cada extracción y volviendo a meter en la bolsa la balota que se ha sacado; es por esta razón que, este procedimiento se denomina ***Muestreo con Reemplazo***. Si no se introdujeran de nuevo en la bolsa las balotas ya sacadas, el ***muestreo sería sin reemplazo***.

Para obtener tamaños de muestra mayores, por ejemplo, tomar 100 observaciones de una muestra de 500, el procedimiento es el mismo, sólo que la numeración de los individuos a muestrear se hace con tres dígitos, iniciando con 000 y terminando con 499. Luego se escogen de la tabla, combinaciones de tres cifras, hasta completar el tamaño requerido.

Es importante reiterar que, la técnica antes detallada, no es el único procedimiento de elegir individuos al azar de una población; existen otros procedimientos alternativos de muestreo, donde se incluye, el uso de calculadora e incluso computadores personales, utilizando una hoja de cálculo para asignar números aleatorios.

Se ha indicado en repetidas ocasiones que, al hacer inferencias estadísticas, el interés radica en hacer afirmaciones de los parámetros de las poblaciones, como la Media μ y la Varianza σ^2, utilizando los estadísticos $\bar{y}$ y S^2, que se calculan a partir de las muestras aleatorias.

Si se tiene una población finita de tamaño $\boldsymbol{N}$, se podrá obtener varias muestras de tamaño $\boldsymbol{n}$. Por lo general, para cada muestra extraída se obtendrá valores diferentes de los estadísticos $\bar{y}$ y S^2, así que, se puede pensar en cada cantidad como una variable aleatoria con cierta Distribución de Probabilidad. Por ejemplo, si el estadístico es la Media, la distribución se conoce como ***Distribución Muestral de Medias***; si se trata de la Varianza, la distribución se denomina ***Distribución Muestral de Varianzas***. La Distribución de Probabilidad de los posibles resultados muestrales, proporciona una base para realizar inferencias.

DISTRIBUCIÓN DE MUESTREO DE LA MEDIA MUESTRAL $\bar{Y}$

Sea $y_1, y_2, y_3, ..., y_n$ una muestra aleatoria de una población, de tal manera que $E[Y_i] = \mu$ y $V[Y_i] = \sigma^2$, $i = 1, 2, 3, ..., n$, entonces el estadístico $\bar{Y} = \frac{y_1+y_2+...+y_n}{n}$ se denomina ***Media Muestral.***

El Valor Esperado o Media de $\bar{Y}$, simbolizado como $\mu_{\bar{Y}}$, y se estima de la siguiente manera:

$$\mu_{\bar{Y}} = E[\bar{Y}] = E\left[\frac{y_1 + y_2 + \ldots + y_n}{n}\right] = \frac{1}{n}(E[y_1] + E[y_2] + \ldots + E[y_n])$$

$$= \frac{1}{n}[\mu + \mu + \ldots + \mu] = \frac{1}{n} n\mu = \mu \qquad (5.1)$$

Es decir, la Media de la distribución de muestreo de la media muestral es la Media Poblacional μ.

La Varianza de la Media Muestral $\bar{Y}$, simbolizada por $\sigma^2_{\bar{Y}}$, se calcula como:

$$\sigma^2_{\bar{Y}} = V[\bar{Y}] = V\left[\frac{y_1 + y_2 + \ldots + y_n}{n^2}\right] = \frac{1}{n^2}(V[y_1] + V[y_2] + \ldots + V[y_n])$$

$$= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \qquad (5.2)$$

Es decir, la Varianza de la distribución muestral de la Media es $\frac{\sigma^2}{n}$

La Desviación Estándar de la distribución muestral de la Media es

$$d.e.(\bar{X}) = \frac{\sigma}{\sqrt{n}} \tag{5.3}$$

Esta expresión se conoce con el nombre de ***Error Estándar de la Media***. Se puede apreciar, que a medida que el tamaño de la muestra crece, el Error Estándar de la Media disminuye; es decir, si el tamaño de la muestra crece, la precisión de la Media Muestral para estimar la Media Poblacional aumenta. Si ***n*** es mayor que el 5% de ***N***, la Varianza $V[\bar{Y}]$ hay que multiplicarla por el factor de corrección para población finita $\frac{N-n}{N-1}$, con lo que:

$$V[\bar{Y}] = \frac{\sigma^2}{n} \times \frac{N-n}{N-1} \quad y \quad d.e.(\bar{Y}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \tag{5.4}$$

Se puede demostrar que, si la muestra aleatoria $y_1, y_2, ..., y_n$ proviene de una población normal con Media μ y Varianza σ^2, entonces el estadístico $\bar{y}$ se distribuye normal con Media μ y Varianza $\frac{\sigma^2}{n}$, es decir, la variable aleatoria

$$\frac{\bar{y}-\mu}{\frac{\sigma}{\sqrt{n}}} = Z, \text{ donde } Z \text{ es normal estándar.} \tag{5.5}$$

Si la población de donde se extrajo la muestra no es normal, el Teorema Central del Límite asegura que la aproximación (5.5) es buena para muestras grandes ($n \geq 30$).

La aproximación (5.5) es buena para muestras pequeñas si la distribución de donde se extrajo la muestra es normal o es bastante parecida a una normal, es

decir, es simétrica y tiene una alta concentración de los datos alrededor del punto de simetría.

Ejemplo 5.1: En una población comercial de cuyes (*Cavia porcellus*), el peso vivo individual promedio a la semana 12 de edad es 850 g y la Desviación Estándar 200 g. Hallar la Probabilidad que, en una muestra de 60 animales, el promedio sea menor que 900 g.

El problema pide calcular $P[\bar{y} < 900]$, conociendo que $n = 60$, $\mu = 850$ y $\sigma = 200$. Dado que ***n*** es grande, se tiene que:

$$\frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = Z, \quad \text{donde } \mu_{\bar{Y}} = \bar{y} \text{ y } \sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$$

Es decir: $\frac{\bar{y}-\mu}{\frac{\sigma}{\sqrt{n}}} = Z$, por lo tanto:

$$P[\bar{y} < 90] = P\left[\frac{\bar{y}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{900-\mu}{\frac{\sigma}{\sqrt{n}}}\right] = P\left[Z < \frac{900-850}{\frac{200}{\sqrt{60}}}\right]$$

$$= P[Z < 1{,}93] = F_Z(1{,}93) = 0{,}9732$$

Según esto, se puede concluir que, el promedio de la muestra de cuyes sea inferior a 900 g, es 0,9732.

DISTRIBUCIÓN DE MUESTREO DE LA PROPORCIÓN MUESTRAL

Sea ***X*** el número de éxitos en una muestra binomial de ***n*** observaciones, donde ***p*** es la probabilidad de éxitos. En la mayoría de aplicaciones, el parámetro ***p*** será la proporción de individuos de una población, que poseen cierta característica de interés, tal como la presencia de una determinada patología.

Un estimador de ***p*** es la proporción muestral $\hat{p} = \frac{x}{n}$, por lo tanto:

$$E[\hat{p}] = E\left[\frac{x}{n}\right] = \frac{E[X]}{n} = \frac{np}{n} = p \qquad (5.6)$$

Es decir, el Valor Esperado o Media de la proporción muestral, es la proporción ***p*** de éxitos en la población:

$$V[\hat{p}] = V\left[\frac{x}{n}\right] = \frac{V[X]}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \qquad (5.7)$$

Esto es: $$\sigma^2{}_{\hat{p}} = \frac{p(1-p)}{n} \qquad (5.8)$$

De la aproximación de la distribución binomial a la normal se tiene que. para ***n*** grande $\frac{x-np}{\sqrt{np(1-p)}}$ es aproximadamente normal estándar, siendo posible demostrar que: $\frac{\hat{p}-p}{\sigma_{\hat{p}}} = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$ es normal estándar.

Ejemplo 5.2: De una población de cíclidos *Herus severus*, se toma una muestra de 200 individuos. Suponiendo que la proporción de machos es *p* = 0,40; hallar la Probabilidad de que en la muestra la proporción de machos esté entre 0,35 y 0,45.

Se tiene que *p* = 0,40 y *n* = 200, entonces:

$$P[0{,}35 \leq \hat{p} \leq 0{,}45] = P\left[\frac{0{,}35-p}{\sigma_{\hat{p}}} \leq Z \leq \frac{0{,}45-p}{\sigma_{\hat{p}}}\right]$$

Ahora: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0{,}40 \times 0{,}60}{200}} = 0{,}035$

$$P[0{,}35 \leq p \leq 0{,}45] = P\left[\frac{0{,}35-0{,}40}{0{,}035} \leq Z \leq \frac{0{,}45-0{,}40}{0{,}035}\right]$$

$$= P[-1{,}43 \leq Z \leq 1{,}43]$$
$$= F_Z(1{,}43) - F_Z(-1{,}43)$$
$$= 0{,}9236 - 0{,}0764 = 0{,}8472$$

DISTRIBUCIÓN MUESTRAL DE LA VARIANZA S^2

Es necesario recordar que, la Varianza muestral $S^2 = \frac{\sum(y_i-\bar{y})^2}{n-1}$ es una medida de la variabilidad e indica la dispersión de los datos alrededor de la Media. La estadística S^2 es un estimador de la Varianza poblacional σ^2.

Para desarrollar la distribución de muestreo S^2 es conveniente hacer una introducción a una nueva distribución, denominada ***Chi-cuadrada***.

Distribución Chi-cuadrada

Si $y_1, y_2, ..., y_\upsilon$ son variables aleatorias independientes y cada una de ellas se distribuye como una normal estándar, entonces la suma $\chi_1^2 + \chi_2^2 + \ldots \chi_\upsilon^2$ se distribuye como Chi-cuadrada con $\boldsymbol{\upsilon}$ grados de libertad χ_υ^2.

Los grados de libertad $\boldsymbol{\upsilon}$ se constituyen en un parámetro asociado a la distribución y representa el número de variables cuyos cuadrados se suman. Esta distribución es asimétrica, con sesgo positivo, como se muestra en la Figura 5.1.

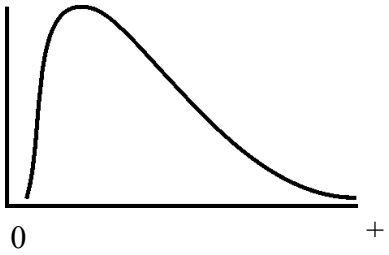


Figura 5.1 Representación gráfica de la distribución χ^2_{υ}

El Rango es el intervalo [0, +∞]

El Valor Esperado o Media $E[X_{\upsilon}^2] = \upsilon$

La Varianza $V[X_{\upsilon}^2] = 2\upsilon$.

Como se puede observar, la forma de la distribución es asimétrica con sesgo positivo y la Función de Distribución Acumulativa está dada por:

$$P[X \leq \chi^2_{1-\alpha,\,\upsilon}] = 1-\alpha \qquad (5.9)$$

Donde $0 \leq \alpha \leq 1$ y $\chi^2_{1-\alpha,\,\upsilon}$ es el valor del cuantil que le corresponde a la χ^2 para $1-\alpha$ y υ dados (Figura 5.2).

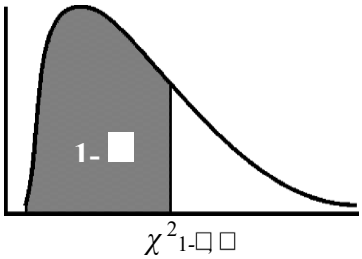


Figura 5.2 Distribución de Chi-cuadrada para $\chi^2_{1-\alpha,\,\upsilon}$

En la Tabla 6 del Apéndice se encuentran tabulados los valores de los cuantiles para diferentes cifras de 1-α y υ.

Ejemplo 5.3: Se requiere calcular el valor de *Y,* correspondiente a una variable distribuida como Chi-cuadrada, en una muestra aleatoria de 11 elementos, con una confiabilidad del 95%.

Simbólicamente, el anterior enunciado se resume de este modo:

$P[Y \leq \chi^2_{0,95;\,10}] = P[Y \leq 18,31] = 0,95.$

Para encontrar el valor pedido, se busca en la Tabla 6 del Apéndice, en el punto de cruce entre $\upsilon = 10$, que son los grados de libertad; es decir, n-1 y 1-α = 0,95, tal como se indica en la sección del Apéndice citado, que se muestra a continuación.

υ	$x_{0,005}$	$x_{0,010}$	$x_{0,025}$	$x_{0,050}$	$x_{0,100}$	$x_{0,900}$	$x_{0,950}$	$x_{0,975}$	$x_{0,990}$	$x_{0,995}$
1	0,00	0,00	0,00	0,00	0,02	2,71	3,84	5,02	6,63	7,88
2	0,01	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	10,60
3	0,07	0,11	0,22	0,35	0,58	6,25	7,81	9,35	11,34	12,84
4	0,21	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	14,86
5	0,41	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	16,75
6	0,68	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81	18,55
7	0,99	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	21,95
9	1,73	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	23,59
1	2,16	2,56	3,25	3,94	4,87	15,99	**18,31**	20,48	23,21	25,19

En este punto, el valor que aparece es 18,31, cuya representación gráfica se indica en la Figura 5.3.

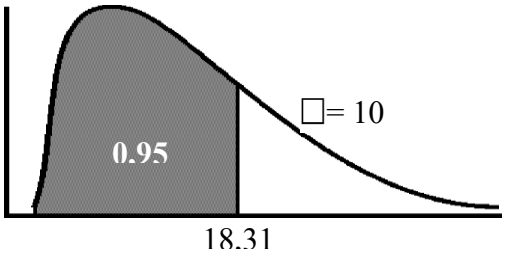


Figura 5.3 Valor de *Y* en una Distribución Chi-cuadrada para una muestra de 11 observaciones con $\alpha = 0{,}5$.

Las aplicaciones más importantes de la Distribución Chi-cuadrada, para resolver problemas específicos aplicados a la Producción y Salud Animal, aparecerán más adelante.

Distribución Muestral de la Varianza S^2

Si S^2 es la Varianza calculada en una muestra aleatoria de tamaño ***n***, extraída de una población con Varianza σ^2, entonces: $E[S^2] = \sigma^2$.

Si la población de donde se extrajo la muestra es normal, entonces se tiene que la estadística

$$(n-1)\frac{S^2}{\sigma^2} \approx \chi^2_{(n-1)} \qquad (5.10)$$

Donde $\chi^2_{(n-1)}$ es la Distribución Chi-cuadrada, con $n - 1$ grados de libertad.

Ejemplo 5.4: Un programa de mejoramiento genético de ovinos africanos pretende seleccionar los reproductores por la conversión alimenticia, de tal manera que, se seleccionan los individuos de mejor valor genético para la característica en cuestión. Se sabe que, el valor genético es una variable que sigue una distribución

normal, con Varianza 1,75. Se extrae una muestra aleatoria de 20 animales y se pide hallar la Probabilidad de que la Varianza Muestral sea superior a 3,10.

Se tiene que $n = 20$ y $\sigma^2 = 1{,}75$.

La Probabilidad que se pide es:

$$P[S^2 > 3{,}10] = P\left[(n-1)\frac{S^2}{\sigma^2} > (n-1)\frac{3{,}10}{1{,}75}\right]$$

Dado que $(n-1)\frac{S^2}{\sigma^2} \approx \chi^2{}_{19}$, entonces

$$P[S^2 > 3{,}10] = P\left[\chi^2_{19} > 19\,\frac{3{,}10}{1{,}75}\right]$$

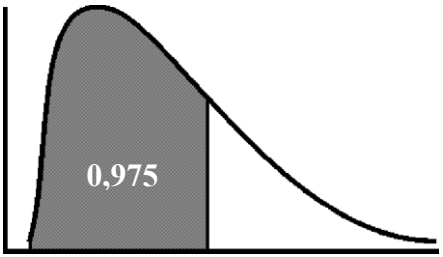
$$= P[\chi^2_{19} > 33{,}6] = 1 - P[\chi^2_{19} \leq 33{,}6]$$

En la Tabla 6 del Apéndice, para 19 grados de libertad, se encuentra que el valor más próximo a 33,66 es 32,87; en la parte superior de esta columna se halla el valor de la Probabilidad 0,975, entonces:

$P[\chi^2_{19} \leq 33{,}6] \approx 0{,}975$ y por lo tanto:

$P[\chi^2_{19} > 33{,}6] \approx 1 - 0{,}975 \approx 0{,}025$, es decir: $P[S^2 > 33{,}66] \approx 0{,}025$.

Este valor de probabilidad se representa de la siguiente manera:



Estimador

En párrafos anteriores de este texto se ha mencionado varias veces que, un aspecto de gran importancia en la Estadística Inferencial, radica en la ***estimación*** de parámetros a partir de sus correspondientes estadísticos, así que es oportuno definir, el término Estimador.

Dado un parámetro θ, un Estimador $\hat{\theta}$ de θ es un estadístico que estima a θ. Por ejemplo, el estadístico $\bar{Y}$ es un Estimador de μ, S^2 es un Estimador de σ^2 y la proporción muestral ***p*** es un Estimador de la proporción poblacional ***P***, que se pueden simbolizar de la siguiente manera:

$\bar{y} = \hat{\mu}$
$S^2 = \hat{\sigma}^2$
$p = \hat{P}$

A esta clase de estimadores se los denomina ***Estimadores Puntuales***.

Para estimar un parámetro es posible definir varios estadísticos; por ejemplo, a la Media Poblacional μ se la puede estimar con la Mediana Muestral. Existen algunas características deseables de los Estimadores Puntuales, que sirven para decidir en términos de éstas, qué Estimador es mejor.

Estimador Insesgado. Un Estimador $\hat{\theta}$ de θ, es insesgado si $E[\hat{\theta}] = \theta$, es decir, el valor medio del Estimador, es igual al parámetro que se busca estimar. Es importante que el Estimador sea insesgado, puesto que la distribución muestral del Estimador se encontrará alrededor del parámetro objetivo y, en consecuencia, los valores de estos Estimadores difieren del verdadero valor del parámetro con menor probabilidad que los otros Estimadores. Entre los Estimadores insesgados se tiene la Media Muestral $\bar{Y}$, la Proporción Muestral ***p***, la Varianza Muestral

$S^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$; es decir,

$E[\bar{Y}] = \mu$, $E[\hat{p}] = p$ y $E[S^2] = \sigma^2$

El Estimador $\hat{S}^2 = \frac{\sum(y_i - \bar{y})^2}{n}$ es sesgado, es decir, $E[\hat{S}^2] \neq \sigma^2$. Es por esto que, para estimar la Varianza σ^2, se utiliza S^2 y no $\hat{S}^2$.

Dado un parámetro θ, pueden existir dos Estimadores Insesgados que estiman al parámetro; en este caso, para decidir qué estimador es "mejor", se utiliza el criterio de eficiencia relativa.

Eficiencia Relativa. Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos Estimadores Insesgados del parámetro θ, la eficiencia relativa de $\hat{\theta}_1$ respecto de $\hat{\theta}_2$, se la define como $\frac{V[\hat{\theta}_2]}{V[\hat{\theta}_1]}$.

Si $V[\hat{\theta}_1] < V[\hat{\theta}_2]$, entonces $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$, en cuyo caso el valor de eficiencia relativa será mayor que 1. Es decir, entre dos Estimadores Insesgados de un parámetro, es más eficiente el que menos variabilidad tenga. Por ejemplo, $\frac{V[media]}{V[mediana]} = 0{,}6369$.

De aquí se deduce que $V[media] < V[mediana]$, o sea, la Media es más eficiente que la Mediana para estimar la media poblacional μ, siendo el valor de eficiencia relativa menor que 1.

Estimadores Insesgados de Varianza Mínima. Sea $\hat{\theta}$ un Estimador Insesgado del parámetro θ. Si $V[\hat{\theta}]$ es menor que la Varianza de cualquier otro Estimador Insesgado de θ, se dice entonces que $\hat{\theta}$ es un Estimador Insesgado de Varianza Mínima de θ. Entre los Estimadores Insesgado de Varianza Mínima más comunes se tienen:

a. La Media Muestral, cuando la muestra proviene de una Distribución Normal.
b. La Varianza Muestral, cuando la muestra proviene de una Distribución Normal.
c. La proporción muestral binomial.

Por lo tanto, se puede concluir que, la Media Muestral, la Varianza Muestral y la Proporción Muestral, son Estimadores aceptables de la Media Poblacional, la Varianza Poblacional y la Proporción Poblacional, respectivamente.

INTERVALOS DE CONFIANZA

En ciertos casos, los Estimadores Puntuales, aunque poseen ciertas propiedades deseables, no coinciden con el parámetro que buscan estimar; algunas veces es preferible estimar un parámetro mediante un ***Intervalo***, es decir, un rango de valores entre los que posiblemente se encuentre el verdadero parámetro que se quiere estimar.

Estimador por Intervalo

Dado un parámetro θ, un Estimador por Intervalo es un intervalo $\hat{\theta}_1 < \theta < \hat{\theta}_2$, donde $\hat{\theta}_1$ y $\hat{\theta}_2$ se denominan Límite Inferior y Límites Superior, respectivamente, y dependen del valor del estadístico $\hat{\theta}$ para una muestra en particular y de la distribución muestral de $\hat{\theta}$. A los Estimadores por Intervalo se los conoce como ***Intervalos de Confianza***.

A la Probabilidad de que un intervalo de confianza contenga al parámetro a estimar, se le denomina ***Coeficiente de Confianza***.

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha \qquad (5.11)$$

donde $0 \leq \alpha \leq 1$ y $1 - \alpha$ es el Coeficiente de Confianza.

Es importante tener en cuenta que, uno o los dos extremos del intervalo varían de manera aleatoria de una muestra a otra, es decir, la longitud y la localización del intervalo son aleatorias y no se podrá estar completamente seguros que, el parámetro objetivo esté contenido en el intervalo.

Si se estima un parámetro θ mediante un Intervalo de Confianza, con confiabilidad $(1-\alpha)100=95\%$, es decir, $P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 95\%$, en la práctica significa que, si se realiza un muestreo repetitivo en la población estudiada, se espera que el 95% de los intervalos que se calculen, deberán contener al parámetro θ y sólo por el azar, un 5% de ellos no contendrá el parámetro.

Intervalo de Confianza para la Media μ en Muestras con Distribución Normal y Varianza conocida

Suponiendo que se tiene una muestra aleatoria procedente de una población normal, con Media μ desconocida y Varianza σ^2 conocida, un Intervalo de Confianza para la Media Poblacional μ al (1 - α)100% de confiabilidad está dado por:

$$\bar{y} \pm \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} \qquad (5.12)$$

donde $Z_{1-\alpha/2}$ es el valor que se encuentra en la tabla de la normal estándar para un Coeficiente de Confianza 1-α, entonces $\bar{y} - \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}$ y $\bar{y} + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}$ son los límites de confianza $\hat{\theta}_1 \ y \ \hat{\theta}_2$, respectivamente (Figura 5.4).

Para calcular un intervalo de confianza para μ, con una confiabilidad del (1-α)100%, se ha supuesto que se conoce σ^2, pero por lo general, en la práctica σ^2 no se conoce; por lo tanto, para la estimación de los límites de confianza, si la muestra es grande ($n \geq 30$), se reemplazará σ^2 por la varianza de la muestra S^2, independientemente de la distribución de probabilidad que tenga la población de donde se extrajo la muestra, puesto que, por el teorema central del límite, para ***n*** grande la distribución de la media muestral $\bar{Y}$, se aproxima a la distribución normal con $\mu_{\bar{Y}} = \mu$ y $\sigma_{\bar{Y}}^2 = \frac{\sigma^2}{n}$. Cuando la muestra es pequeña, se explica más adelante.

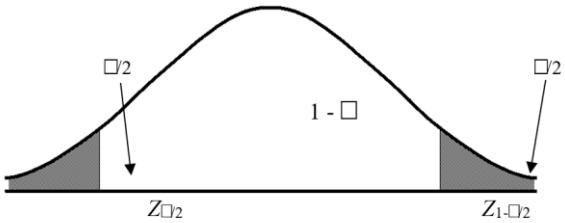


Figura 5.4 Intervalos de confianza para la media μ

Ejemplo 5.5: En la Granja de Botana de la Universidad de Nariño, se tomó una muestra aleatoria de 39 cuyes (*Cavia porcellus*), con el fin de estimar el peso promedio a las 12 semanas, mediante un Intervalo de Confianza, al 90% de confiabilidad. Los datos obtenidos fueron los siguientes: 666, 861, 702, 733, 720, 708, 985, 1030, 812, 798, 800, 968, 1002, 981, 700, 721, 1100, 1090, 1032, 847, 580, 920, 710, 830, 945, 975, 875, 810, 1005, 1050, 945, 925, 940, 936, 1050, 884, 1025, 975, 885.

De los anteriores datos se tiene que: $\bar{y} = 885{,}15$

$$S = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n-1}} = \sqrt{17625} = 132{,}76$$

$\hat{\sigma} = S = 132{,}76$

Al calcular $Z_{1-\alpha/2}$ para una confiabilidad del 90%, se tiene que:

$1 - \alpha = 0{,}90$

$\alpha = 0{,}10$

$\alpha/2 = 0{,}05$

$1-\alpha/2 = 0{,}95$

En la Tabla 4 del Apéndice, para 1-$\alpha/2$ = 0,95 le corresponde un valor Z = 1,64, es decir, $Z_{1-\alpha/2} = Z_{0,95} = 1,64$.

$$\bar{y} - \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} = 885,15 - \frac{132,76}{\sqrt{39}} 1,64 = 850,28 \approx 850,30$$

$$\bar{y} + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} = 885,15 + \frac{132,76}{\sqrt{39}} 1,64 = 920$$

Expresado en términos de probabilidad se tiene que: $P[850,3 < \mu < 920] = 0,90$.

Esto significa que, la verdadera Media del peso, a las 12 semanas de edad de los cuyes de la granja estudiada, está en el intervalo comprendido entre 850,3 g y 920 g, con una probabilidad del 90%.

Tamaño de la Muestra para estimar μ

Cuando se realiza un experimento, surge la pregunta: ¿***Qué tan grande debe ser la muestra?*** Si se toma una muestra demasiado grande, se puede ocasionar desperdicio de tiempo y dinero, en cambio, si la muestra es demasiado pequeña, los resultados podrían carecer de validez, desde el punto de vista de la Inferencia Estadística.

El tamaño de la muestra necesario para estimar la media poblacional μ a partir de la media muestral $\bar{y}$, con una confiabilidad (1-α)100% y un error de estimación ***e***, está dado por:

$$n = \frac{Z_{1-\alpha/2}^2 \, \sigma^2}{e^2} \qquad (5.13)$$

Si se conoce el tamaño de la población ***N*** y éste es pequeño, se puede utilizar:

$$n = \frac{NZ_{1-\alpha/2}^2\ \sigma^2}{e^2(N-1)+Z_{1-\alpha/2}^2\ \sigma^2} \qquad (5.14)$$

Donde $Z_{1-\alpha/2}$ es el valor que se encuentra en la Tabla Normal Estandarizada para una confiabilidad de $1-\alpha$. Esta confiabilidad la fija el investigador y, por lo general, se trabaja con niveles del 95% y 99%, dependiendo de los alcances de las conclusiones y recomendaciones. Para los niveles antes indicados, los valores *Z* son 1,96 y 2,57, respectivamente.

El Error Máximo de Estimación ***e***, es la distancia máxima que podría admitirse entre la estimación de la verdadera media con respecto al parámetro $|\bar{y} - \mu|$. Esta cantidad es arbitraria y la fija el investigador.

La Desviación Estándar σ corresponde al valor de este parámetro en la población de donde se extraerá la muestra, lo que indica que, (5.13) y (5.14) se podrán utilizar para calcular el tamaño de muestra, siempre y cuando σ, sea conocida. Sin embargo, como se ha mencionado, en la mayoría de los casos, este valor no se conoce, por lo tanto, habrá necesidad de estimarlo, utilizando cualquiera de los procedimientos indicados a continuación:

a. De la población objeto de estudio se extrae una ***Muestra Piloto o Preliminar***, cuyo tamaño debe ser $n \geq 30$. Con los datos obtenidos se calcula *S* y este valor "reemplaza" a σ en las fórmulas correspondientes.

b. Se puede utilizar σ de estudios similares, realizados previamente.

c. Emplear una estimación burda de σ con la expresión ($\frac{1}{6}rango$), para lo cual se requiere conocer los valores máximo y mínimo de la variable en la población.

Ejemplo 5.6: Un estudio pretende estimar el peso medio del barbudo negro (*Rhamdia quelem*) a la edad de un año, en un estanque experimental ubicado en la Cuenca Media del Río Cauca. De antemano, se debe establecer el número de animales a capturar, sabiendo que la Desviación Estándar, para la característica en cuestión, es de 50 g, la Confiabilidad del 95% y el Error de Estimación máximo admitido de 5 gramos.

De acuerdo con los anteriores datos, ¿Cuál debe ser el tamaño de muestra para estimar la Media?

Teniendo en cuenta que la población es grande y desconocida, se utiliza (5.13), sabiendo que: σ= 50, 1-α es 95%, por lo tanto, $Z_{\alpha/2}$ = 1,96, e = 5 g:

$n = \frac{1{,}96^2 \times 50^2}{5^2}$ = 384,16 ≈ 384 animales

Ejemplo 5.7: Con los mismos datos del Ejemplo 5.6, pero utilizando un error de estimación del 10%, ¿Cuál sería el tamaño de muestra?

$n = \frac{1{,}96^2 \times 50^2}{10^2}$ = 96,04 ≈ 96 animales.

Ejemplo 5.8: Con los mismos datos del Ejemplo 5.7, pero utilizando 1-α = 99%, ¿Cuál sería el tamaño de muestra?

$n = \frac{2{,}57^2 \times 50^2}{10^2} = 165{,}12 \approx 165$ animales.

Ejemplo 5.9: Con los mismos datos del Ejemplo 5,8, pero utilizando error de estimación de 5 g, ¿Cuál sería el tamaño de muestra?

$n = \frac{2{,}57^2 \times 50^2}{5^2} = 660{,}49 \approx 660$ animales

De los ejemplos anteriores se puede establecer claramente que, entre mayor sea el nivel de confiabilidad y menor sea el error de estimación requerido, el tamaño de la muestra aumenta.

Muchas veces se trata de estimar la Media de una población, cuando se desconoce la Varianza σ^2 y es imposible obtener una muestra de tamaño $n \geq 30$, es decir, se trata de poblaciones pequeñas. En estos casos ($n < 30$), no es apropiado pretender estimar σ^2 con S^2, puesto que: $\frac{\bar{y}-\mu}{\frac{S}{\sqrt{n}}}$ no se distribuye normal estándar.

Para enfrentar este tipo de situaciones, es necesario introducir una nueva distribución denominada ***Distribución t de Student.***

Distribución *t* de Student. Dada una muestra aleatoria de tamaño ***n***, procedente de una población normal con Media μ, la estadística:

$$\frac{\bar{y}-\mu}{\frac{S}{\sqrt{n}}} = t \qquad (5.15)$$

sigue una distribución ***t de Student***, con n-1 grados de libertad.

Entre las propiedades de la Distribución *t de Student,* se tienen las siguientes:

a. Es simétrica respecto al centro (media).

b. La forma es similar a la normal, pero es menos espigada en el centro (Mesocúrtica) y tiene colas más largas (Figura 5.6).

c. El rango está entre $-\infty$ y $+\infty$.

d. La Distribución *t de Student* se aproxima a la distribución normal a medida que los grados de libertad n-1 se aproximan a infinito.

e. El Valor Esperado o Media es cero, $E[t] = 0$.

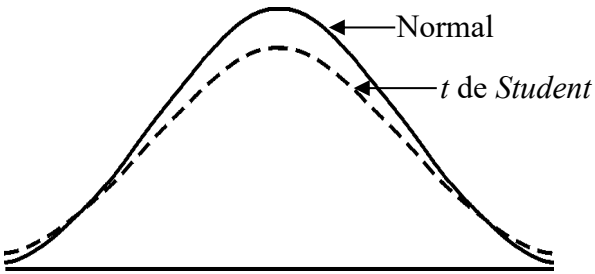


Figura 5.6 Curva de Densidad de *t* de *Student*

La Varianza de una variable con distribución t, con υ grados de libertad, es

$$V(t_\upsilon) = \frac{\upsilon}{\upsilon-2}, \quad \upsilon > 2 \qquad (5.16)$$

La ***Función de Distribución Acumulativa*** está dada por $P[T \leq t_{1-\alpha,\ \upsilon}]=1-\alpha$, donde $0 \leq \alpha \leq 1$ y $t_{1-\alpha,\ \upsilon}$ es el valor cuantil que le corresponde a la distribución t para $1-\alpha$ y υ dados (Figura 5.7).

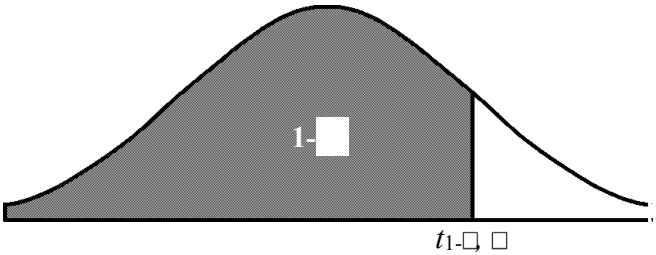


Figura 5.7 Función de Distribución Acumulativa de *t de Student*

Así, para $P[T \leq t_{0,90,\ 15}] = 0,90$ y $t_{0,90\ ;\ 15} = 1,341$

En la Tabla 7 del Apéndice, en el cruce entre $\upsilon = 15$ y $1\text{-}\alpha = 0,90$, se encuentra el valor 1,341, como se ilustra en el aparte requerido de dicho anexo, como se muestra a continuación y se representa en la Figura 5.8.

υ	$t_{0,800}$	$t_{0,900}$	$t_{0,950}$	$t_{0,975}$	$t_{0,990}$	$t_{0,995}$	$t_{0,999}$
1	1,376	3,078	6,314	12,706	31,821	63,657	318,309
2	1,061	1,886	2,920	4,303	6,965	9,925	22,327
3	0,978	1,638	2,353	3,182	4,541	5,841	10,215
4	0,941	1,533	2,132	2,776	3,747	4,604	7,173
5	0,920	1,476	2,015	2,571	3,365	4,032	5,893
6	0,906	1,440	1,943	2,447	3,143	3,707	5,208
7	0,896	1,415	1,895	2,365	2,998	3,499	4,785
8	0,889	1,397	1,860	2,306	2,896	3,355	4,501
9	0,883	1,383	1,833	2,262	2,821	3,250	4,297
10	0,879	1,372	1,812	2,228	2,764	3,169	4,144
11	0,876	1,363	1,796	2,201	2,718	3,106	4,025
12	0,873	1,356	1,782	2,179	2,681	3,055	3,930
13	0,870	1,350	1,771	2,160	2,650	3,012	3,852
14	0,868	1,345	1,761	2,145	2,624	2,977	3,787
15	0,866	**1,341**	1,753	2,131	2,602	2,947	3,733

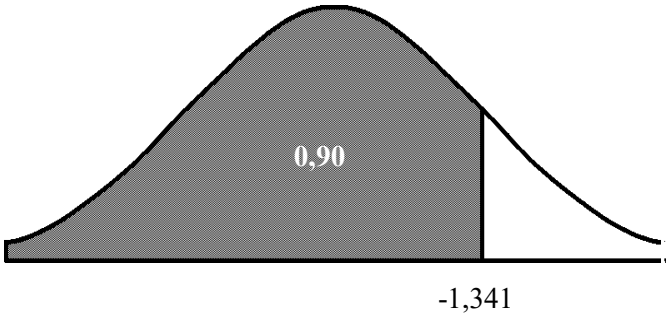


Figura 5.8 Ubicación del Valor de *t* en la Curva para 1-α = 0,90 y υ = 15

De esta manera, $P[T \leq t_{0,10,\ 15}] = P[T \leq -1,341] = 0,10$ (Figura 5.9).

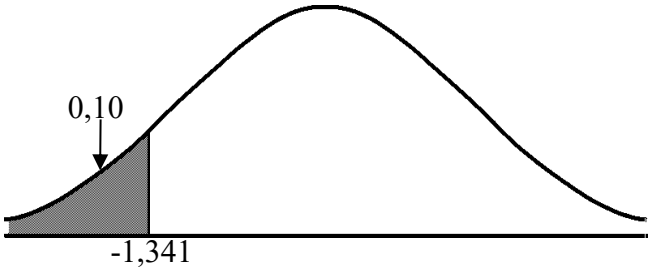


Figura 5.9 Ubicación del Valor de *t* para $P[T \leq t_{0,10,\ 15}]$

Intervalo de Confianza para μ con una muestra pequeña y Varianza σ^2 desconocida

Dada una muestra aleatoria de tamaño n<30, con Media $\bar{y}$ y Varianza S^2, procedente de una población aproximadamente normal con Media μ y Varianza σ^2 desconocida, un Intervalo de Confianza para μ al (1-α) 100% de confiabilidad viene dado por:

$$\bar{y} - t_{\alpha/2,\ v}\ \frac{S}{\sqrt{n}} < \mu < \bar{y} + t_{1-\frac{\alpha}{2};\ v}\ \frac{S}{\sqrt{n}} \qquad (5.17)$$

Donde $t_{1-\alpha/2,\ v}$ es el valor de t con $v = n - 1$ grados de libertad y $P[T \leq t_{1-\frac{\alpha}{2};\ v}] = 1 - \alpha/2$.

Ejemplo 5.10: Un veterinario tomó una muestra al azar, de 15 perros de raza Labrador Dorado, para estimar el consumo medio de alimento en un periodo de un mes, calculando una Media de 22,5 kg y una Varianza de 35 kg^2; se pregunta ¿Cuál es el intervalo de confianza, al 95% de confiabilidad, para el verdadero consumo medio? De acuerdo con (5.17), la respuesta está dada por el desarrollo de las siguientes operaciones:

$$\bar{y} = 22{,}5$$
$$S^2 = 35$$
$$S = \sqrt{35} \ = 5{,}916$$
$$IC = \bar{y} \pm \frac{S}{\sqrt{n}} t_{1-\alpha/2} = 22{,}5 \pm \frac{5{,}916}{\sqrt{14}}\ 2{,}160 = 22{,}5 \pm 3{,}415$$

Este mismo Intervalo de Confianza expresado como una Probabilidad, según lo indicado en (5.17) se indica del siguiente modo:

$$P\left[22{,}5 - \frac{5{,}916}{\sqrt{14}} \times 2{,}16 < \mu < 22{,}5 + \frac{5{,}916}{\sqrt{14}} \times 2{,}16\right] = 95\%$$

$$P[19{,}085 < \mu < 25{,}915] = 95\%$$

Así se indica que, la Probabilidad de que la verdadera media de consumo se encuentre entre 19,085 y 25,915 kg es del 95%, es decir, en una serie grande de muestreos sucesivos sobre la misma población, el valor medio calculado se encontrará entre los límites indicados, en el 95% de las veces.

Ejemplo 5.11: En el problema planteado en el Ejemplo 5.10, asumiendo que no es posible tomar una muestra de más de 10 animales, con los cuales se calcula un consumo de 19,7 kg y una Varianza de 8 kg^2, ¿Cuál es el Intervalo de Confianza con un 95% de confiabilidad?

El procedimiento para resolverlo es exactamente igual. Se dispone del valor medio y de la Varianza; con este último dato, se calcula el error estándar que, en este caso, es de 0,2828, por lo tanto, el Intervalo de Confianza se calcula de la siguiente manera:

$$P(19{,}7 - (0{,}2828 \times 2.262) < \mu < 19{,}7 + (0{,}2828 \times 2{,}262)) = 95\%$$

$$P(19{,}0603 < \mu < 20{,}3396) = 95\%$$

En este caso, se interpreta este resultado como una Probabilidad, cuyo significado es que, la verdadera media se encuentra entre 19,0603 y 20,3396, es decir, en una serie grande de muestreos sucesivos sobre la misma población, el valor medio calculado se encontrará entre los límites indicados, en el 95% de las veces.

Intervalo de Confianza para la Diferencia de Medias

Un problema que se presenta con frecuencia en Inferencia Estadística se refiere a la comparación entre dos medias poblacionales.

Para comparar las medias de dos poblaciones, es necesario extraer dos muestras aleatorias independientes, una de cada población, con tamaños n_1 y n_2 y la inferencia sobre μ_1 - μ_2 se basa en los resultados muestrales de $\bar{y}_1 - \bar{y}_2$.

Si las poblaciones de donde provienen las muestras son normales e independientes, con Varianzas σ_1^2 y σ_2^2 conocidas, la distribución muestral de $(\bar{y}_1 - \bar{y}_2)$ es normal con media (μ_1 - μ_2) y Varianza $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$. Por lo tanto, la variable aleatoria

$$\frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = Z \qquad (5.18)$$

es normal estándar.

Entonces, si n_1 y n_2 son grandes, la distribución de $(\bar{y}_1 - \bar{y}_2)$ será aproximadamente normal, con Media (μ_1 - μ_2) y Varianza $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$, sin tener en cuenta las distribuciones de las poblaciones de donde se extrajeron las muestras.

Intervalo de Confianza para la Diferencia de Medias (μ_1 - μ_2), con Varianzas σ_1^2 y σ_2^2 conocidas

Dadas dos muestras aleatorias con tamaños n_1 y n_2 y con medias muestrales $\bar{y}_1$ y $\bar{y}_2$ respectivamente, procedentes de dos poblaciones normales e independientes, con Medias μ_1 y μ_2 y Varianzas conocidas σ_1^2 y σ_2^2 respectivamente, un intervalo de confianza para μ_1 y μ_2 al (1-α)100% de confiabilidad viene dado por:

$$(\bar{y}_1 - \bar{y}_2) - Z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < (\mu_1 - \mu_2) < (\bar{y}_1 - \bar{y}_2) + Z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (5.19)$$

Donde $Z_{1-\alpha/2}$ es tal que $P[Z<z_{1-\alpha/2}] = 1 - \alpha/2$.

Si las poblaciones no son normales, se obtiene un intervalo de confianza bastante aceptable cuando n_1 y $n_2 \geq 30$; además, es posible reemplazar σ_1^2 y σ_2^2 por S_1^2 y S_2^2, respectivamente.

Ejemplo 5.12: Un profesional especialista en mejoramiento genético, asegura que, gracias a su programa de selección genética, ha obtenido un nuevo fenotipo de *Oreochromis aureus*, cuyo rendimiento medio en peso final comercial a la cosecha, es superior al fenotipo estándar de la misma especie. Se sabe que las Varianzas para cada una de las especies son $\sigma_1^2 = 0{,}012$ y $\sigma_2^2 = 0{,}018$. Se tomó una muestra al azar de 90 ejemplares para cada uno de los grupos, calculando un valor medio de $\bar{y}_1 = 0{,}523$ kg y $\bar{y}_2 = 0{,}475$ kg. Se pregunta ¿Cuál es el límite de confianza para la verdadera diferencia de medias entre el grupo mejorado y el grupo estándar?

$$(0{,}523 - 0{,}475) - 1{,}96\sqrt{\frac{0{,}012}{90} + \frac{0{,}018}{90}} < (\mu_1 - \mu_2) < (0{,}523 - 0{,}475) + 1{,}96\sqrt{\frac{0{,}012}{90} + \frac{0{,}018}{90}}$$

$$P(0{,}048 - 1{,}96(0{,}01825) < (\mu_1 - \mu_2) < 0{,}048 + 1{,}96(0{,}01825))$$

$P(0{,}0122 < (\mu_1 - \mu_2) < 0{,}0838) = 95\%$

Según esto, con una confiabilidad del 95%, se asegura, que la diferencia de las medias de los dos grupos, está entre 0,0122 y 0,0837 kg.

Por otra parte, puesto que el intervalo de confianza no incluye el 0 (cero) y es superior a éste, en el 95% de los casos, es posible afirmar que existen diferencias significativas entre las medias del peso final comercial a la cosecha de las dos poblaciones, mostrando superioridad del nuevo fenotipo de la especie en evaluación.

Si σ_1^2 y σ_2^2 no se conocen, pero se pueden suponer iguales $\sigma_1^2 = \sigma_2^2 = \sigma^2$ y las muestras son pequeñas n_1 y $n_2 < 30$, se recurre a la distribución ***t de Student***.

Si el muestreo se lleva a cabo sobre poblaciones normales e independientes, con Varianzas iguales pero desconocidas $\sigma_1^2 = \sigma_2^2 = \sigma^2$, la variable aleatoria

$$\frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = T \qquad (5.20)$$

se distribuye ***t*** con (n_1 + n_2 – 2) grados de libertad. S_p^2 es un estimador combinado de la Varianza común σ^2, así:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \qquad (5.21)$$

Intervalo de Confianza para la Diferencia de Medias (μ_1 - μ_2) con Muestras pequeñas y Varianzas iguales $\sigma_1^2 = \sigma_2^2$ pero desconocidas

Dadas dos muestras aleatorias con tamaños n_1 y $n_2 < 30$ y medias muestrales $\bar{y}_1$ y $\bar{y}_2$, respectivamente, procedentes de dos poblaciones normales e independientes con medias μ_1 y μ_2 y Varianzas desconocidas pero iguales $\sigma_1^2 = \sigma_2^2$ respectivamente, con un intervalo de confianza μ_1 - μ_2 al (1-α)100% de confiabilidad, viene dado por:

$$(\bar{y}_1 - \bar{y}_2) - t_{1-\alpha/2;\,\upsilon}\ \ S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < (\mu_1 - \mu_2) < (\bar{y}_1 - \bar{y}_2) + t_{1-\alpha/2;\,\upsilon}\ \ S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \qquad (5.22)$$

Donde: $t_{1-\alpha/2;\,\upsilon}$ es el valor de ***t*** con $\upsilon = n_1 + n_2 - 2$ grados de libertad y

$$P[T \leq t_{1-\alpha/2;\,\upsilon}] = 1\text{-}\alpha/2.$$

S_p^2 es un estimador combinado de la Varianza común $\sigma_1^2 = \sigma_2^2 = \sigma^2$ que se expresa como:

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \qquad (5.23)$$

El intervalo es bastante aceptable, si las poblaciones de donde provienen las muestras son aproximadamente normales.

Ejemplo 5.13: En una región con poca tradición en la ovinocultura, se tomó dos muestras conformadas por 11 ejemplares machos, en dos fincas. En cada una de las explotaciones se calculó las

respectivas Medias y Varianzas, para la variable **peso a la edad adulta**, encontrando los siguientes resultados:

Finca 1: $n_1 = 11$
$\bar{y}_1 = 56{,}5$ kg
$S_1^2 = 146{,}8$ kg^2

Finca 2: $n_1 = 11$
$\bar{y}_1 = 57{,}8$ kg
$S_1^2 = 154{,}2$ kg^2

Se pregunta, ¿Cuál es el límite de confianza para la diferencia de Medias de las dos poblaciones?

Para responder este interrogante, se aplica la fórmula contenida en (5.22), donde será necesario, además, utilizar (5.23) para obtener un estimativo del error estándar de la diferencia de medias S_p^2.

$$S_p^2 = \frac{(11-1)146{,}8 + (11-1)154{,}2}{11+11-2} = 150{,}5$$

$$S_p = \sqrt{S_p^2} = \sqrt{150{,}5} = 12{,}27 \text{ kg}$$

Como ya se dijo, el intervalo de confianza se calcula con (5.22):

$$(56{,}5 - 57{,}8) - (2{,}086 \times 12{,}27)\sqrt{\frac{1}{11}+\frac{1}{11}} < (\mu_1 - \mu_2)$$

$$< (56{,}5 - 57{,}8) + (2{,}086 \times 12{,}27)\sqrt{\frac{1}{11}+\frac{1}{11}}$$

$$-1{,}3 - 10{,}914 < (\mu_1 - \mu_2) < -1{,}3 + 10{,}914$$

> -12,214 < $(\mu_1 - \mu_2)$< 9,614
>
> Esto indica que, la verdadera diferencia de las dos Medias en toda la población estudiada, en una serie grande de muestreos el 95% de las muestras estará contenida en un rango que va de –12,214 a 9,614 kg.
>
> También, como cero pertenece al intervalo, entonces $\mu_1 - \mu_2 = 0$, es decir $\mu_1 = \mu_2$, se puede interpretar, como la inexistencia de diferencias significativas entre las medias del peso a la edad adulta de las dos poblaciones de ovinos, con el 95% de confiabilidad.

Intervalo de Confianza para Proporciones

Es útil, cuando se está interesado en estimar la proporción de individuos de una población, con una determinada característica, como es el caso de establecer la proporción de animales vacunados contra la fiebre aftosa que no desarrollaron la enfermedad, las vacas positivas a cualquier enfermedad o los peces que sobreviven a una variación térmica brusca y muchos otros casos similares.

Se debe recordar que, ***p*** es la proporción de individuos poseedores de cierta característica en una población de interés y un estimador de ***p*** es la proporción muestral $\hat{p} = \frac{X}{n}$, donde *X* es el número de éxitos en ***n*** ensayos independientes, siempre y cuando ***n*** sea grande y la variable sea aleatoria; entonces:

$$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = Z \qquad (5.24)$$

es normal estándar.

Intervalo de Confianza para *p* con Muestra grande

Sea *p* la proporción observada de éxitos en una muestra aleatoria de tamaño ***n***, un intervalo de confianza para ***p*** al (1-α)100% de confiabilidad, viene dado por:

$$\hat{p} - Z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + Z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad (5.25)$$

Donde $Z_{1-\alpha/2}$ es tal que $P[Z < z_{1-\alpha/2}] = 1-\alpha/2$.

El intervalo es confiable cuando ***n*** es grande y *np* *n*(1-*p*) están demasiado cercanos a 1 o a 0. Generalmente se considera que, estas condiciones se satisfacen si *np* y *n*(1-*p*) son mayores que 5 (Canavos, 1986).

Ejemplo 5.14: En un estudio de prevalencia por muestreo, realizado en Colombia, 150 vacas de la raza Guernesey resultaron positivas a brucelosis, en una muestra de 355 animales. ¿Cuál es el límite de confianza, para la verdadera proporción de animales enfermos con un α = 5%?

En primer lugar se debe calcular $\hat{p} = \frac{150}{355} = 0{,}42$

Por lo tanto 1-*p* = *q* = 1 – 0,42 = 0,58.

Para establecer el límite de confianza se aplica (5.5.18), por lo tanto:

$$0{,}42 - 1{,}96\sqrt{\frac{0{,}42 \times 0{,}48}{355}} < p < 0{,}42 + 1{,}96\sqrt{\frac{0{,}42 \times 0{,}48}{355}}$$

0,42 – 0,047 < *p* < 0,42 + 0,047

0,373 < *p* < 0,467

> Por consiguiente, se puede asegurar, que en la zona estudiada, la prevalencia se encuentra entre 0,373 y 0,467. Estos valores, expresados en términos porcentuales, indican que las hembras positivas a brucelosis, están entre 37,3 y 46,7%, afirmación que se hace con el 95% de confiabilidad.

Tamaño de Muestra para estimar p

El tamaño de muestra necesario para estimar la proporción poblacional $\boldsymbol{p}$ a partir de la proporción muestral $\hat{p}$, con una confiabilidad del (1-α)100% y un error de estimación $\boldsymbol{e}$ está dado por:

$$n = \frac{Z_{1-\alpha/2}^2 \; p(1-p)}{e^2} \qquad (5.26)$$

Donde $Z_{1-\alpha/2}$ es el valor tabular de la distribución normal estándar para un valor α dado.

$\boldsymbol{p}$ es la proporción de individuos con la característica de interés para el investigador.

$\boldsymbol{e}$ es el error máximo admitido para la estimación del parámetro.

Como en el caso del tamaño de muestra para estimar la Media, la confiabilidad 1-α y el error de estimación $\boldsymbol{e}$ son arbitrarios y los fija el investigador, dependiendo de los alcances e implicaciones de sus resultados y conclusiones. Para proporciones, el error máximo de estimación $\boldsymbol{e}$ es una medida que indica, la distancia a entre la proporción estimada y la verdadera $|p - \hat{p}|$.

Cuando ***p*** no se conoce, se asume que ***p* = 0,5** y, de esta manera, se obtendrá el máximo valor de ***n***, procedimiento que finalmente permitirá alcanzar la confiabilidad y la dimensión del intervalo deseado.

Si el tamaño de la población es relativamente pequeño y ***n* > 0,05*N***, se puede utilizar el factor de corrección para poblaciones finitas:

$$n_1 = \frac{n}{1+\frac{n}{N}} \qquad (5.27)$$

Ejemplo 5.15: Un Médico Veterinario debe establecer la prevalencia de la neumonía enzoótica en una población de cerdos. Los registros de una entidad oficial indican que, en total existen 2.800 animales. Se asume que la enfermedad ataca por igual a machos y hembras de cualquier edad y no existe información previa sobre la proporción de animales enfermos ($\hat{p}$). El profesional pregunta: ¿Cuál es el tamaño de muestra para estimar la prevalencia de la enfermedad, con una confiabilidad del 95% y un error de estimación del 5%?

En primer lugar, se debe aplicar (5.27), con p = 0,5 y α = 0,05:

$$n = \frac{Z^2_{1-\alpha/2}\ p(1-p)}{e^2} = \frac{1{,}96^2\ (0{,}5)\ \ (0{,}5)}{0{,}05^2} = \frac{0{,}9604}{0{,}0025} = 384{,}16$$

Debido a que no es posible muestrear fracciones de animales, es necesario aproximar el tamaño de muestra requerido a 384 cerdos que se escogerán irrestrictamente al azar, lo que indica que, todos los individuos tienen la misma probabilidad de ser elegidos y formar parte de la muestra.

Por otra parte, $N \times 0{,}05 = 2.800 \times 0{,}05 = 140$ y $n = 384$; estas cifras indican, que el tamaño inicial de muestra n es mayor que $N(0{,}05)$, por lo tanto, es necesario hacer la corrección por tamaño finito, aplicando (5.27):

$$n_1(\text{tamaño final de la muestra}) = \frac{384}{1+\frac{384}{2.800}} = \frac{384}{1{,}13714} = 337{,}69$$

En consecuencia, el Veterinario deberá tomar muestras en 338 animales.

Ejemplo 5.16: Con los mismos datos del Ejemplo 5.15, pero asumiendo que se conoce un dato de prevalencia esperada del 20%, ¿Cuál será el tamaño de la muestra?

$$n = \frac{1{,}96^2 \times 0{,}2 \times 0{,}8}{0{,}05^2} = \frac{0{,}614656}{0{,}0025} = 245{,}86 \approx 246$$

$$n_1(tamaño\ final\ de\ la\ muestra) = \frac{246}{1+\frac{246}{2.800}} = \frac{246}{1{,}0878}$$

$$= 226{,}13$$

En este caso, la cantidad de animales a muestrear es 226.

Ejemplo 5.17: Con los mismos datos del Ejemplo 5.16, pero con una confiabilidad del 90% y error de estimación del 10%, ¿Cuál será el tamaño de la muestra?

$$n = \frac{1{,}64^2 \times 0{,}2 \times 0{,}8}{0{,}1^2} = \frac{0{,}430336}{0{,}01} = 43{,}034$$

En este caso, no se corrige por tamaño finito, dado que $n < 0{,}05N$, en consecuencia, el tamaño final de muestra será de 43 animales.

Estimación de la Diferencia entre dos Proporciones

Muchas veces, el investigador está interesado en comparar dos proporciones provenientes de dos poblaciones independientes, para establecer si estadísticamente son iguales o diferentes. Así, en el caso de probar la eficacia de dos tratamientos para el control de la saprolegnia en truchas, cuando se utiliza yodo o azul de metileno, el interés se centra en estimar la diferencia entre las proporciones de animales curados en cada uno de los tratamientos. Para cumplir con este objetivo, se selecciona dos muestras aleatorias independientes, de tamaño n_1 y n_2, a partir de dos poblaciones binomiales, ya que los resultados, son que el animal se cure de la enfermedad o no.

Un estimador de $p_1 - p_2$ está dado por $\hat{p}_1 - \hat{p}_2$, donde $\hat{p}_1 = \frac{X_1}{n_1}$ es la proporción muestral de individuos, que poseen la característica de interés en la primera población y $\hat{p}_2 = \frac{X_2}{n_2}$ es la proporción muestral en la segunda población.

Para muestras grandes n_1 y n_2, la distribución de $\hat{p}_1 - \hat{p}_2$ se aproxima a la normal, con proporción $p_1 - p_2$ y varianza $\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}$, es decir,

$$\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} = Z \qquad (5.28)$$

Intervalo de Confianza entre dos Proporciones $p_1 - p_2$ con muestras grandes

Sea $\hat{p}_1$ la proporción de éxitos observados en una muestra aleatoria de tamaño n_1 procedente de una población con proporción de éxitos p_1, y sea $\hat{p}_2$ la proporción de éxitos observados de una muestra aleatoria independiente, de tamaño n_2, procedente de una población con proporción de éxitos p_2. Si los

tamaños de muestras n_1 y n_2 son grandes, un intervalo de confianza para $p_1 - p_2$ al (1-α)100% de confiabilidad, viene dado por:

$$(\hat{p}_1 - \hat{p}_2) - Z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + Z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

(5.29)

Donde $Z_{1-\alpha/2}$ es tal que $P\left(Z < z_{1-\alpha/2}\right) = 1 - \frac{\alpha}{1}$

Ejemplo 5.18: Un Ingeniero en Producción Acuícola utilizó dos tratamientos para el control de la Aspergillosis en Molinesias, uno con sal marina y otro con formaldehído. Para este propósito, distribuyó cada uno de los dos tratamientos en cinco acuarios, con 50 ejemplares afectados en cada acuario, es decir, en total fueron tratados 250 animales en cada grupo. Al final del experimento se constató, que la sal marina curó 96 peces en el primer grupo y el formaldehído 103 en el segundo grupo. ¿Cuál es el límite de confianza para la verdadera diferencia de las proporciones de animales curados en cada tratamiento con confiabilidad del 95%?

En primer lugar, se calculan las proporciones de éxito en cada tratamiento, así:

$$\hat{p}_1 = \frac{96}{250} = 0{,}384$$

$$\hat{p}_2 = \frac{103}{250} = 0{,}412$$

De acuerdo con (5.29):

$$(0{,}384 - 0{,}412) - 1{,}96\sqrt{\frac{0{,}384(0{,}616)}{250} + \frac{0{,}412(0{,}588)}{250}} < (p_1 - p_2) < (0{,}384 - 0{,}412) + 1{,}96\sqrt{\frac{0{,}384(0{,}616)}{250} + \frac{0{,}412(0{,}588)}{250}}$$

$-0{,}028 - 1{,}96(0{,}04376) < (p_1 - p_2) < -0{,}028 + 1{,}96(0{,}04376)$

$-0{,}1137696 < (p_1 - p_2) < 0{,}0577696$

En conclusión, con el 95% de confianza, se asegura que la verdadera diferencia de las proporciones se encuentra entre –0,1138 y 0,058.

Adicionalmente se puede concluir que, en el 95% de los casos, no existen diferencias significativas en el efecto de los dos tratamientos sobre el control de la Aspergilosis en Molinesias, dado que el intervalo de confianza incluye el cero.

Intervalo de Confianza para la Varianza

Es necesario recordar que si $S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$ es la Varianza de una muestra aleatoria de tamaño ***n***, procedente de una población normal con Varianza σ^2, entonces, la distribución de muestreo $(n-1)\frac{S^2}{\sigma^2}$ corresponde a una distribución de Chi-cuadrado, con $(n - 1)$ grados de libertad.

Intervalo de Confianza para σ^2. Si S^2 es la Varianza de una muestra aleatoria de una población normal con Media μ y Varianza σ^2, un intervalo de confianza para σ^2 al $(1-\alpha)100\%$ de confiabilidad, está dado por:

$$\frac{(n-1)S^2}{\chi^2{}_{1-\alpha/2;\, n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2{}_{\alpha/2;\, n-1}} \qquad (5.30)$$

Donde $\chi^2{}_{1-\alpha/2}$ es tal que $P(\chi^2 \leq \chi^2{}_{1-\alpha/2}) = 1\text{-}\alpha/2$.

Ejemplo 5.19: Con el fin de establecer un Programa de producción piscícola con *Oreochromis sp*, se midió el caudal de una fuente de agua durante un período de seis meses, tomando una muestra semanal. Al final del período, se estimó una varianza de 42 L^2/seg. ¿Cuál es el límite de confianza para la varianza, con una confiabilidad del 95%?

Teniendo en cuenta que se tomó 24 muestras, de acuerdo con (5.30); dado que $1 - \alpha = 0{,}95$; entonces $\alpha = 0{,}05$; $\alpha/2 = 0{,}025$; $1 - \alpha/2 = 0{,}975$.

Así, consultando la Tabla 6 del Apéndice, en el aparte correspondiente que se muestra enseguida, para $\chi^2_{0{,}975}$ y $v = 23$, el valor cuantil es 38,09. De igual manera, para $\chi^2_{0{,}025}$ y $v = 23$, el valor cuantil es 11,69.

v	$x_{0,005}$	$x_{0,010}$	$x_{0,025}$	$x_{0,050}$	$x_{0,100}$	$x_{0,900}$	$x_{0,950}$	$x_{0,975}$	$x_{0,990}$	$x_{0,995}$
1	0,00	0,00	0,00	0,00	0,02	2,71	3,84	5,02	6,63	7,88
2	0,01	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	10,60
3	0,07	0,11	0,22	0,35	0,58	6,25	7,81	9,35	11,34	12,84
4	0,21	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	14,86
1	6,84	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	38,58
2	7,43	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	40,00
2	8,03	8,90	10,28	11,59	13,24	29,62	32,67	35,48	38,93	41,40
2	8,64	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29	42,80
2	9,26	10,20	**11,69**	13,09	14,85	32,01	35,17	**38,08**	41,64	44,18

$\chi^2_{1-\alpha/2;\, n-1} = \chi^2_{0,975;\, 23} = 38,09$

$\chi^2_{\alpha/2;\, n-1} = \chi^2_{0,025;\, 23} = 11,69$

Aplicando (5.5.20) $\frac{(n-1)S^2}{\chi^2_{1-\alpha/2;\, n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2;\, n-1}}$

$$\frac{(24-1)42}{38,09} < \sigma^2 < \frac{(24-1)42}{11,69}$$

$25,36 < \sigma^2 < 82,63$

Con el 95% de confianza, el intervalo para la varianza está entre 25,36 y 82,63 L^2/seg.

Intervalo de Confianza para el Cociente de dos Varianzas

Muchas veces, surge la necesidad de comparar la variabilidad de dos procesos distintos, es decir, se está interesado en comparar σ_1^2 y σ_2^2. Una manera de hacerlo, es obtener su cociente σ_1^2/σ_2^2, entonces, si las dos Varianzas son iguales, dicho cociente será igual a 1.

Para la construcción de un intervalo de confianza, para σ_1^2/σ_2^2, es necesario introducir la ***Distribución de Probabilidad de F***, donde, si S_1^2 y S_2^2 son las Varianzas de dos muestras aleatorias independientes de tamaños n_1 y n_2, procedentes de dos poblaciones normales e independientes con Varianzas σ_1^2 y σ_2^2, respectivamente, a partir de lo cual, se tiene una variable aleatoria *F*, tal que:

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = F \qquad (5.31)$$

que sigue una distribución ***F*** con (n_1 – 1) grados de libertad en el numerador y (n_2 – 1) grados de libertad en el denominador (Figura 5.10).



Figura 5.10 Distribución de *F*

En esta distribución de Probabilidad:

$E[F_{\upsilon 1,\, \upsilon 2}] = \frac{\upsilon_2}{\upsilon_2 - 2}, \quad \upsilon_2 > 4$

$V[F_{\upsilon 1,\, \upsilon 2}] = \frac{\upsilon_2^2(2\upsilon_2 + 2\upsilon_1 - 4)}{\upsilon_1(\upsilon_2 - 2)^2(\upsilon_2 - 4)}, \quad \upsilon_2 > 4$

En la Tabla 8 del Apéndice, se muestran los valores $F_{1-\alpha,\, \upsilon 1,\, \upsilon 2}$, tales que:

$P[F \leq f_{1-\alpha,\, \upsilon 1,\, \upsilon 2}] = 1-\alpha$, que corresponde a la Función de Distribución Acumulada (Figura 5.11).

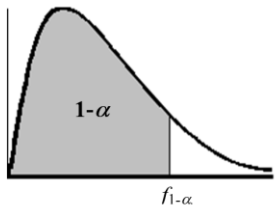


Figura 5.11 Representación de la Función de Distribución Acumulada de *F*.

Ejemplo 5.20: Calcular $P[F \leq f_{0,90,\,5,\,10}]$.

En la Tabla 8 del Apéndice, para 1-α = 0,90 con 5 grados de libertad en el numerador y 10 grados de libertad en el denominador, se encuentra que:

$f_{0,90,\,5,\,10}$ = 2,52, por lo tanto $P[F \leq 2{,}52] = 0{,}90$, como se muestra en el siguiente esquema.



En la citada tabla se encuentran valores para $\alpha < 0{,}5$. Si se desea para $\alpha > 0{,}5$, se aprovecha la siguiente propiedad:

$$f_{1-\alpha,\,\upsilon 1,\,\upsilon 2} = \frac{1}{f_{\alpha,\,\upsilon 2,\,\upsilon 1}}.$$

En el siguiente caso:

$$P[f \leq f_{0,10,\,10,\,5}] = P\left[F \leq \frac{1}{f_{0,9,\,5,\,10}}\right] = P\left[F \leq \frac{1}{2{,}52}\right] = P[F \leq 0{,}3968] = 0{,}10$$

Los anteriores conceptos sirven de base para calcular el ***Intervalo de Confianza para el Cociente de Dos Varianzas*** σ_1^2/σ_2^2. Cuando S_1^2 y S_2^2 son las Varianzas de dos muestras aleatorias independientes de tamaños n_1 y n_2, respectivamente, procedentes de dos poblaciones normales con Varianzas σ_1^2 y σ_2^2, respectivamente, un intervalo de confianza para σ_1^2/σ_2^2, al (1-α)100% de confiabilidad, está dado por:

$$\frac{S_1^2}{S_2^2}\frac{1}{f_{1-\alpha/2,\,\upsilon 1,\,\upsilon 2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2,\,\upsilon 1,\,\upsilon 2}} \qquad (5.32)$$

Donde $f_{\alpha/2,\,\upsilon 1,\,\upsilon 2}$ es un valor f con $\upsilon_1 = n_1 - 1$ y $\upsilon_2 = n_2 - 1$ grados de libertad y $f_{\alpha/2}$ y $f_{1-\alpha/2}$ son tales que $P[F \leq f_{\alpha/2}] = 1 - \alpha/2$ y $P[F \leq f_{1-\alpha/2}] = \alpha/2$.

Ejemplo 5.21: En una población de *Cavia porcellus,* se estimó la Varianza para el peso al destete en 64,49 g^2, medida calculada sobre una muestra de 25 individuos de una línea mejorada; paralelamente, en otra población, con muestra de 31 individuos, el valor de la Varianza fue de 61,72 g^2. Se pide calcular el intervalo de confianza para el cociente de Varianzas, con un $\alpha = 0{,}10$.

El procedimiento para resolver el problema propuesto aplicando (5.32):

$\alpha = 0{,}10$

$\alpha/2 = 0{,}05$

$1 - \alpha/2 = 0{,}95$

$f_{1-\frac{\alpha}{2};\,\upsilon 1;\,\upsilon 2} = f_{0{,}95;\,24;\,30} = 1{,}89$

$f_{1-\frac{\alpha}{2};\,\upsilon 2;\,\upsilon 1} = 1/f_{0{,}95;\,30;\,24} = 1{,}94$

Con esto se tiene:

$$\frac{64{,}49}{61{,}72} \times \frac{1}{1{,}89} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{64{,}49}{61{,}72} \times 1{,}94$$

$$0{,}528 < \frac{\sigma_1^2}{\sigma_2^2} < 2{,}027$$

> De acuerdo con los resultados, el Intervalo de Confianza para el cociente de las dos Varianzas está entre 0,5528 y 2,0227, con un 95% de confiabilidad. Dado que 1 está dentro del intervalo, es decir, $\sigma_1^2/\sigma_2^2 = 1$, se puede afirmar que, no existen diferencias significativas, por lo tanto, las dos Varianzas son iguales.

CAPÍTULO 6

Prueba o contraste de hipótesis

En el capítulo anterior, se analizó la estimación puntual y por intervalo de un parámetro. En el presente capítulo se considera otro enfoque de la Inferencia Estadística, referida a la toma de decisiones sobre las poblaciones, basada en el análisis de una muestra. Los dos conceptos se fundamentan en la evidencia experimental que se obtiene a través de un muestreo aleatorio. Muchas de estas decisiones tienen que ver con aceptar o rechazar una afirmación sobre un parámetro o sobre la forma de la distribución de la población de interés, las cuales se denominan ***Hipótesis Estadísticas*** y al procedimiento, mediante el cual, se toma la decisión sobre la hipótesis se lo conoce como ***Prueba*** o ***Contraste de Hipótesis***.

Una Hipótesis Estadística es una suposición sobre un parámetro, verdadera a la luz del conocimiento científico existente sobre el tema en estudio, la cual, requiere ser probada para que pueda verificarse su validez.

Ejemplo 6.1: Un procesador de leche afirma que, el producto de su empresa contiene en promedio 0,36 kilogramos de grasa por litro. Para verificar esta afirmación, se toma una muestra aleatoria de la leche procesada en la planta y se determina la cantidad de grasa por litro, con base en cuyos resultados muestrales, se generaliza a toda la población, en este caso, a toda la leche procesada en la planta.

Ejemplo 6.2: Un profesional tiene que definir, sobre la base de experimentos, la efectividad del yodo y el azul de metileno en el control de Saprolegnia en trucha Arco Iris. Los animales afectados por el hongo se dividen en dos grupos al azar: el uno es tratado con yodo y el otro con azul de metileno. Al terminar el ensayo, es posible establecer la proporción de animales curados en cada grupo y, mediante la Prueba de Hipótesis, determinar si existen diferencias significativas entre dichas proporciones, como una medición de la efectividad de los dos fármacos.

Para abordar este tema con mayor claridad, es necesario recurrir a ciertas definiciones como son: Hipótesis Nula, Hipótesis Alterna, Región Crítica, Error Tipo I, Error Tipo II, Nivel de Significancia y Potencia de la Prueba.

En primer lugar, se debe tener en cuenta que, en el análisis de este tipo de problemas, se considera dos clases de Hipótesis Estadísticas: la Hipótesis Nula (H_0) y la Hipótesis Alterna (H_1).

La ***Hipótesis Nula*** es una afirmación o explicación tentativa sobre un fenómeno particular, la cual debe ser comprobada y puede referirse al conocimiento o experiencia actual o existente. El rechazo de una Hipótesis Nula conduce a la formulación de una ***Hipótesis Alternativa***. Generalmente, la Hipótesis Alternativa coincide con la hipótesis planteada por el investigador y la Hipótesis Nula es la negación de la Hipótesis Alternativa; de este modo, uno de los propósitos de todo experimento, es decidir si los datos obtenidos en la prueba de campo o laboratorio tienden a refutar o apoyar la Hipótesis Nula, aclarando que “rechazar una hipótesis significa, concluir que es falsa”, mientras que “aceptar una hipótesis significa, no tener suficiente evidencia para rechazarla”.

En la mayoría de los casos, es deseable que una Hipótesis Nula referida a un parámetro de la población, sea planteada, de tal manera que, se especifique un valor exacto del parámetro, mientras que la Hipótesis Alternativa permite la posibilidad de contrastar varios valores.

En el Ejemplo 6.1, la Hipótesis Nula sería:

H_0: El contenido promedio de grasa por litro es de 0,36 kg.
Se puede expresar de la siguiente forma: Ho: $\mu = 0{,}36$

La Hipótesis Alternativa podría ser una de las siguientes:

H_1: El contenido promedio de grasa por litro es mayor que 0,36 kg.
H_1: El contenido promedio de grasa por litro es menor que 0,36 kg.
H_1: El contenido promedio de grasa por litro es diferente de 0,36 kg.

Las Hipótesis Nula y Alternativa, podrían ser escritas así:

H_0: $\mu = 0{,}36$
H_1: $\mu > 0{,}36$

H_0: $\mu = 0{,}36$
H_1: $\mu < 0{,}36$

H_0: $\mu = 0{,}36$
H_1: $\mu \neq 0{,}36$

Para el Ejemplo 6.2, la Hipótesis Nula se plantea como:

H_0: La proporción de truchas curadas al recibir el tratamiento con yodo (*Py*), es igual a la proporción de truchas curadas al recibir un tratamiento con azul de metileno (*Pa*).

H_0: $Py = Pa$

De la misma forma que en el caso anterior, las Hipótesis Nula y Alternativa, se podrían plantear como:

H_0: $Py = Pa$
H_1: $Py > Pa$

H_0: $Py = Pa$
H_1: $Py < Pa$

H_0: $Py = Pa$
H_1: $Py \neq Pa$

Después de plantear las Hipótesis Nula y Alternativa y de recoger información muestral generada en el experimento, debe tomarse una decisión sobre la Hipótesis Nula. Las dos posibles decisiones respecto a Ho son aceptarla o rechazarla y, al tomar una de estas decisiones, se pueden cometer dos tipos de error.

El primero o ***Error Tipo I***, consiste en rechazar H_0, cuando en realidad ésta es verdadera. El otro tipo de error, denominado ***Error Tipo II***, consiste en aceptar H_0 cuando en realidad ésta es falsa.

En la Tabla 6.1 se puede apreciar los estados de la naturaleza y las decisiones sobre la hipótesis nula.

Tabla 6.1 Estados de la naturaleza y decisiones sobre la Hipótesis Nula

Decisiones sobre H_0	**Estados de la naturaleza**	
	H_0 es Verdadera	**H_0 es Falsa**
Aceptar H_0	Decisión Correcta	Error tipo II
Rechazar H_0	Error tipo I	Decisión Correcta

Como se mencionó anteriormente, la decisión de aceptar o rechazar una Hipótesis Estadística tiene como base los datos generados a partir de una muestra, es decir, es un proceso inferencial en el cual siempre existirán riesgos de tomar una decisión falsa, los que se expresan en términos probabilísticos. Así, la Probabilidad de rechazar Ho cuando H_0 es verdadera, se define como la ***Probabilidad del Error Tipo I*** y se denota por $\boldsymbol{\alpha}$, donde $0 \leq \alpha \leq 1$.

$$P[\textit{rechazar } H_0 | \ H_0 \textit{ es verdadera}] = \alpha$$

La Probabilidad de aceptar H_0, cuando H_0 es falsa, se define como la ***Probabilidad del Error Tipo II*** y se denota por $\boldsymbol{\beta}$, donde $0 \leq \beta \leq 1$.

$$P[\textit{aceptar } H_0 | \ H_0 \textit{ es falsa}] = \beta$$

El valor que toma α se denomina ***Nivel de Significancia*** y, entre más grande sea α, mayor será la Probabilidad de rechazar H_0 equivocadamente, es decir, existe mayor Probabilidad de cometer Error Tipo I.

Como puede entenderse, al verificar una Prueba de Hipótesis, es difícil establecer la gravedad de los Errores Tipo I y II, además, para un tamaño de muestra dado, una disminución del tamaño del Error Tipo I o α, implica un aumento en el tamaño del Error Tipo II o β. Hasta ahora el único

procedimiento eficaz para evitar incurrir en errores tipo α o β, consiste en realizar técnicamente el muestreo y utilizar el máximo tamaño de muestra.

En la práctica, comúnmente se fija el máximo tamaño del Error Tipo I a tolerar en una situación dada, a criterio del investigador, dependiendo de los alcances y de la gravedad de las recomendaciones inherentes al error que se pueda cometer; por ejemplo, si se trata de lanzar al mercado un nuevo producto que potencialmente afecte la salud de los humanos, el valor de α debe ser lo más pequeño posible. En la mayoría de los trabajos de investigación, incluidos los de salud y producción animal, α oscila entre 1% y 5%. Cuando se escoge un nivel de significancia α y se rechaza una Hipótesis Nula, se dice que los resultados de la muestra son significativos al $\alpha \times 100\%$.

En revistas técnicas y libros especializados, se utiliza el valor $\boldsymbol{p}$ para representar el mínimo nivel de significancia α que permite rechazar la Hipótesis Nula, con los datos analizados. Por ejemplo, si en un ensayo de fertilización del *Brachiaria decumbens* se reporta que, el rendimiento de materia verde por hectárea con dos niveles de fósforo no presentó diferencias significativas, se reportará como ($p > 0,05$); si por el contrario se observaron diferencias estadísticas significativas en las medias de los rendimientos, se reportará como ($p < 0,05$).

REGIÓN CRÍTICA O REGIÓN DE RECHAZO

Como se mencionó anteriormente, al plantear una hipótesis, el procedimiento mediante el cual se toma la decisión de aceptarla o rechazarla, es la Prueba o Contraste de Hipótesis; dicha decisión se basa en un estadístico de prueba, definido como una cantidad numérica calculada con la información obtenida de una muestra, teniendo en cuenta el parámetro sobre el cual se hace la inferencia y la distribución muestral del estadístico correspondiente.

La Hipótesis Nula se rechaza para ciertos valores del estadístico de prueba, tales valores constituyen lo que se conoce como ***Región Crítica*** o ***Región de Rechazo*** de la prueba; al complemento se lo denomina ***Zona de no Rechazo*** o ***Zona de Aceptación***.

Ejemplo 6.3: En el Ejemplo 6.1, el procesador de leche afirma que, su producto tiene un contenido promedio de grasa por litro de 0,36 kg.

Así se formula la Hipótesis Nula como:
H_0: $\mu = 0{,}36$.

Para un tamaño de muestra dado, supóngase que se decide rechazar H_0 si se obtiene un promedio muestral $\bar{y}$ menor que 0,34.

En este ejemplo, $\bar{y} = 0{,}34$ es el valor crítico y el conjunto de valores menores que 0,34 constituyen la región crítica o región de rechazo de la prueba.

Para mostrarlo de una forma gáfica, se supone que el contenido de grasa por litro se distribuye aproximadamente como una normal o que ***n*** es lo suficientemente grande, de tal manera que, la distribución de muestreo $\bar{y}$ se aproxime a una distribución normal.

En la Figura 6.1 se observa que , la región crítica es la parte sombreada a la izquierda del punto crítico; la región situada a la derecha de dicho punto, es la región de aceptación o no rechazo de Ho. El área de la región crítica, está dada por α que constituye, como se ha indicado, el nivel de significancia de la prueba, es decir,

$$P[\textit{rechazar Ho}|\textit{Ho es verdadera}] = \alpha \; o \; P[\bar{y} < 0{,}34|\mu = 0{,}36] = \alpha.$$

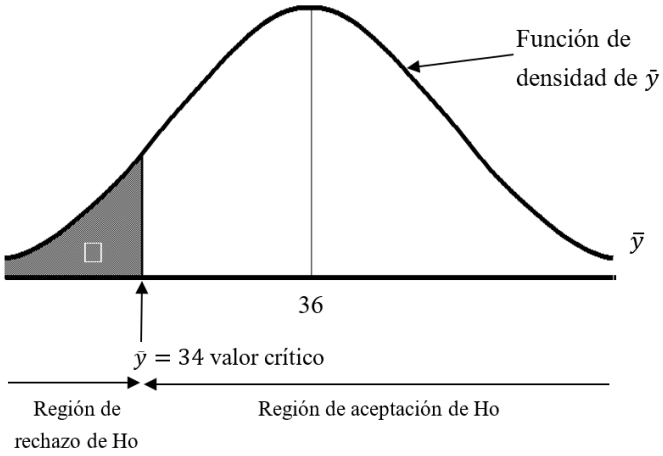


Figura 6.1 Regiones de aceptación y de rechazo de la Hipótesis Nula en una Distribución Normal.

Por ejemplo, si $\alpha = 0{,}05$, entonces el tamaño de la región crítica es el 5,0% del área total comprendida bajo la curva normal.

$P[\bar{y} < 0{,}34|\mu = 0{,}36] = 0{,}05$ significa que hay aproximadamente cinco ocasiones en 100 en que se rechazaría la hipótesis, cuando en realidad debería ser aceptada; es decir, la confiabilidad es del 95% para tomar una decisión adecuada. En este caso, se dice que, la hipótesis ha sido rechazada al nivel de significancia del 5,0%.

HIPÓTESIS UNILATERALES Y BILATERALES

En algunas pruebas de hipótesis se involucran ***Hipótesis Alternativas Unilaterales***, también conocidas como Pruebas de Hipótesis ***A Una Cola***, tal como se indican a continuación:

H_0: $\theta = \theta_0$ o H_0: $\theta = \theta_0$

H_1: $\theta > \theta_0$ H_1: $\theta < \theta_0$

Se dice que H_1 es una ***Hipótesis Alternativa Unilateral***, debido a que los posibles valores de θ bajo H_1 se encuentran a un solo lado del valor propuesto bajo Ho.

Si H_1: $\theta > \theta_0$, la región crítica se encuentra en la cola superior de la distribución del estadístico de prueba, mientras que si H_1: $\theta < \theta$o, la región crítica se encuentra en la cola inferior de la distribución del estadístico de prueba. En cierto sentido, el símbolo de desigualdad señala la dirección en la que se ubica la región crítica.

A una prueba de la forma H_0: $\theta = \theta_0$, H_1: $\theta \neq \theta_0$ se la denomina ***Prueba Bilateral*** o ***De Dos Colas***, puesto que la región crítica está dividida en dos partes iguales ubicadas en cada cola de la distribución del estadístico de prueba. La hipótesis alterna H_1: $\theta \neq \theta_0$ implica que pueden presentarse dos situaciones, ya sea que $\theta > \theta_0$ o que $\theta < \theta_0$.

Como se explicó antes, al realizar un procedimiento para probar hipótesis, la Hipótesis Nula, siempre se plantea como una igualdad, de tal manera que, se especifica un solo valor para poder controlar la Probabilidad de Error Tipo I, mientras que la Hipótesis Alternativa puede ser Unilateral o Bilateral, dependiendo del interés del investigador.

Resumiendo, una prueba o contraste de hipótesis consta, esquemáticamente, de los siguientes pasos:

a. Identificación del parámetro de interés.
b. Planteamiento de la Hipótesis Nula y la Alternativa (H_0 y H_1).

c. Elección del nivel de significancia.
d. Definición del estadístico de prueba.
e. Establecimiento de las regiones de aceptación y de rechazo para el estadístico.
f. Cálculo del valor del estadístico de prueba a partir de una muestra aleatoria.
g. Decisión sobre el rechazo o aceptación de la hipótesis, dependiendo si el estadístico de prueba se ubica en la región de rechazo o en la de aceptación.
h. Interpretación y conclusión.

CONTRASTES RELACIONADOS CON LA MEDIA μ

El objetivo fundamental de estas pruebas, consiste en establecer, si una media poblacional μ es igual o diferente de cierto valor supuesto μ_0, para lo cual se contrasta una de las siguientes hipótesis respecto al parámetro μ:

H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$
H_1: $\mu \neq \mu_0$	H_1: $\mu > \mu_0$	H_1: $\mu < \mu_0$

Los contrastes con respecto a μ varían en su procedimiento, dependiendo fundamentalmente del conocimiento de σ^2, en consecuencia, se pueden presentar las situaciones que se describen a continuación.

Contraste de Hipótesis para la Media Poblacional μ con Varianza Poblacional σ^2 conocida

Sea $\bar{y}$ la Media de una muestra aleatoria de tamaño $\boldsymbol{n}$, procedente de una población normal, con Media μ y Varianza σ^2 conocida. En este caso se pueden contrastar las siguientes hipótesis:

Ho: $\mu = \mu_0$

H_1: $\mu \neq \mu_0$

donde μ_0 es un valor específico.

El contraste sobre la Media Poblacional tiene como estadístico de prueba la Media Muestral $\bar{y}$ que, por lo general, se estandariza para utilizar un estadístico de prueba basado en la Distribución Normal Estándar. Bajo H_0: $\mu = \mu_0$, la Media Muestral $\bar{y}$ tiene una Distribución Normal con Media μ_0 y Desviación Estándar $\frac{\sigma}{\sqrt{n}}$. Esto es, la variable aleatoria

$$Z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} \qquad (6.1)$$

donde (6.1) corresponde a una variable normal estándar.

La hipótesis nula H_0: $\mu = \mu_0$ se rechaza si la Media Muestral $\bar{y}$ es mayor o menor que el valor postulado μ_0, para la Media Poblacional μ. Por lo tanto, H_0 se rechaza si se observa un valor alto o un valor bajo para el estadístico de prueba (6.1).

Si α es la probabilidad de rechazar H_0, cuando H_0 es verdadera, la probabilidad de que Z esté ubicado en la región $Z > Z_{1-\alpha/2}$ o $Z < Z_{\alpha/2}$ es α.

Por lo tanto, H_0 se rechaza si: $Z > Z_{1-\alpha/2}$ o $Z < Z_{\alpha/2}$, donde $Z_{\alpha/2}$ y $Z_{1-\alpha/2}$ corresponden a los valores cuantiles de la normal estándar (Figura 6.2).

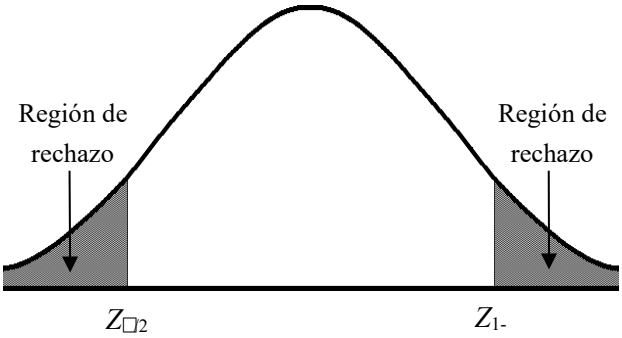


Figura 6.2 Regiones de rechazo de la normal estándar en un contraste de hipótesis para la media.

Ejemplo 6.4 Un grupo investigador asegura que, mediante cruzamientos, ha logrado obtener un híbrido de conejo, cuyo rendimiento promedio en canal es de 2,28 kg. Por experiencia se sabe que, $\sigma = 0{,}65$ kg. Un productor está interesado en adquirir ejemplares del mencionado híbrido; para ello, solicita a un profesional pecuario, un estudio que confirme si efectivamente se alcanza el rendimiento indicado. El profesional toma una muestra aleatoria de 25 animales y calcula un rendimiento promedio de 2,02 kg. Con base en esta información y utilizando un nivel de significancia de $\alpha = 0{,}05$, ¿Se puede afirmar que el rendimiento promedio es el indicado por los investigadores?

Siguiendo el esquema planteado anteriormente, para resolver este interrogante se procede del siguiente modo:

a. El parámetro de interés es la Media Poblacional μ.

b. Las Hipótesis Nula y Alternativa son

H_0: $\mu = 2{,}28$

H_1: $\mu \neq 2{,}28$.

En este caso, la Hipótesis Alterna es bilateral, ya que se desea rechazar la H_0 si la Media es significativamente mayor o menor que 2,28.

c. Se define $\alpha = 0{,}05$.

d. El estadístico de prueba es $\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}} = Z$.

e. Ho se rechaza si: $Z > Z_{1-\alpha/2} = Z_{0.975} = 1{,}96$ o

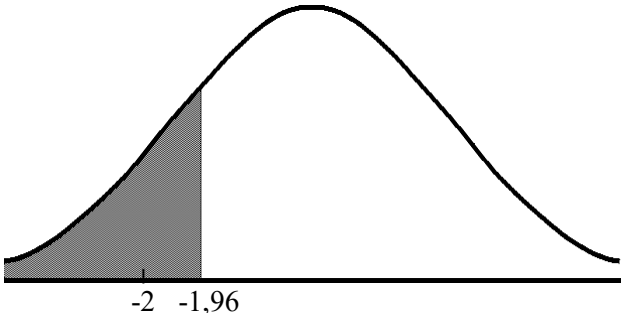
$$Z < Z_{\alpha/2} = Z_{0{,}025} = -1{,}96$$

f. El valor del estadístico de prueba es: $\frac{2{,}02-2{,}28}{0{,}65/\sqrt{25}} = -2$

g. Decisión: Dado que $Z = -2 < -1{,}96$, es decir, $Z = -2$ se ubica en la zona de rechazo, por lo tanto, H_0: $\mu = 2{,}28$ debe rechazarse.

Para informar sobre los resultados de la prueba, puede recurrirse a una explicación planteada de la siguiente manera: *con un nivel de significancia del* 5% *se asegura que, el rendimiento promedio en canal de los híbridos de conejo es diferente de* 2,28 *kg.*

En realidad, los datos proporcionan evidencia suficiente, para asegurar que existe un rendimiento promedio menor de 2,28 kg, tal como se puede apreciar en el siguiente esquema demostrativo, donde se observa que el valor crítico cae en la región donde se encuentran valores negativos:



Por otra parte, es necesario recordar que el valor ***p*** es el mínimo nivel de significancia α, para el cual los datos observados indican la necesidad de rechazar la Hipótesis Nula. En este caso, el mínimo valor α con el cual se rechaza la hipótesis H_0: μ = 2,28 kg, viene dado por la Probabilidad de encontrar un Z < -2, es decir, $P[Z < -2] = 0{,}0228$; como el contraste es a dos colas $\alpha/2 = 0{,}0228$, por lo tanto, $\alpha = 0{,}0456$. Esto significa que, si la Hipótesis Nula fuera cierta, la Probabilidad de encontrar una Media Muestral alejada de 2,28 kg es 0,0456.

Relación entre Contraste de Hipótesis e Intervalo de Confianza

Para las hipótesis H_0: $\mu = \mu_0$ y H_1: $\mu \neq \mu_0$, los puntos críticos son (Figura 6.3):

$$a = \mu_0 - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{y}$$

$$b = \mu_o + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Ho se rechaza si $\bar{y} < a$ o $\bar{y} > b$; es decir, H_0: $\mu = \mu_0$ frente a la alternativa H_1: $\mu \neq \mu_0$ se rechaza si en el intervalo de confianza para μ al (1-α)100% de confiabilidad no contiene el valor μ_o.

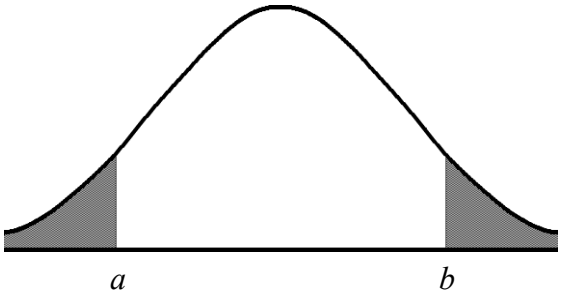


Figura 6.3 Puntos Críticos en un Intervalo de Confianza.

Considerando el contraste H_0: $\mu = \mu_0$, H_1: $\mu > \mu_0$, la hipótesis nula H_0: $\mu = \mu_0$ se rechaza si la media muestral $\bar{Y}$ es mayor que el valor postulado μ_o para la media poblacional μ. Por lo tanto, H_0 se rechaza si se observa un valor mayor para el estadístico de prueba

$$Z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}$$

Es decir, si α es el nivel de significancia, H_0 se rechaza si $Z > Z_{1-\alpha}$ (Figura 6.4).

Donde $Z_{1-\alpha}$ es tal que $P[Z > Z_{1-\alpha}] = \alpha$.

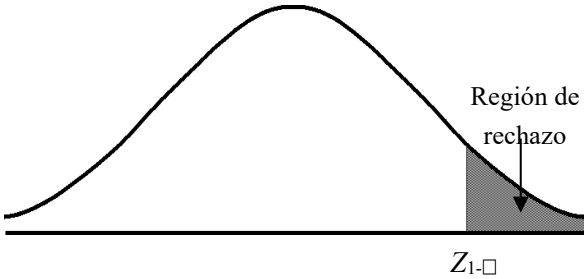
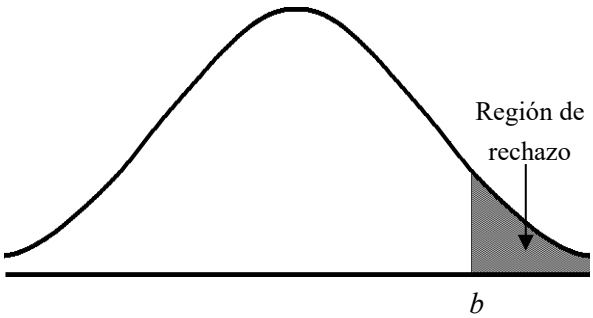


Figura 6.4 Región de Rechazo para H_1: $\mu > \mu_0$.

Para la hipótesis planteada, el punto crítico es $b = \mu_o + Z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$ y, por lo tanto, H_0 se rechaza si la Media Muestral $\bar{y} > \mu_o + Z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$, como se muestra en el siguiente esquema, equivalente a la Figura 6.4.



De la misma forma, para contrastar H_0: $\mu = \mu_0$, H_1: $\mu < \mu_0$

H_0 se rechaza si se observa un valor menor para el estadístico de prueba

$$Z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}$$

Si α es el nivel de significancia, H_0 se rechaza si $Z < Z_\alpha$.

Donde $P[Z < Z_\alpha] = \alpha$.

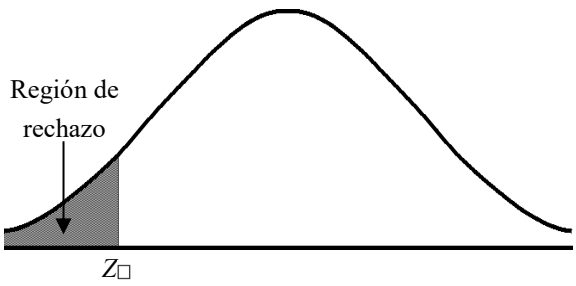


Figura 6.5 Región de rechazo para H_1: $\mu < \mu_0$.

De manera equivalente H_0 se rechaza si $\bar{y} < \mu_o - Z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$

Ejemplo 6.5: En el Ejemplo 6.4 se dijo que existe una fuerte evidencia de que, el rendimiento promedio en canal de los híbridos de conejo es menor que 2,28 kg, dado que el valor crítico se ubicó en la región de la curva con valores negativos.

Para afianzar esa afirmación se debe contrastar la hipótesis:

Ho: μ = 2,28
H_1: μ < 2,28.

Siguiendo el esquema planteado anteriormente, se tiene:

a. El parámetro de interés es la Media Poblacional μ.

b. Las hipótesis son:

   H_0: μ = 2,28
   H_1: μ < 2,28

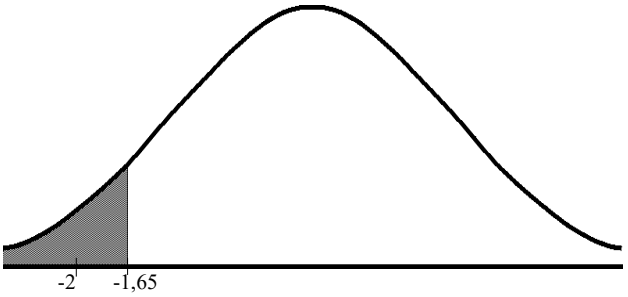
c. α = 0,05

d. El estadístico de prueba es: $Z = \frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}}$.

e. Ho se rechaza si $Z < Z_\alpha = Z_{0,05}$ = -1,65.

f. El valor de este estadístico de prueba es: $\frac{2,02-2,28}{0,65/\sqrt{25}} = -2$

g. Dado que Z = -2 < $Z_{0,05}$ = -1,65, como se muestra en el siguiente esquema, H_0 se rechaza, es decir, con un nivel de significancia de 5%, se concluye que, el rendimiento promedio en canal de los conejos híbridos es menor de 2,28 kg.



Es importante recordar que, por lo general, cuando se hacen pruebas de hipótesis relacionadas con variables productivas, se asume que la variable sigue una Distribución Normal y, si bien, esta afirmación es confirmada por múltiples trabajos de investigación, es conveniente analizar los datos de cada caso particular, para confirmar la normalidad de las variables analizadas. En caso de encontrarse una desviación marcada de la normalidad en los datos, los procedimientos indicados para las Pruebas de Hipótesis, no serían los más adecuados.

Contraste de Hipótesis para la Media Poblacional μ, con Muestras grandes ($n \geq 30$) y Varianza σ^2 desconocida

En la mayoría de las situaciones prácticas no se conoce σ^2, pero por lo general, si el tamaño de muestra es grande ($n \geq 30$), el Teorema Central del Límite, permite concluir que, la Distribución de la Media Muestral es aproximadamente normal y, por lo tanto, se pueden utilizar los mismos procedimientos de contraste desarrollados para el caso en que σ^2 es conocida, para lo cual, se puede sustituir σ^2 por S^2.

Ejemplo 6.6: Un investigador asegura que, un rebaño selecto de vacas Holstein, cuyo promedio de producción es 34 L/día, podría llegar a producir 41 L/día siempre y cuando se incorpore en la dieta proteínas de origen animal y vitaminas liposolubles, en las

cantidades que él sugiere. Con una muestra aleatoria de 50 vacas, sometidas a una dieta como la indicada, se calculó un promedio de producción de 39 L, con una Desviación Estándar de seis litros. Se pregunta: ¿Existen razones suficientes para creer que el promedio de producción, es menor de 41 L?.

Esta pregunta se puede responder, desarrollando el procedimiento para Prueba de Hipótesis. Como σ^2 es desconocida y el tamaño de muestra es mayor de 30, se aplica el mismo proceso general, así:

a. El parámetro de interés es la Media μ.

b. H_0: $\mu = 41$
   H_1: $\mu < 41$.

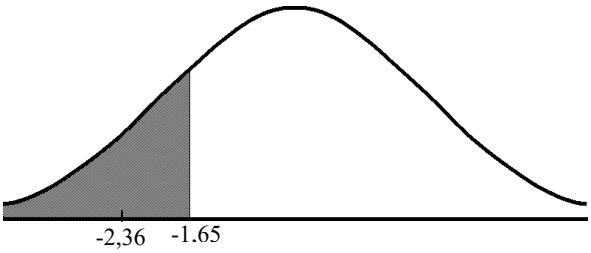
c. $\alpha = 0{,}05$.

d. El estadístico de prueba es $\frac{\bar{y}-\mu_o}{\sigma/\sqrt{n}} = Z$, donde $\sigma \cong S$.

e. Ho se rechaza si $Z < Z_\alpha = Z_{0,05} = -1{,}65$.

f. El valor del estadístico de prueba es: $\frac{39-41}{6/\sqrt{50}} = -2{,}36$.

g. Decisión: Dado que $Z = -2{,}36 < -1{,}65$, como se muestra en el esquema siguiente:



> H_0 se rechaza y, por lo tanto, se concluye que, el promedio de producción de las vacas que fueron sometidas a dicha dieta es menor que 41 L, afirmación que se hace con un nivel de significancia del 5%.

Contraste de Hipótesis para la Media Poblacional μ con Muestras pequeñas ($n < 30$) y σ^2 desconocida

Sean, $\bar{y}$ la Media y S la Desviación Estándar de una muestra aleatoria de tamaño n, procedente de una población normal, con Media μ y Varianza σ^2 desconocida, el contraste de hipótesis bilateral:

H_0: $\mu = \mu_o$
H_1: $\mu \neq \mu_o$

se basa en el estadístico de prueba:

$$T = \frac{\bar{y} - \mu_0}{S/\sqrt{n}} \qquad (6.2)$$

El cual tiene una distribución t con n-1 grados de libertad. Utilizando el mismo procedimiento general de Prueba de Hipótesis anteriormente indicado, H_0 se rechaza si: $T > t_{1-\alpha/2,\, n-1}$ o $T < t_{\alpha/2,\, n-1}$

Para el contraste de hipótesis:

H_0: $\mu = \mu_0$
H_1: $\mu > \mu_0$
H_0 se rechaza si $T > t_{1-\alpha,\, n-1}$

Para el contraste de hipótesis:

H_0: $\mu = \mu_0$

H_1: $\mu < \mu_0$

H_0 se rechaza si $T < t_{\alpha,\ n\text{-}1}$

Ejemplo 6.7: Una compañía dedicada a la comercialización de camarón congelado imprime las etiquetas de las bandejas "contiene 12 oz". Con una muestra de 22 bandejas se calculó una Media de $\bar{y} = 11{,}84$ onzas y una Desviación Estándar (S) de 0,5 onzas. Suponiendo que, los pesos de las bandejas siguen en forma adecuada una Distribución Normal, con un nivel de significancia del 5,0%, ¿Aseguraría que la afirmación del productor es correcta?

Al igual que en los ejemplos anteriores, debe seguirse el procedimiento general de Prueba de Hipótesis:

a. El parámetro de interés es la Media Poblacional μ, en este caso, el peso promedio de la población de bandejas.

b. Se plantea la hipótesis:

   H_0: $\mu = 12$
   H_1: $\mu < 12$ (a una cola).

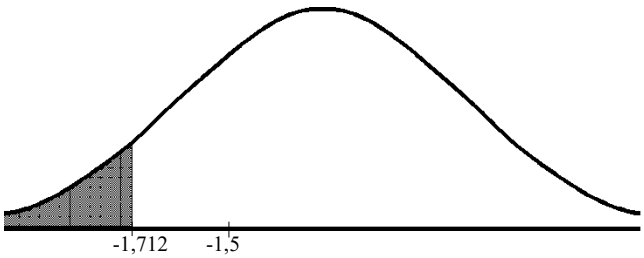
c. $\alpha = 0{,}05$.

d. El estadístico de prueba es: $\frac{\bar{y}-\mu_0}{S/\sqrt{n}} = t$.

e. H_0 se rechaza si $T < t_{\alpha\,;\ n\text{-}1} = t_{0{,}95,\ 21} = -1{,}712$.

f. El valor del estadístico de prueba: $\frac{11{,}84-12}{0{,}5/\sqrt{22}} = -1{,}5$.

g. Decisión, interpretación y conclusión: Dado que $T = -1,5 > t_{0,05,\ 21} = -1,712$, es decir, $t = -1,712$ cae en zona de aceptación, como se muestra en el esquema siguiente, en cuyo caso, no existen elementos suficientes para rechazar Ho, por lo tanto, $\mu = 12$ onzas, y la afirmación sobre el contenido de los envases (12 onzas) es correcta.



CONTRASTES RELACIONADOS CON LA DIFERENCIA DE MEDIAS

El objetivo fundamental de estos contrastes, consiste en establecer si existe diferencia entre dos Medias μ_1 y μ_2 de dos poblaciones independientes, para ello se recurre a uno de los siguientes contrastes:

$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$	$H_1: \mu_1 > \mu_2$	$H_1: \mu_1 < \mu_2$

A veces, es de interés analizar si la diferencia $\mu_1 - \mu_2$ es igual, mayor o menor que un número real cualquiera δ, es decir, se analizan los siguientes contrastes:

$H_0: \mu_1 - \mu_2 = \delta$	$H_0: \mu_1 - \mu_2 = \delta$	$H_0: \mu_1 - \mu_2 = \delta$
$H_0: \mu_1 - \mu_2 \neq \delta$	$H_1: \mu_1 - \mu_2 > \delta$	$H_1: \mu_1 - \mu_2 < \delta$

Contraste de Hipótesis para la Diferencia de Medias con Varianzas σ_1^2 y σ_2^2 conocidas

Sean dos muestras aleatorias con tamaños n_1 y n_2 grandes, con Medias Muestrales $\bar{y}_1$ y $\bar{y}_2$, respectivamente, procedentes de dos poblaciones normales e independientes, con Medias μ_1 y μ_2 y Varianzas σ_1^2 y σ_2^2 conocidas. En casos como este, la mayoría de veces, el investigador pretende establecer la igualdad o diferencia de las Medias, para lo cual se recurre a realizar la correspondiente Prueba de Hipótesis que, se basa en el siguiente estadístico de prueba:

$$Z = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad (6.3)$$

En (6.3) Z tiene una Distribución Normal Estándar y los contrastes de hipótesis se apoyan en este resultado.

En el contraste bilateral:

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

el estadística de prueba indicado en (6.3) se puede expresar del siguiente modo:

$$Z = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad (6.4)$$

La regla de decisión consiste en rechazar H_0 si $Z > z_{1-\alpha/2}$ *o* $Z < z_{\alpha/2}$.

Para contrastar la Hipótesis Alternativa Unilateral:

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 > \mu_2$

la regla de decisión consiste en rechazar H_0 cuando $Z > z_{1-\alpha}$

Para la otra Hipótesis Unilateral:

H_0: $\mu_1 = \mu_2$
Ho: $\mu_1 < \mu_2$

H_1 se rechaza cuando $Z < Z_{\alpha}$

Es muy común, en Producción y Salud Animal, el uso de la Prueba de Hipótesis, para establecer la igualdad de dos Medias, ya que, por motivos de costos, sólo se pueden comparar dos tratamientos o probar la efectividad de dos sistemas de producción y muchos casos similares.

Ejemplo 6.8: En los peces de la tribu Tilapinae, se recurre a la técnica de la triploidía inducida, cuyo objetivo consiste en producir animales con un juego cromosómico adicional que, en muchas ocasiones, presentan un rendimiento medio superior a los diploides normales y como ventaja adicional no se reproducen, lo que evita que los ejemplares utilicen parte de la energía en este proceso. En un experimento donde se conocían previamente las Varianzas, se comparó el peso final de los diploides normales y los triploides, encontrándose los siguientes resultados:

Diploides: $n_1 = 60$, $\bar{y}_1 = 0{,}320$ kg, $\sigma_1^2 = 0{,}86$ kg^2.
Triploides: $n_2 = 60$, $\bar{y}_2 = 0{,}360$ kg, $\sigma_2^2 = 0{,}94$ kg^2.

Con base en los anteriores datos, con $\alpha = 0{,}05$, se debe decidir si las Medias de los dos grupos son iguales o diferentes.

a. El parámetro de interés es la diferencia de Medias $\mu_1 - \mu_2$.

b. Las hipótesis son:

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

c. $\alpha = 0{,}05$.

d. El estadístico de prueba es: $Z = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

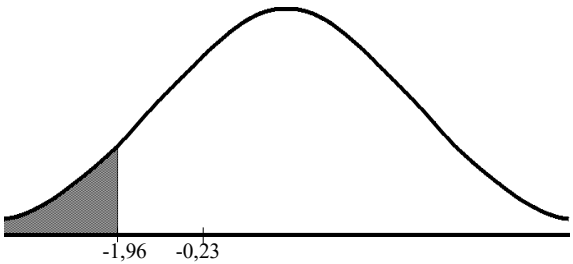
e. Ho se rechaza si: $Z > z_{1-\alpha/2} = z_{0{,}975} = 1{,}96$ o

$Z < z_{\alpha/2} = z_{0{,}025} = -1{,}96$

f. El valor del estadístico de prueba es:

$$\frac{0{,}320 - 0{,}360}{\sqrt{\frac{0{,}86}{60} + \frac{0{,}94}{60}}} = \frac{-0{,}04}{\sqrt{0{,}03}} = \frac{-0{,}04}{0{,}1732} = -0{,}23$$

g. Decisión, interpretación y conclusión: Puesto que $z = -0{,}23 > z_{\alpha/2} = -1{,}96$, representado en el siguiente esquema:



Por lo tanto, no existen elementos suficientes para rechazar H_0; en consecuencia, con un nivel de significancia del 5%, se asegura que las medias son iguales.

Contraste de Hipótesis para la Diferencia de Medias con Muestras grandes ($n_1, n_2 \geq 30$) y Varianzas σ_1^2 y σ_2^2 desconocidas

Si los tamaños de muestras n_1 y n_2 son grandes, el Teorema Central del Límite garantiza que la distribución de la diferencia de Medias Muestrales es aproximadamente normal y, en consecuencia, se puede utilizar el procedimiento empleado cuando σ_1^2 y σ_2^2 se conocen, reemplazando S_1^2 y S_2^2 por σ_1^2 y σ_2^2.

Ejemplo 6.8: En las razas bovinas para leche, luego de muchos años de investigación, se aprobó la aplicación de hormona del crecimiento (GH) como factor importante para el incremento de la producción láctea. En un experimento se comparó, la producción promedio de dos hatos de vacas Guernesey de segundo parto, uno que no recibió ningún tratamiento y otro tratado con GH. Al final de la lactancia se encontraron los siguientes resultados:

- Grupo Control (sin tratamiento): $n_1 = 43$, $\bar{y}_1 = 3.200$ kg, $S_1^2 = 126.000$ kg^2.
- Grupo experimental (con tratamiento): $n_1 = 48$, $\bar{y}_2 = 3.260$ kg, $S_2^2 = 109.000$ kg^2.

Con base en los anteriores datos muestrales, debe decidirse si se justifica el uso de GH, e igualmente debe recurrirse a una prueba de hipótesis, con $\alpha = 0{,}05$.

a. El parámetro de interés es $\mu_1 - \mu_2$.
b. H_0: $\mu_1 = \mu_2$ (La producción promedio de los dos grupos no presenta diferencias significativas).

H_1: $\mu_1 < \mu_2$ (La producción promedio sin tratamiento es menor que la producción promedio con tratamiento).

c. $\alpha = 0{,}05$.

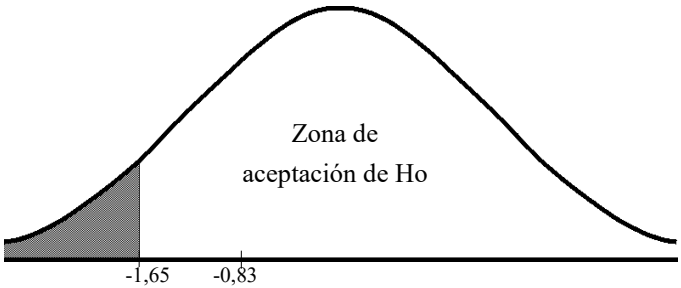
d. El estadístico de prueba es: $Z = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

e. H_0 se rechaza si $Z < z_\alpha$ donde $z_{0,05} = -1{,}65$

f. El valor del estadístico de prueba es

$$Z = \frac{3.200 - 3.260}{\sqrt{\frac{126.000}{43} + \frac{109.00}{48}}} = \frac{-60}{\sqrt{5.201}} = \frac{-60}{72{,}11} = -0{,}83$$

g. Decisión, interpretación y conclusión: Dado que (Z = -0,83) > (Z_α = -1,65)



Se concluye: no existen elementos suficientes para rechazar Ho, con el 95% de confiabilidad; se puede decir, la producción promedio de los dos grupos, no presenta diferencias significativas, lo que conduce a no recomendar la aplicación de la hormona del crecimiento (GH).

Contraste de Hipótesis para la Diferencia de Medias con Muestras pequeñas (n_1, $n_2 < 30$) y Varianzas desconocidas pero iguales ($\sigma_1^2 = \sigma_2^2$)

Dadas dos muestras aleatorias con tamaños n_1, $n_2 < 30$ y medias maestrales $\bar{y}_1$ y $\bar{y}_2$, respectivamente, procedentes de dos poblaciones normales e independientes con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2 desconocidas pero que se suponen iguales $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Si las Varianzas muestrales son S_1^2 y S_2^2, un estimador de la ***Varianza Común*** σ^2 es:

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \qquad (6.5)$$

Bajo estas condiciones, tal como se vio en el capítulo anterior, la variable aleatoria

$$T = \frac{(\bar{y}_1 - \bar{y}_2)-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_1}}} \qquad (6.6)$$

tiene una distribución ***t*** con $n_1 + n_2$ - 2 grados de libertad. Este es el estadístico de prueba, en el cual se apoya el contraste de hipótesis.

El contraste bilateral
H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

Tiene como estadístico de prueba

$$\frac{\bar{Y}_1-\bar{Y}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = T \qquad (6.7)$$

y la regla de decisión es rechazar Ho si:

$T > t_{1-\alpha/2;\ n1+n2-2}$ ó
$T < t_{\alpha/2;\ n1+n2-2}$

Donde $t_{1-\alpha/2}$ es el valor de ***t*** con n_1+n_2-2 grados de libertad y $P[T \leq t_{1-\alpha/2}] = 1 - \alpha/2$.

Para el contraste unilateral:

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 > \mu_2$

La regla de decisión es rechazar Ho si: $T > t_{1-\alpha;\ n1+n2-2}$

Para el otro contraste unilateral:

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 < \mu_2$

La regla de decisión es rechazar Ho si: $T < t_{\alpha;\ n1+n2-2}$

Ejemplo 6.9: En el *Prochilodus magdalenae* (bocachico), como en todas las especies ícticas, el diámetro de los huevos es un indicador de mucha importancia, ya que de él depende en buena medida el tamaño de las larvas y, por consiguiente, su capacidad de adaptación al medio, mejorando así la supervivencia. La utilización de hormonas estimulantes de la maduración gonadal, además de producir mayor cantidad de ovocitos, incrementan el tamaño de los huevos, por esta razón, se planteó un experimento consistente en probar el efecto del extracto pituitaria de carpa (*Cyprinus carpio*) (EPC) y la hormona coriónica humana (HCG)

en el tamaño de los huevos de la especie en mención, midiendo el diámetro en mm. Los resultados del citado experimento fueron los siguientes:

EPC	**HCG**
1,5	1,3
1,2	1,0
1,8	1,6
1,1	0,9
1,4	0,7
1,0	0,9
1,6	1,4
1,6	1,4
1,6	1,4
1,5	1,5
1,8	1,5
1,9	1,7
1,6	1,4
1,6	1,4
1,7	1,5

Con base en los datos anteriores, se pide decidir cuál de los dos tratamientos es más recomendable para incrementar el tamaño de las ovas.

El procedimiento para realizar una recomendación técnicamente formulada, consiste en efectuar una prueba de hipótesis.

a. El parámetro de interés es μEPC - μHCG.

b. H_0: μEPC = μHCG
H_1: μEPC ≠ μHCG

Esto indica que, H_0 asegura que el diámetro medio de las ovas es igual en los dos tratamientos, mientras que en H_1 se plantea que el diámetro medio es diferente.

c. $\alpha = 0{,}05$

d. Se asume que, las Varianzas son iguales (efectivamente lo son), así que el estadístico de prueba es:

$$\frac{\bar{Y}_1 - \bar{Y}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = T \qquad \text{Donde } S_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

e. Ho se rechaza si:

$T > t_{1-\alpha/2,\, n1+n2-2} = t_{0{,}975,\, 28} = 2{,}05$ o
$T < t_{\alpha/2,\, n1+n2-2} = t_{0{,}025,\, 28} = -2{,}05$

f. Para calcular el estadístico de prueba, en primer lugar, se calculan los estadígrafos más importantes, así:

	EPC	**HCG**
n	15	15
Media	1,5260	1,3060
Varianza	0,0664	0,0849
Desviación estándar	0,2576	0,2915

$$S_p = \sqrt{\frac{(15-1)0{,}0664 + (15-1)0{,}0849}{28}} = \sqrt{\frac{1{,}1182}{28}}$$

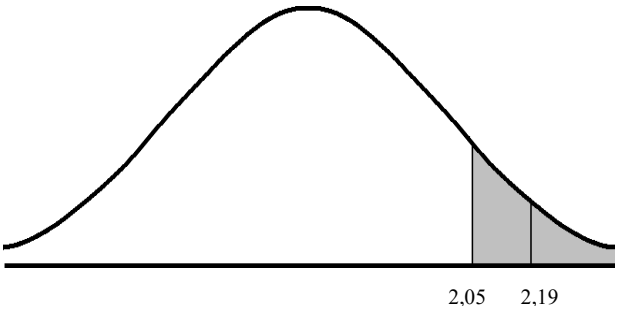
$$= \sqrt{0{,}7565} = 0{,}2750$$

$$\sqrt{\frac{1}{n1} + \frac{1}{n2}} = \sqrt{\frac{1}{15} + \frac{1}{15}} = 0{,}3651$$

El estadístico de prueba es:

$$T = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1{,}526 - 1{,}306}{0{,}2750 \times 0{,}3651} = 2{,}19$$

g. Dado que $T = 2{,}19 > t_{0{,}975,\ 28} = 2{,}05$ como se indica en el siguiente esquema.



Teniendo en cuenta que el valor de T (1,19) es mayor que el valor de $t_{0{,}95,\ 28}$ (2,05), H_0 se rechaza y se concluye que, existen diferencias significativas entre las medias de las dos poblaciones, razón por la cual, se puede decir que hay suficiente evidencia para recomendar el extracto pituitario de carpa (EPC) como tratamiento para incrementar el tamaño de las ovas.

Ejemplo 6.10: Algunos autores aseguran que, el selenio es uno de los principales oligoelementos para una adecuada producción de la avena forrajera (*Avena sativa*). En un experimento se evaluó el efecto de la adición de este mineral en los programas de fertilización midiendo la producción de materia seca por hectárea (Tn/Ha) con el objetivo de recomendar su adición en las prácticas culturales. Los resultados, al finalizar del trabajo de investigación, fueron los siguientes:

Con selenio: $n_1 = 9$, $\bar{y}_1 = 9{,}69$ Tn, $S_1^2 = 0{,}336$ Tn^2
Sin selenio: $n_2 = 10$, $\bar{y}_2 = 7{,}33$ Tn, $S_2^2 = 1{,}073$ Tn^2

Con base en estos resultados, ¿Se debe recomendar la incorporación del selenio en los programas de fertilización?

El procedimiento para resolver esta pregunta, mediante una prueba de hipótesis, es el siguiente:

b. H_0: μ con selenio = μ sin selenio
   H_1: μ con selenio > μ sin selenio.

c. $\alpha = 0{,}05$.

d. Si se supone que las Varianzas son iguales, el estadístico de prueba es como en (6.6):

$$T = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

e. H_0 se rechaza si $T > t_{1-\alpha,\, n1+n2-2} = t_{0{,}95,\, 17} = 1{,}74$

f. Se calcula el estadístico de prueba:

$$S_p = \sqrt{\frac{(9-1)0{,}336 + (10-1)1{,}073}{17}} = \sqrt{0{,}7262}$$

$$= 0{,}8522$$

$$\sqrt{\frac{1}{n1} + \frac{1}{n2}} = \sqrt{\frac{1}{9} + \frac{1}{10}} = 0{,}4595$$

$$T = \frac{9{,}69 - 7{,}33}{0{,}8522 \times 0{,}4595} = 6{,}03$$

g. Dado que $T = 6{,}03 > t_{0{,}95;\,17} = 1{,}74$, H_0 se rechaza, por lo tanto, se puede concluir, con el 95% de confiabilidad, que el rendimiento medio de materia seca por hectárea, cuando se utiliza selenio, es significativamente mayor que cuando no se utiliza.

Cuando las Varianzas σ_1^2 y σ_2^2 son significativamente diferentes, el estadístico de prueba se calcula de la siguiente manera:

$$T' = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad (6.8)$$

Se utiliza la fórmula (6.8) para indicar que, el estadístico de prueba no se distribuye exactamente como una ***t*** de Student, sino en una forma suficientemente aproximada a ésta, con υ grados de libertad, donde υ se calcula como:

$$\upsilon = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1+1}+\frac{\left(s_2^2/n_2\right)^2}{n_2+1}} \qquad (6.9)$$

Ejemplo 6.11: Los carbohidratos, en la dieta, son la fuente más importante de energía y de los principales precursores de grasa y lactosa en la leche de las vacas. Un experimento, tuvo como objetivo, evaluar dos dietas con diferentes fuentes de carbohidratos, una en la que se utilizó trigo y otra con maíz, con el propósito de recomendar la que produjera mayor contenido graso en la leche (kg/L) de vacas Jersey de segundo parto. Los resultados al final de experimento fueron los siguientes:

Dieta con trigo: $n_1 = 12$, $\bar{y}_1 = 0{,}31$ kg, $S_1^2 = 0{,}001454$ kg^2.
Dieta con maíz: $n_2 = 12$, $\bar{y}_2 = 0{,}26$ kg, $S_2^2 = 0{,}011509$ kg^2.

Con estos valores, se debe decidir, cuál de las dos fuentes de carbohidratos es la más recomendable, para lo cual se realiza una Prueba de Hipótesis, asumiendo que las Varianzas son diferentes.

De acuerdo con los datos, parece que la dieta con trigo produce un promedio mayor en el contenido graso de la leche, entonces contrastamos esta hipótesis.

a. $H_0: \mu_1 = \mu_2$

   $H_1: \mu_1 > \mu_2$ (La dieta con trigo produce mayor contenido graso en la leche).

b. $\alpha = 0{,}05$

c. Si las Varianzas son diferentes, el estadístico de prueba es como en (6.4.6):

$$T' = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

d. Ho se rechaza si $T' > t_{1-\alpha,\, \upsilon}$

e. Se calcula el estadístico de prueba con (6.4.6):

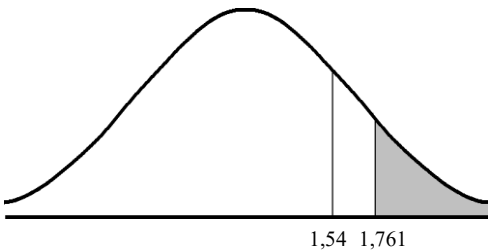
$$T' = \frac{0{,}31 - 0{,}26}{\sqrt{\frac{0{,}001454}{12} + \frac{0{,}011509}{12}}} = 1{,}5213$$

Se calculan los grados de libertad con (6.9):

$$\upsilon = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1+1} + \frac{\left(s_2^2/n_2\right)^2}{n_2+1}} = \frac{(0{,}0010802)^2}{0{,}000000084956} = 13{,}74 \approx 14$$

Entonces se encuentra el valor ***t*** de la tabla, cuyos valores se representan en el siguiente esquema

$t_{1-\alpha;\, \upsilon} = t_{0{,}95;\, 14} = 1{,}761$



f. De aquí se deduce que $T' = 1{,}521 < t_{0{,}95;\, 14} = 1{,}7613$

g. No existen elementos suficientes para rechazar Ho debiendo aceptarla; es decir, con el 95% de confiabilidad, se puede afirmar que, la producción media de grasa cuando se utiliza trigo en la dieta no es mayor que cuando se utiliza maíz. Por consiguiente, se recomienda utilizar cualquiera de las dos dietas y evaluar otros factores, tales como, el costo o la disponibilidad de la materia prima en la zona para tomar la decisión final.

Contraste de Hipótesis para la Diferencia de Medias: Datos Pareado

En los contrastes de hipótesis referidos a la diferencia de medias, se asumió que las muestras son independientes. En ocasiones, se presentan casos en los que las muestras no son independientes; esto, por lo general, sucede cuando las observaciones de las dos muestras se asocian por pares, por ejemplo, cuando los mismos individuos escogidos al azar se comparan antes y después de haber recibido algún tratamiento o cuando los grupos control y experimental deban parearse considerando alguna característica de interés.

El objetivo de la asociación por pares es hacer que las unidades experimentales dentro de cada nivel o grupo sean, en lo posible, lo más homogéneas, de tal manera que las diferencias entre los dos niveles o grupos puedan ser atribuidos a las diferencias entre tratamientos y no a efectos externos.

Así, sea (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) un conjunto de observaciones pareadas, donde (x_i, y_i) representa la respuesta de la i-ésima observación, antes y después del tratamiento, como se muestra a continuación:

Par	Antes	Después	Diferencia
1	x_1	y_1	$D_1 = x_1 - y_1$
2	x_2	y_2	$D_2 = x_2 - y_2$
. .	. .	. .	. .
n	x_n	y_n	$D_n = x_n - y_n$

Donde la Media y la Varianza muestrales de las diferencias D_1, D_2, ..., D_n están dadas por:

$$\bar{D} = \sum_{i=1}^{n} \frac{D_i}{n} \qquad (6.10)$$

$$S_D^2 = \sum_{i=1}^{n} \frac{(D_1 - \bar{D})^2}{n-1} = \frac{\sum_{i=1}^{n} D_i^2 - \frac{\left(\sum_{i=1}^{n} D_i\right)^2}{n}}{n-1} \qquad (6.11)$$

Si se supone que las diferencias D_1, D_2, ..., D_n constituyen una muestra aleatoria procedente de una población normal con Media $\mu_D = \mu_1 - \mu_2$ y Varianzas $\sigma_D^2 = V(x_i - y_i)$, $i = 1, 2, ..., n$, el contraste

H_0: $\mu_1 - \mu_2 = 0$
H_1: $\mu_1 - \mu_2 \neq 0$,

es equivalente a

H_0: $\mu_D = 0$
H_1: $\mu_D \neq 0$,

y tiene como estadístico de prueba $T = \frac{\bar{D}}{S_D/\sqrt{n}}$ (6.12)

y la regla de decisión es rechazar Ho si $T > t_{1-\alpha/;\, n-1}$ o $T < t_{\alpha/2;\, n-1}$.

Para el contraste unilateral Ho: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 > 0$

es equivalente a H_0: $\mu_D = 0$

H_1: $\mu_D > 0$

y la regla de decisión es rechazar Ho si $T > t_{1-\alpha;\, n-1}$.

Para el otro contraste unilateral H_0: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 < 0$

es equivalente a H_0: $\mu_D = 0$

H_1: $\mu_D < 0$,

y la regla de decisión es rechazar H_0 si $T < t_{\alpha,\, n-1}$.

Si el tamaño de muestra es grande ($n \geq 30$), independientemente de la distribución muestral de las diferencias $D_1, D_2, ..., D_n$, el Teorema Central del Límite permite utilizar como estadística de prueba

$$\frac{\bar{D}}{\sigma_D/\sqrt{n}} = Z \qquad (6.13)$$

Si σ_D no se conoce, se reemplaza por S_D.

Es recomendable realizar un contraste de hipótesis, para observaciones pareadas, cuando exista una correlación grande entre las parejas de observaciones. En caso contrario, es mejor utilizar un contraste basado en muestras independientes.

Ejemplo 6.12: Con el propósito de establecer el efecto de la dopamina sobre la frecuencia cardiaca de los perros Labrador Retriever, un grupo de machos de esta raza recibió una dosis de 40 mg en 500 ml de Ringer lactato, EV a razón de 1,5 mcg/kg/minuto y se midió la frecuencia cardiaca (latidos/minuto). Luego, al mismo grupo de animales se le midió la frecuencia cardiaca sin recibir la dopamina. Los resultados de este experimento fueron los siguientes:

Frecuencia cardiaca			
Par	**Con dopamina**	**Sin dopamina**	**Diferencia**
1	154	149	5
2	159	151	8
3	144	142	2
4	148	145	3
5	145	142	3
6	159	145	14
7	160	154	6
8	148	145	3

El ensayo pretende determinar si la dopamina afecta de manera significativa la frecuencia cardiaca.

a. El parámetro de interés es:
   μ_D = μ con dopamina - μ *sin dopamina*

b. H_0: μ con dopamina = μ sin dopamina
   H_1: μ con dopamina $\neq$ μ sin dopamina

c. $\alpha = 0{,}05$

d. El estadístico de prueba es como en (6.13):

$$T = \frac{\bar{D}}{S_D/\sqrt{n}}$$

e. Ho se rechaza si: $T > t_{1-\alpha/2,\, n-1} : t_{0,975,\, 7} = 2{,}365$ o
$T < t_{\alpha/2,\, n-1}: t_{0,025,\, 7} = -2{,}365$

f. Para calcular el valor del estadístico de prueba, en primer lugar, se calcula:

$$\bar{D} = \sum_{i=1}^{n} \frac{D_i}{n} = \frac{44}{8} = 5{,}5$$

y la Varianza:

$$S_D^2 = \frac{\sum_{i=1}^{n} D_i^2 - \frac{(\sum_{i=1}^{n} D_i)^2}{n}}{n-1} = \frac{352 - \frac{(44)^2}{8}}{8-1} = \frac{110}{7} = 15{,}714$$

$$S_D = \sqrt{S_D^2} = \sqrt{15{,}714} = 3{,}964$$

El estadístico de prueba se calcula así:

$$T = \frac{\bar{D}}{\frac{S_D}{\sqrt{n}}} = \frac{5{,}5}{\frac{3{,}964}{\sqrt{8}}} = \frac{5{,}5}{1{,}40} = 3{,}92$$

g. Dado que $T = 3{,}92 > t_{1-\alpha/2,\, n-1} = t_{0,975,\, 7} = 2{,}365$, Ho se rechaza, es decir, con el 95% de confianza, se puede concluir, que la frecuencia

> cardiaca en los perros Labrador Retriever se ve afectada significativamente por la dopamina.

Cálculo del Tamaño de la Muestra para Estimar la Diferencia de Medias

Muestras Independientes. Cuando los tamaños de las muestras son suficientemente grandes, se trabaja con igual tamaño de muestra en los dos grupos y, además, se asume que la distribución de la variable cuantitativa en los dos grupos es normal, entonces una forma bastante aceptable para calcular qué tan grande debe ser la muestra, está dado por:

$$n = \frac{2(Z_\alpha + Z_\beta)^2 \sigma^2}{d^2} \qquad (6.14)$$

si la prueba es unilateral y las Varianzas poblacionales son iguales a σ^2.

Donde:

n = Elementos necesarios en cada una de las muestras

Z_α = Valor Z correspondiente al valor α fijado

Z_β = Valor Z correspondiente al valor β fijado

σ^2 = Varianza de la variable cuantitativa analizada

d = Valor mínimo de la diferencia que se desea detectar entre las dos Medias (datos cuantitativos)

Si la prueba es bilateral y las Varianzas de las dos poblaciones son iguales a σ^2, n se calcula como:

$$n = \frac{2(Z_{\alpha/2} + Z_\beta)^2 \sigma^2}{d^2} \qquad (6.15)$$

Los valores $Z_{\alpha/2}$ según el nivel de confianza y Z_{β}, según el poder de la prueba, se indican en la Tabla 4 del Apéndice, en la forma como se había indicado antes.

Cuando se pretende calcular el tamaño de muestra para el contraste de hipótesis de dos Medias, será necesario establecer de antemano los siguientes aspectos:

- La magnitud de la diferencia que se desea detectar.
- El riesgo de cometer error tipo α (tipo I).
- La potencia o poder estadístico de la prueba que se calcula como (1 - β), donde β es el riesgo de cometer error tipo II
- Definir si la hipótesis es unilateral o bilateral.

Ejemplo 6.13: En un sistema de ceba intensiva de novillos Angus, se utilizó dos fuentes de proteína: la torta de soya y la torta de canavalia, siendo ésta la única variante de las dietas. Para recomendar una de ellas, se realizará una Prueba de Hipótesis para comparar la ganancia media de peso por día. Se pregunta ¿Cuántos novillos de cada grupo deben ser pesados para realizar la Prueba de Hipótesis?

Para responder esta pregunta, se deben considerar los siguientes aspectos:

- La torta de soya, es la fuente de proteína que tradicionalmente se ha utilizado, alcanzando rendimientos de ganancia de peso medios de 1,5 kg/día, bajo ceba intensiva (feed lot).
- Los productores de la zona estarían dispuestos a variar la fuente proteínica, si alcanza un rendimiento medio superior a

1,8 kg. Así que la magnitud de la diferencia a detectar, será 0,3 kg/día (1,8 - 1,5).

- Estudios previos indican, que la Desviación Estándar para la variable de interés, es 0,74 kg.
- El poder estadístico de la prueba será del 95% y el error tipo α que, se está dispuesto a aceptar, será del 5%
- La Prueba de Hipótesis será bilateral

Con base en lo anterior, se procede a la aplicación de (6.15)

$$n = \frac{2(Z_{\alpha/2} + Z_{\beta})^2 \sigma^2}{d^2}$$

$$n = \frac{2(Z_{0,025} + Z_{0,95})^2 \sigma^2}{d^2}$$

$$n = \frac{2(1,96 + 1,64)^2 (0,74)^2}{(0,3)^2} = \frac{14,19}{0,09} = 157,71 \approx 158$$

Como resultado de estos cálculos se afirma que se requieren 158 animales por cada grupo.

CONTRASTES RELACIONADOS CON PROPORCIONES

Contraste para la Proporción Poblacional: Muestras grandes.

En muchas ocasiones, el interés se centra en contrastar hipótesis sobre la proporción de individuos de una población que poseen una determinada característica. Como se mencionó anteriormente, un estimador de la

proporción ***P*** es la proporción muestral $\hat{p} = \frac{x}{n}$, donde ***x*** es el número de éxitos. Para un tamaño de muestra ***n*** grande, la variable aleatoria:

$$\frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}} = Z$$

es normal estándar.

Para el contraste de hipótesis bilateral

H_0: $P = P_0$
H_1: $P \neq P_0$

el estadístico de prueba es

$$\frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = Z \qquad (6.16)$$

y la regla de decisión es rechazar H_0, si $Z > z_{1-\alpha/2}$ o $Z < z_{\alpha/2}$.

Para los contrastes unilaterales

H_0: $P = P_0$
H_1: $P > P_0$

y la regla de decisión es rechazar Ho, si $Z > z_{1-\alpha}$.

Para

H_0: $P = P_0$
H_1: $P < P_0$

y la regla de decisión es rechazar Ho, si $Z < z_{\alpha}$.

Ejemplo 6.14: Una casa comercial lanza una campaña, para que los productores de huevos de gallina utilicen una premezcla vitamínica y mineral. La meta de la empresa es lograr que la mitad de los productores de una determinada región utilicen el producto. Luego de un período de tiempo, se toma una muestra al azar de 680 productores, de los cuales, 298 responden que efectivamente están utilizando la premezcla. ¿La compañía está alcanzado el nivel de aceptación del producto?

a. El interés está en probar si el 50% de los productores utilizan el producto, entonces $P_0 = 0{,}5$.

b. La Prueba de Hipótesis toma la forma:

H_0: $P = 0{,}5$
H_1: $P \neq 0{,}5$

c. $\alpha = 0{,}05$.

d. El estadístico de prueba, es según la fórmula (6.16):

$$\frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = Z$$

e. Ho se rechaza si $Z > z_{1-\alpha/2}$ o $Z < z_{\alpha/2}$

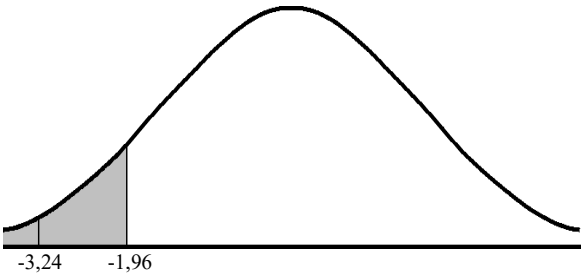
f. Para hallar el valor del estadístico de prueba, antes se requiere calcular:

$$\hat{p} = \frac{x}{n} = \frac{298}{680} = 0{,}438$$

$$Z = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0{,}438 - 0{,}50}{\sqrt{\frac{0{,}50(1-0{,}50)}{680}}} = \frac{-0{,}062}{0{,}0191} = -3{,}24$$

g. Dado que $Z = -3{,}24 < z_{\alpha/2} = z_{0{,}25} = -1{,}96$

H_0 se rechaza, por lo tanto, con el 95% de confiabilidad, se puede concluir que, la proporción de usuarios del producto, es diferente a la establecida por la empresa fabricante, es decir, la campaña no alcanzó el éxito previsto.



Contraste para la Diferencia entre dos Proporciones: Muestras grandes.

Cuando se quiere comparar dos proporciones poblacionales P_1 y P_2, se selecciona dos muestras aleatorias independientes de tamaños n_1 y n_2. Si $\hat{p}_1 = \frac{x_1}{n_1}$ y $\hat{p}_2 = \frac{x_2}{n_2}$ son las proporciones muestrales de éxitos en cada población, se tiene que la variable aleatoria, cuando para n_1 y n_2 son grandes, está dada por:

$$\frac{(\hat{p}_1 - \hat{p}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}} = Z \qquad (6.17)$$

Si $P_1 = P_2 = P$, se tiene que esta estadística se reduce a:

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = Z \qquad (6.18)$$

Donde la proporción común $\hat{p}$ se estima mediante $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$

El contraste bilateral

Ho: $P_1 = P_2$
H_1: $P_1 \neq P_2$

Tiene como estadístico de prueba (6.18): $\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = Z$

La regla de decisión es rechazar H_0 si $Z > z_{1-\alpha/2}$ o $Z < z_{\alpha/2}$.

Ejemplo 6.15: En un experimento se pretende comparar el efecto del Fosfato de Ácido Ascórbico (FFA) y del β-glucán, suministrados en el alimento, sobre la mortalidad de alevinos de trucha arco iris. Al final de la investigación se obtuvo los siguientes resultados:

Con FFA: Cantidad inicial de alevinos: 1.200 = n_1
Cantidad final de alevinos: 1.098
Cantidad de alevinos muertos: 120 = x_1

Con β-glucán: Cantidad inicial de alevinos: 980 = n_2
Cantidad final de alevinos: 891
Cantidad de alevinos muertos: 89 = x_2

b. H_0: $P_1 = P_2$

H_1: $P_1 \neq P_2$ (La proporción de alevinos muertos alimentados con FFA es diferente a la proporción de alevinos muertos alimentados con β-glucan)

c. $\alpha = 0{,}05$

d. El estadístico de prueba es según (6.18)

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = Z$$

e. H_0 se rechaza si:

$Z > z_{1-\alpha/2} = z_{0,975} = 1{,}96$ o
$Z > z_{\alpha/2} = z_{0,025} = -1{,}96$

f. Se requiere calcular:

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{120}{1200} = 0{,}1$$

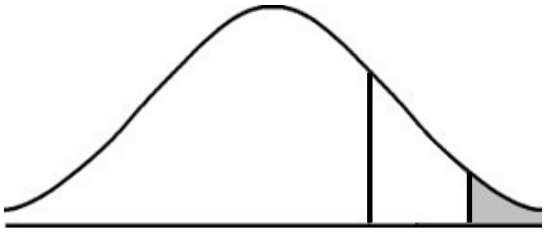
$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{89}{980} = 0{,}0908$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{120 + 89}{1200 + 980} = \frac{209}{2180} = 0{,}096$$

De esta manera, el estadístico de prueba es:

$$\frac{0{,}1 - 0{,}0908}{\sqrt{0{,}096(1 - 0{,}096)\left(\frac{1}{1200} + \frac{1}{980}\right)}} = \frac{0{,}0092}{0{,}000609}$$
$$= 57{,}1872$$

g. Dado que $Z = 57{,}1872$ es mayor que $z_{1-\alpha/2} = z_{0{,}975} = 1{,}96$



Es decir, Z se ubica en zona de rechazo; entonces, H_0 se rechaza y, por lo tanto, con el 95% de confiabilidad, hay evidencia suficiente, para recomendar el uso de Fosfato de Ácido Ascórbico, suministrado en el alimento para disminuir la mortalidad de alevinos de trucha arco iris.

CONTRASTE DE HIPÓTESIS PARA LA VARIANZA POBLACIONAL σ^2

En algunas ocasiones, es de interés, contrastar hipótesis sobre la Varianza poblacional σ^2. Este contraste se basa en la Varianza muestral:

$$S^2 = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n - 1}$$

Es importante recordar que, si S^2 es la Varianza de una muestra aleatoria procedente de una población normal, con Media μ y σ^2 desconocida, la variable aleatoria:

$$\frac{(n-1)S^2}{\sigma^2} = \chi^2$$

tiene una Distribución Chi-cuadrado con n-1 grados de libertad.

El contraste bilateral

H_0: $\sigma^2 = \sigma_0^2$
H_1: $\sigma^2 \neq \sigma_0^2$

tiene como estadístico de prueba $\frac{(n-1)S^2}{\sigma_0^2} = \chi^2$ (6.19)

La regla de decisión es rechazar Ho si $\chi^2 > \chi^2_{1-\frac{\alpha}{2};\, n-1}$ o c

Donde: $\chi^2_{1-\frac{\alpha}{2}}$ es tal que $P\left(\chi^2 > \chi^2_{1-\frac{\alpha}{2}}\right)$ = 1-α/2 y

$\chi^2_{\frac{\alpha}{2}}$ es tal que $P\left(\chi^2_{\frac{\alpha}{2}}\right)$ = α/2.

Ejemplo 6.16: Un profesional experto en mejoramiento animal asegura que, en una línea selecta de conejos blancos de Nueva Zelanda, la Varianza para la característica peso a los 30 días es de 0,012 kg^2. Con una muestra aleatoria de 32 animales, se calculó una Varianza de 0,0101 kg^2. Con α = 0,05 y un contraste bilateral, ¿Se puede afirmar que el profesional tiene razón?

Siguiendo el procedimiento, descrito anteriormente, para una Prueba de Hipótesis:

a) Parámetro de interés: Varianza σ^2

b) H_0: $\sigma^2 = 0{,}012$
   H_1: $\sigma^2 \neq 0{,}012$

c) $\alpha = 0{,}05$

d) El estadístico de prueba es (6.19): $\frac{(n-1)S^2}{{\sigma_0}^2} = \chi^2$

e) Ho se rechaza si $\chi^2 > \chi^2_{0{,}975;\,31} = 46{,}99$.

f) El valor del estadístico de prueba es:

$$\chi^2 = \frac{(n-1)S^2}{{\sigma_0}^2} = \frac{(32-1)\ 0{,}0101}{0{,}012} = 26{,}09$$

g) Dado que $\chi^2 = 26{,}09$ no es mayor que $\chi^2_{0{,}975;\,31} = 46{,}99$, no existen elementos suficientes para rechazar H_0, debiendo aceptarla, por lo tanto, con un nivel de significancia del 5%, el profesional tiene razón; es decir, la línea selecta de conejos blancos de Nueva Zelanda tiene una Varianza de 0,012 kg^2, para la característica peso a los 30 días.

En este ejemplo, es conveniente recordar que, por lo general, los planes de mejoramiento genético por selección buscan incrementar la Media de los caracteres productivos y tener grupos homogéneos, es decir, de Varianza menor con respecto al grupo

control no seleccionado, de allí la importancia de probar si la Varianza es igual al parámetro o no.

Contraste de Hipótesis para la Igualdad de dos Varianzas $\boldsymbol{\sigma_1^2}$ y $\boldsymbol{\sigma_2^2}$

En el capítulo anterior se estableció, que $\boldsymbol{S_1^2}$ y $\boldsymbol{S_2^2}$ son las Varianzas de dos muestras aleatorias independientes, con tamaños n_1 y n_2, procedentes de dos poblaciones normales, con Varianzas σ_1^2 y σ_2^2, respectivamente, entonces, la variable aleatoria

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = F$$

sigue una distribución F de Fisher, con n_1 -1 grados de libertad en el numerador y n_2 -1 grados de libertad en el denominador.

El contraste bilateral
Ho: $\sigma_1^2 = \sigma_2^2$
H_1: $\sigma_1^2 \neq \sigma_2^2$

Tiene como estadística de prueba: $\frac{S_1^2}{S_2^2} = F$ (6.20)

La regla de decisión es rechazar Ho si $F > f_{1-\alpha/2,\, n1-1,\, n2-1}$ o $F < f_{\alpha/2,\, n1-1,\, n2-1}$.

Ejemplo 6.17: En el Ejemplo 6.9, para realizar la Prueba de Hipótesis, se asumió que las Varianzas $\boldsymbol{\sigma_1^2}$ y $\boldsymbol{\sigma_2^2}$ eran iguales. El procedimiento correcto para realizar la Prueba de Hipótesis debió incluir, como paso previo, establecer estadísticamente la igualdad de las Varianzas, lo cual se realiza mediante una Prueba de Hipótesis, utilizando como estadístico de prueba (6.20).

Siguiendo el proceso para Pruebas de Hipótesis, se tiene:

a) Parámetro de interés: Cociente de Varianzas $\frac{\sigma_1^2}{\sigma_2^2}$

b) H_0: $\sigma_1^2 = \sigma_2^2$

H_1: $\sigma_1^2 \neq \sigma_2^2$

c) $\alpha = 0{,}05$

d) El estadístico de prueba es (6.20): $\frac{S_1^2}{S_2^2} = F$

e) H_0 se rechaza si

$F > f_{1-\alpha/2,\, n1-1,\, n2-1} = f_{0{,}975,\, 14,\, 14} = 2{,}9786$ o
$F < f_{\alpha/2,\, n1-1,\, n2-1} = f_{0{,}025,\, 14,\, 14} = 0{,}3357$

f) El valor del estadístico de prueba es:

$$F = \frac{S_1^2}{S_2^2} = \frac{0{,}0849}{0{,}0664} = 1{,}274$$

g) Dado que F (1,279) no es mayor que $f_{0{,}975,\, 14,\, 14}$ (2,9786) y tampoco menor que $f_{0{,}025,\, 14,\, 14}$ (0,3357), entonces, H_0 se rechaza, es decir, se asume que las Varianzas del diámetro de las ovas de bocachico, de la población que recibió EPC no presenta diferencias significativas con respecto a la que recibió HCG, considerando que son iguales, con el 95% de confiabilidad.

Ejemplo 6.18: En el Ejemplo 6.11, para realizar la Prueba de Hipótesis, se asumió que las Varianzas $\boldsymbol{\sigma_1^2}$ y $\boldsymbol{\sigma_2^2}$ eran diferentes. El

procedimiento correcto para realizar la Prueba de Hipótesis debió incluir, como paso previo, establecer estadísticamente la igualdad de las Varianzas, lo cual se realiza, mediante una Prueba de Hipótesis, utilizando como estadístico de prueba (6.20).

Siguiendo el proceso para Pruebas de Hipótesis, se tiene:

a) Parámetro de interés: Cociente de Varianzas $\frac{\sigma_1^2}{\sigma_2^2}$

b) H_0: $\sigma_1^2 = \sigma_2^2$
H_1: $\sigma_1^2 \neq \sigma_2^2$

d) $\alpha = 0{,}10$

e) El estadístico de prueba es (6.20): $\frac{S_1^2}{S_2^2} = F$

f) Ho se rechaza si

$F > f_{1-\alpha/2,\, n1-1,\, n2-1} = f_{0{,}95,\, 11,\, 11} = 2{,}879$ o
$F < f_{\alpha/2,\, n1-1,\, n2-1} = f_{0{,}05,\, 11,\, 11} = 0{,}3549$

g) El valor del estadístico de prueba es:

$$F = \frac{S_1^2}{S_2^2} = \frac{0{,}011509}{0{,}001454} = 7{,}9154$$

h) Dado que F (7,92) es mayor que $f_{0{,}95,\, 11,\, 11}$ (2,879), entonces, H_0 se rechaza, con lo cual se pudo establecer, con el 95% de confiabilidad, que las Varianzas de las dos poblaciones, de vacas alimentadas con trigo y con maíz, presentaron diferencias significativas, es decir, son diferentes ($\sigma_1^2 \neq \sigma_2^2$).

CONTRASTE DE BONDAD DE AJUSTE

Los contrastes de hipótesis, vistos anteriormente, se refieren a parámetros de poblacionales, como μ, σ^2 y P, en los que se conoce la forma de la distribución de la población de interés. En muchas ocasiones, no se conoce la Función de Distribución y, se desea probar la hipótesis de que dicha población se ajusta a un modelo o tiene una Función de Distribución dada; así, se puede tener interés en probar, si una muestra procede de una Distribución Normal, Poisson, o de cualquier otra.

Este tipo de pruebas se denominan ***Pruebas de Bondad de Ajuste***, puesto que comparan los resultados de la muestra, con aquellos que se esperaría observar, siempre y cuando los resultados muestrales se ajusten a una forma de distribución supuesta. El procedimiento consiste, en clasificar los valores que se observan en categorías o intervalos de clase mutuamente excluyentes y luego comparar las frecuencias observadas y esperadas.

El contraste se puede definir de la siguiente manera: suponiendo que se tiene una muestra aleatoria de ***n*** observaciones, cada una de las cuales podrá ser clasificada en una y sólo en una de las posibles ***k*** categorías, sean O_1, O_2, ..., O_k, las frecuencias observadas de las ***k*** categorías que, según la Distribución de Probabilidad propuesta, se espera que ocurran con frecuencia e_1, e_2, ..., e_k, denominadas Frecuencias Esperadas, tal como se muestra a continuación:

Categorías	1	2	3	....	k
Frecuencias observadas	O_1	O_2	O_3	....	O_k
Frecuencias esperadas	E_1	E_2	E_3	....	E_k

Una medida de la discrepancia existente entre las frecuencias observadas y las frecuencias esperadas o teóricas, se encuentra definida por:

$$\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \chi^2 \qquad (6.21)$$

Donde χ^2 es el valor de una variable aleatoria, cuya Distribución Muestral, se aproxima a la Distribución Chi-cuadrado con k-1 grados de libertad.

O_i y E_i representan las frecuencias observada y esperada, respectivamente, para la i-ésima categoría.

$$\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i = n \qquad (6.22)$$
$$E_i = np_i, \quad i = 1,\ 2,\ \ldots,\ k$$

Donde $\boldsymbol{p_i}$ es la Probabilidad de que una observación se encuentre en la i-ésima categoría.

El estadístico (6.21) será pequeño, si las frecuencias observadas y esperadas son muy cercanas, por el contrario, será muy grande si las diferencias son altas. La Regla de Decisión es rechazar la hipótesis de que la Distribución de la Población en la distribución propuesta, si el valor del estadístico de prueba χ^2 es mayor que $\chi^2_{1-\alpha;\, k-1}$.

Este criterio de decisión se utiliza, si cada una de las frecuencias esperadas es igual o mayor que cinco; en caso contrario, deben unirse las categorías adyacentes.

Ejemplo 2.19: Se pide probar si el peso al destete de una muestra de 50 cuyes, los que se agruparon en intervalos de clase de acuerdo con lo indicado en la Tabla 2.1, cuyos datos se vuelven a presentar, para comodidad del lector, puede ser aproximada mediante una distribución normal con media μ = 240,34 y desviación estándar σ = 55,32.

Clase	Intervalos de Clase Peso al destete (g)	Frecuencia Absoluta	Frecuencia Relativa (%)	Frecuencia Absoluta Acumulada	Frecuencia Relativa Acumulada (%)
1	128-159	6	12	6	12
2	159-190	5	10	11	22
3	190-221	8	16	19	38
4	221-252	4	8	23	46
5	252-283	11	22	34	68
6	283-314	16	32	50	100
Total		$n = 50$	100		

Utilizando los límites reales de clase se tiene:

Intervalos reales de clase	O_i	E_i
127,5 – 158,5	6	3,476
158,5 – 189,5	5	5,476
189,5 – 220,5	8	9,044
220,5 – 251,5	4	11,000
251,5 – 282,5	11	9,853
282,5 – 313,5	16	11,150

15

10,427

Las frecuencias esperadas, de los seis intervalos de clase, se hallan calculando el área bajo la curva normal, comprendida entre los límites reales de cada clase. Para que las frecuencias esperadas sumen $\boldsymbol{n}$, se considera que la primera clase no tiene límite inferior y la última clase no tiene límite superior.

En consecuencia, para la primera clase se tiene:

$$P[y < 158{,}5] = P\left[z < \frac{158{,}5-240{,}34}{55{,}32}\right] = P[z < -1{,}4794] = 0{,}06952 = p_1$$

Por lo tanto, la frecuencia esperada $E_1 = np_1 = 50 \times 0{,}06952 = (3{,}476)$.

Del mismo modo para la segunda clase se tiene:

$$P[158{,}5 < y < 189{,}5] = P\left[\frac{158{,}5-240{,}34}{55{,}32} < z < \frac{189{,}5-240{,}34}{55{,}32}\right]$$
$$= P[-1{,}4794 < z < -0{,}919]$$
$$= F_Z(-0{,}919) - F_Z(-1{,}4794)$$
$$= 0{,}1790 - 0{,}0695 = 0{,}1095$$

En consecuencia: $E_2 = np_2 = 50 \times 0{,}1095 = 5{,}476$.

Las frecuencias esperadas E_3, E_4 y E_5 se hallan de manera similar.

Para la última clase:

$$P[y > 282{,}5] = P\left[Z > \frac{282{,}5-240{,}34}{55{,}32}\right] = P[Z > 0{,}7621] = 1 - F_Z(0{,}7621)$$

$$= 1 - 0{,}7770 = 0{,}2230$$

De aquí se tiene que $E_6 = np_6 = 50 \times 0{,}2230 = 11{,}150$.

En la clase cuatro, la frecuencia observada es 4, por lo tanto se unen las clases cuarta y quinta.

Para efectuar el contraste de hipótesis, se sigue el esquema ya explicado anteriormente:

b. La hipótesis se plantea de la siguiente manera:

H_0: La muestra proviene de una población con distribución s normal.

H_1: La muestra proviene de una población que no posee distribución normal.

c. Se establece el $\alpha = 0{,}01$.

d. El estadístico de prueba está indicado en (6.7.1).

e. H_0 se rechaza si $\chi^2 > \chi^2_{1-a;\ k-1-m} = \chi^2_{0{,}99;\ 2} = 9{,}2103$, puesto que hay k=5 clases y se estimaron dos parámetros μ y σ^2, del conjunto de datos, se tiene $m = 2$.

f. El valor del estadístico de prueba es:

$$\chi^2 = \frac{(6-3{,}347)^2}{3{,}347} + \frac{(5-5{,}476)^2}{5{,}476} + \frac{(8-9{,}044)^2}{9{,}044} + \frac{(15-10{,}427)^2}{10{,}427} + \frac{(16-11{,}150)^2}{11{,}15} = 6{,}1106$$

f. Dado que $\chi^2 = 6{,}1106 < \chi^2_{0{,}99;\ 2} = 9{,}2103$, no existen elementos suficientes para rechazar H_0, debiendo aceptarla y, por lo tanto, con confiabilidad del 99%, se concluye que la distribución se ajusta a una normal, con $\mu = 240{,}34$ y $\sigma = 55{,}3$.

Otra interesante y útil aplicación de la Prueba de Bondad de Ajuste, se presenta con frecuencia, en estudios clásicos de la Genética Mendeliana, cuando se pretende establecer, si dos o más rasgos se heredan de forma independiente. Una manera bastante difundida para determinar esta condición consiste, en probar si las frecuencias observadas de los fenotipos en la segunda generación filial F2, se ajustan a las frecuencias esperadas o teóricas de los mismos. Si la prueba estadística χ^2 permite concluir el ajuste de las frecuencias, se concluye, que los

rasgos se heredan de modo independiente, de lo contrario se deben proponer y comprobar otros mecanismos hereditarios.

Ejemplo 6.20: En los Bovinos *Bos taurus,* la ausencia de cuernos es dominante sobre la presencia de los mismos y, a su vez, en ciertas razas como la Holstein, el color negro del pelo es dominante sobre el color rojo. Para probar si estos dos rasgos son o no son independientes, se procede del siguiente modo:

- Se cruzan individuos sin cuernos y de color negro, cuyo genotipo es CCNN con animales con cuernos y de color rojo, con genotipo ccnn. Todos los descendientes de este cruce constituyen la primera generación filial F1 y, su fenotipo es sin cuernos, con pelo de color negro, por ende, su genotipo será CcNn.

- Los individuos F1 se aparean entre sí. Los descendientes de este cruzamiento, constituyen la segunda generación filial F2, donde se observan cuatro fenotipos: sin cuernos y color negro, sin cuernos y color rojo, con cuernos y color negro, con cuernos y color rojo.

- Si los dos caracteres se heredan en forma independiente, los cuatro fenotipos antes señalados, seguirán una distribución 9:3:3:1.

- Ho: los cuatro fenotipos de la F2 siguen una distribución 9:3:3:1 y por lo tanto las dos características (cuernos y color del pelo) se heredan en forma independiente.

H_1: los cuatro fenotipos de la F2 no siguen una distribución 9:3:3:1 y por lo tanto, las dos características (cuernos y color del pelo) no se heredan en forma independiente.

La Regla de Decisión consiste en rechazar H_0 si el estadístico de prueba χ^2es mayor que el χ^2de la tabla.

En un experimento específico, se obtuvo los resultados indicados enseguida, como fruto del cruce entre individuos F1; se pide probar, con $\alpha = 0{,}5$, si las frecuencias observadas se ajustan a las frecuencias teóricas 9:3:3:1.

Sin cuernos y negros: 649
Sin cuernos y rojos: 198
Con cuernos y negros: 216
Con cuernos y rojos: 83

El estadístico de prueba se obtiene con (6.7.1) $\chi_c^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$

Las frecuencias esperadas, para casos como éste, se obtienen multiplicando el total de individuos (n) por la Probabilidad de ocurrencia de cada fenotipo.

Para mayor claridad, es recomendable construir una tabla como la indicada a continuación:

Fenotipo	**Frecuencia Observada (O_i)**	**Frecuencia Esperada (E_i)**
Sin cuernos y negro	649	1146 × (9/16) = 644,625
Sin cuernos y rojo	198	1146 × (3/16) = 214,875
Con cuernos y negro	216	1146 × (3/16) = 214,875
Con cuernos y rojo	83	1146 × (1/16) = 71,825
Total (n)	**1146**	**1146**

$$\chi_c^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \frac{(649 - 644{,}625)^2}{644{,}625} + \frac{(198 - 214{,}875)^2}{214{,}875}$$

$$+ \frac{(216 - 214{,}875)^2}{214{,}875} + \frac{(83 - 71{,}825)^2}{71{,}825} = 3{,}1673$$

χ_c^2 = 3,1673. El χ^2de la tabla, para el α = 5% y (c-1) × (f-1) = 3 grados de libertad es de 7,81, lo que conduce a la aceptación de Ho y, por consiguiente, se concluye, que las frecuencias observadas se ajustan a las frecuencias esperadas 9:3:3:1, con el 95% de confiabilidad.

PRUEBAS DE INDEPENDENCIA

Muchas veces, el interés recae en probar la Hipótesis Nula, donde dos criterios de clasificación que se aplican al mismo conjunto de elementos son independientes. La clasificación de una muestra de elementos, en dos criterios de clasificación, se representa mediante una Tabla de Contingencia, en la que las filas representan las categorías o niveles de uno de los criterios y las columnas representan las categorías o niveles del otro criterio.

Para mayor claridad, se considera el caso de un experimento para estudiar el efecto de la vacunación contra el New Castle en Broilers. Los resultados se muestran a continuación:

Característica B

Característica		Enfermos	Sanos
	Vacunados	10	80
	No vacunados	20	38

Para esta situación particular, las categorías de la característica A son animales vacunados y no vacunados y, las categorías o niveles de la característica B son animales que enfermaron y no enfermaron; el interés, es constatar la hipótesis de que no hay diferencia significativa entre los grupos con o sin vacuna, es decir, la presencia de la enfermedad y la vacunación son independientes.

De manera general, supongamos que se estudian dos atributos o características A y B, donde A tiene ***r*** categorías y B tiene ***c*** categorías, de tal manera que haya $r \times c$ clasificaciones cruzadas, donde, además, O_{ij} representa el número de observaciones correspondiente al cruce de las categorías ***i*** de A y ***j*** de B para $i = 1, 2, ..., r$ y $j = 1, 2, ..., c$, como se indica en la Tabla 6.2, denominada ***Tabla de Contingencia r × c***:

Tabla 6.2

Tabla de Contingencia para dos atributos A y B con *r* y *c* Categorías.

Característica A	**Característica B**						**Total**
Categorías	1	2	...	*j*	...	*c*	
1	O_{11}	O_{12}	...	O_{1j}	...	O_{1c}	$O_{1.}$
2	O_{21}	O_{22}	...	O_{2j}	...	O_{2c}	$O_{2.}$
. .	. .	. .	. .	. .	. .	. .	. .
i	O_{i1}	O_{i2}	...	O_{ij}	…	O_{ic}	$\mathbf{O}_{i.}$
. .	. .	. .	. .	. .	. .	. .	. .
R	O_{r1}	O_{r2}	…	O_{rj}	…	O_{rc}	$O_{r.}$
Total	$\boldsymbol{O}_{.1}$	$\boldsymbol{O}_{.2}$	**…**	$\boldsymbol{O}_{.j}$	**…**	$\boldsymbol{O}_{.c}$	***N***

$O_{.j} = O_{1j} + O_{2j} + ... + O_{rj}$

$O_{i.} = O_{i1} + O_{i2} + ... + O_{ic}$

A estos totales de filas o columnas se los denomina ***Frecuencias Marginales***.

Las hipótesis que se plantean son:
H_0: Los dos atributos son independientes.
H_1: Los dos atributos no son independientes.

El estadístico de prueba es: $\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} = \chi^2$ que tiene una Distribución Chi-cuadrado con $(r-1)\,(c-1)$ grados de libertad.

Un estimador de E_{ij} es $\hat{E}_{ij} = \frac{O_{i.}\,O_{.j}}{n}$.

La hipótesis de independencia se rechaza si $\chi^2 > \chi^2_{1-a;\ (r-1)\,(c-1)}$.

Ejemplo 6.21: Un veterinario desea probar si la infestación por parásitos gastrointestinales en caninos es independiente de la raza de perros. Para cumplir con este objetivo, tomó al azar, una muestra de animales de las razas Bóxer, Dálmata y Doberman. Al realizar los correspondientes exámenes coprológicos, encontró los siguientes resultados, que representan el número de animales positivos al parásito:

Raza	**Parásito**			**Total**
Categorías	***Ancilostoma sp***	***Tiphylidium canis***	***Toxocara canis***	
Bóxer	27	24	18	69
Dálmata	19	32	22	73
Doberman	18	26	17	61
Total	**64**	**82**	**57**	**203**

Para desarrollar la Prueba de Hipótesis, se empieza por estimar las frecuencias esperadas:

$\hat{E}_{(\text{bóxer, ancilostoma})} = \frac{69 \times 64}{203} = 21{,}75$

$\hat{E}_{(\text{bóxer, tiphylidium})} = \frac{69 \times 82}{203} = 27{,}87$

$\hat{E}_{(\text{bóxer, toxocara})} = \frac{69 \times 57}{203} = 19{,}37$

$\hat{E}_{(\text{dálmata, ancilostoma})} = \frac{73 \times 64}{203} = 23{,}01$

$\hat{E}_{(\text{dálmata, tiphylidium})} = \frac{73 \times 82}{203} = 29{,}49$

$\hat{E}_{(\text{dálmata, toxocara})} = \frac{73 \times 57}{203} = 20{,}50$

$\hat{E}_{(\text{doberman, ancilostoma})} = \frac{61 \times 64}{203} = 19{,}23$

$\hat{E}_{(\text{doberman, tiphylidium})} = \frac{61 \times 82}{203} = 24{,}64$

$\hat{E}_{(\text{doberman, toxocara})} = \frac{61 \times 57}{203} = 17{,}13$

Posteriormente, se sigue el procedimiento general del contraste de hipótesis:

- Las hipótesis se plantean de la siguiente manera:

  H_0: El tipo de parásito es independiente de la raza del perro (Afecta por igual a todas las razas).

  H_1: El tipo de parásito no es independiente de la raza del perro (alguna(s) razas(s) son más susceptibles).

- Se fija $\alpha = 0{,}05$.
- Se calcula el estadístico de prueba: $\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} = \chi^2$
- Ho se rechaza si $\chi^2 > \chi^2_{1-a;\ (3-1)\,(3-1)} = \chi^2_{0{,}95;\,4} = 9{,}49$.
- El valor del estadístico de prueba es:

$$\chi^2 = \frac{(27-21{,}75)^2}{21{,}75} + \frac{(24-27{,}87)^2}{27{,}87} + \ldots + \frac{(17-17{,}13)^2}{17{,}13}$$
$$= 3{,}08$$

- Dado que $\chi^2 = 3{,}08$ es menor que $\chi^2_{0{,}95;\,4} = 9{,}49$, no existen elementos suficientes para rechazar H_0, por lo tanto, se concluye, con un nivel de significancia del 5%, que la especie de parásito es independiente de la raza de los perros, es decir, el parásito afecta por igual a las razas Bóxer, Dálmata y Doberman, o también podría decirse que ninguna de las razas muestra una mayor capacidad para defenderse del ataque de los parásitos estudiados.

PRUEBA DE HOMOGENEIDAD

Esta prueba es similar a la Prueba de Independencia, sólo que en esta última, los totales de las filas y columnas de una Tabla de Contingencia, se determinan al azar, en cambio, en la Prueba de Homogeneidad, los totales de las filas o de las columnas se escogen de antemano, es decir, la población se divide en subpoblaciones o estratos, de acuerdo con las categorías de clasificación y se toma una muestra de cada una de éstas.

En este caso, el interés radica en probar si las muestras extraídas provienen de poblaciones homogéneas, considerando algún criterio específico de clasificación; en otras palabras, se quiere probar la hipótesis que las proporciones de la población en cada fila no presentan diferencias significativas. A ésta se la denomina ***Prueba de Homogeneidad*** y, como ya se dijo, se basa en los mismos principios de la Prueba de Independencia. Es conveniente recalcar, que, en este caso, el término homogeneidad hace referencia, a una igualdad en las proporciones de individuos que poseen la

misma característica, en dos poblaciones diferentes y la única diferencia con relación con la Prueba de Independencia, consiste en que los totales marginales de las filas son previamente determinados por el investigador.

Ejemplo 6.22: El desplazamiento de abomaso a la izquierda (DAI) es uno de los problemas más serios en las ganaderías especializadas en producción lechera con vacas de alta producción e intensamente seleccionadas. Los expertos en nutrición creen que, el alto contenido de granos y la disminución de fibra en las dietas, está asociada a mayores tasas de incidencia de DAI. Para confirmar esta explicación, se determinó la presencia de desplazamiento de abomaso en 180 vacas provenientes de una zona, donde las dietas se balancean adecuadamente, es decir, sin exceso de granos; en otra zona donde las dietas se caracterizan por altas cantidades de grano, se estudió 160 animales, encontrando los siguientes resultados:

Con desplazamiento de abomaso			
Zona	**Si**	**No**	**Total**
1	$O_{11} = 18$	$O_{12} = 162$	$n_1 = 180$ (fijo)
2	$O_{21} = 48$	$O_{22} = 112$	$n_2 = 160$ (fijo)
Total	$\boldsymbol{n_1 = 66}$ **(aleatorio)**	$\boldsymbol{n_2 = 274}$	$\boldsymbol{n = 340}$

Con base en estas cifras, se solicita determinar si las dos zonas son homogéneas en cuanto a presentación de casos de DAI o si esta patología depende de la dieta suministrada.

Las dos poblaciones serán homogéneas si la proporción de animales enfermos es igual, es decir, que H_0: $P_1 = P_2$ y la hipótesis alterna plantea que las poblaciones no son homogéneas, ya que

las proporciones de animales enfermos es diferente, es decir, H_1: $P_1 \neq P_2$.

Las frecuencias esperadas E_i se calculan del siguiente modo:

$$E_{11} = n_1 \frac{n_1}{n} = 180\left(\frac{66}{340}\right) = 34{,}97$$

$$E_{12} = n_1 \frac{n_2}{n} = 180\left(\frac{274}{340}\right) = 145{,}05$$

$$E_{21} = n_2 \frac{n_1}{n} = 160\left(\frac{66}{340}\right) = 31{,}05$$

$$E_{22} = n_2 \frac{n_2}{n} = 160\left(\frac{274}{340}\right) = 128{,}94$$

El estadístico de prueba se calcula con (6.7.1)

$$\chi_c^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \frac{(18 - 34{,}97)^2}{34{,}97} + \frac{(162 - 145{,}05)^2}{145{,}05}$$

$$+ \frac{(48 - 31{,}05)^2}{31{,}05} + \frac{(112 - 128{,}94)^2}{128{,}94} = 21{,}70$$

χ_c^2= 21,70 es mayor que el valor tabular para α = *5%* y (*c*-1) (*f*-1) = 3 grados de libertad (7,81), por lo tanto, se rechaza H_0 y se concluye que, las dos poblaciones no son homogéneas, con el 95% de confiabilidad. De acuerdo con este estudio, existen suficientes evidencias, para afirmar, que la mayor proporción de casos de DAI se encuentra en la zona dos influenciadas por la dieta excesiva en granos.

INTRODUCCIÓN AL ANÁLISIS DE VARIANZA DE DATOS

En esta Sección, se describe un método para el análisis de un experimento que tiene más de dos grupos de observaciones. La finalidad es determinar si existen diferencias significativas entre las Medias Poblacionales de los grupos, los cuales se supone, son muestras aleatorias independientes de poblaciones con Distribución Normal. El análisis se basa, en un examen de la variación entre los grupos y dentro de ellos, el que suele denominarse Análisis de Varianza de una vía (ANOVA de una vía) y puede considerarse, una extensión de las Pruebas de Hipótesis para dos Medias, en un experimento, en el que hay más de dos grupos que comparar.

Entonces, una pregunta válida es "¿Por qué es necesario construir otra Prueba de Hipótesis, para comparar la diferencia entre tres o más Medias, cuando es posible hacer muchas pruebas, comparando pares de Medias?". Para acercarse a una respuesta adecuada, se debe recordar, que cada prueba estadística que se realice, corre el riesgo de cometer error tipo I o α, entonces, si se efectúan muchas Pruebas de Hipótesis con los mismos datos, se acumula cada vez más riesgo de cometer error tipo I, por lo tanto, es preferible hacer una prueba estadística que responda a todas las hipótesis simultáneamente, para aumentar la confiabilidad, disminuyendo α.

Análisis de Varianza de un Factor

Se desea comparar las Medias de k poblaciones, para las cuales, se tiene muestras aleatorias independientes de tamaño n_1, n_2, n_3, … n_k respectivamente; esto es, datos relativos a k grupos o tratamientos. De forma convencional, se utiliza el símbolo y para designar los valores muestrales, adicionando dos subíndices, de manera que y_{ij} designa la observación j-ésima

en el i-ésimo grupo o tratamiento, de esta manera, es posible representar los dados como se indica en la Tabla 6.3.

Tabla 6.3

Concentración de Observaciones Muestrales de Muestras Aleatorias Independientes de *k* Poblaciones.

Tratamiento 1	**Tratamiento 2 …**	**Tratamiento k**
y_{11}	Y_{21}	y_{k1}
y_{12}	y_{22}	y_{k2}
y_{13}	y_{23}	y_{k3}
.	.	.
.	.	.
.	.	.
y_{1n1}	y_{2n2}	y_{knk}

Se puede observar en la tabla anterior, que hay *k* tratamientos y, en el i-ésimo tratamiento hay *ni* observaciones, donde $i = 1, 2, \ldots, k$; $j = 1, 2, 3, \ldots, ni$.

Asumiendo que los datos corresponden a una muestra aleatoria, donde cada y_{ij} es una observación de una variable aleatoria Y_{ij}; además, asumiendo que todas las variables aleatorias Y_{ij} son independientes y se distribuyen normal con Media μ_i y Varianza común σ^2 $(\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2)$, el Análisis de Varianza de un factor está diseñado para contrastar la siguiente hipótesis nula:

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$, contra la respectiva alterna:
H_1: por lo menos una media tiene diferencia significativa.

Estas hipótesis nula y alterna se pueden expresar por las hipótesis equivalentes:

H_0: $\tau_1 = \tau_2 = \ldots = \tau_k = 0$
H_1: por lo menos un τ_i es significativamente diferente de 0.

Procedimiento para el ANOVA

En el procedimiento se introducen los valores o cantidades a utilizar para verificar la hipótesis. La j-ésima Media muestral que corresponde al i-ésimo tratamiento, es:

$$\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}; i = 1, 2, \dots, k$$

Donde n_i representa el número de observaciones del grupo o tratamiento *i*.

La hipótesis nula especifica, que las *k* poblaciones tienen una Media común. El proceso lógico, sería construir un estimador de esta Media común, a partir de los datos muestrales, es decir, la suma total de observaciones muestrales dividida entre el número total de observaciones, entonces:

$$n = \sum_{i=1}^{k} n_i$$

La media global de todas las observaciones muestrales se puede escribir como

$$\bar{y} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}}{n} \qquad (6.23)$$

Donde la doble sumatoria indique que se suma sobre todos los grupos y sobre todas las observaciones de cada grupo. Una expresión equivalente para la media global es:

$$\bar{y} = \frac{\sum_{i=1}^{k} n_i \bar{y}_i}{n} = \frac{n_1 \bar{y}_1 + n_2 \bar{y}_2 + \cdots + n_k \bar{y}_k}{n_1 + n_2 + \cdots + n_k} \qquad (6.24)$$

Las observaciones consignadas en la Tabla 6.3 pueden escribirse mediante el modelo lineal:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad (6.25)$$

$i = 1, 2, \ldots, k$

$j = 1, 2, \ldots, n$

Donde y_{ij} es una variable aleatoria, que denota la ij-ésima observación; μ es el parámetro común a todos los tratamientos, denominada Media Global; τ_i es un parámetro asociado al i-ésimo tratamiento; ε_i es un componente de error aleatorio.

El modelo también puede ser expresado en la siguiente forma: $y_{ij} = \mu_i + \varepsilon_{ij}$, donde $\mu_i = \mu + \tau_i$ es la media del i-ésimo tratamiento; τ_i también es denominado como efecto del i-ésimo tratamiento.

Por consiguiente, cada tratamiento se puede considerar, como una población normal con Media μ_i y Varianza común σ^2.

Prueba de Hipótesis

La prueba se basa en la comparación de dos estimadores independientes de la Varianza poblacional σ^2. Dichas estimaciones se obtienen, dividiendo la variabilidad total de los datos en dos componentes: variabilidad dentro de grupos y variabilidad entre grupos, así:

$$\text{SCT} = \text{SCE} + \text{SCTr} \qquad (6.26)$$

Donde:
SCT: Suma de Cuadrados Total o Variabilidad Total.

SCE: Suma Cuadrado del Error o Variabilidad Dentro de Grupos.
SCTr: Suma Cuadrado de Tratamientos o Variabilidad Entre Grupos.

La Variabilidad Dentro de Grupos o Suma Cuadrado del Error (SCE), se calcula como:

$$SCE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \qquad (6.27)$$

La Variabilidad Entre Grupos o Suma Cuadrado de Tratamientos (SCTr) se calcula como:

$$SCTr = \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2 \qquad (6.28)$$

La Variabilidad Total o Suma de Cuadrados Total (SCT) se calcula como:

$$SCT = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 \qquad (6.29)$$

Una estimación de σ^2 basado en k grados de libertad, está dada por:

$$S_1^2 = \frac{SCTr}{k-1} = CMTr \qquad (6.30)$$

Cuadrado medio de tratamientos

Una segunda estimación independiente de σ^2, basada en $n - k$ grados de libertad, está dada por:

$$S_2^2 = \frac{SCE}{n-k} = CME \qquad (6.31)$$

Cuadrado medio del error

En este caso, la hipótesis nula a contrastar es:

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$, contra la respectiva alterna:
H_1: por lo menos una media tiene diferencia significativa.

La Regla de Decisión que se utiliza en este contraste es:

$$\frac{CMTr}{CME} = Fc > f_{1-\alpha;\, k-1;\, n-k}$$

Los cálculos de un Análisis de Varianza de un factor, para realizar la Prueba de Hipótesis, se resumen como se indica en la Tabla 6.4.

Tabla 6.4 Resumen de los cálculos del Análisis de Varianza

Fuentes de Variación	**Suma de Cuadrados**	**Grados de Libertad**	**Cuadrado Medio**	***Fc***
Tratamientos	*SCTr*	*k* - 1	*CMTr*	$\frac{CMTr}{CME}$
Error	*SCE*	*n* - *k*	*CME*	
Total	*SCT*	*n* - 1		

Para facilitar los cálculos del Análisis de Varianza de un factor, se utilizan los totales, así:

- Suma de todas las observaciones bajo el tratamiento *i*:

$$T_i = \sum_{j=1}^{n_i} y_{ij} \qquad (6.32)$$

- Suma de todas las observaciones:

$$T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij} \qquad (6.33)$$

De aquí se deducen los cálculos para el Análisis de Varianza (Tabla 6.4).

$SCT = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{n}$ Donde $n = \sum_{i=1}^{k} n_i$

$$SCTr = \frac{\sum_{i=1}^{k} T_i^2}{n_i} - \frac{T^2}{n}$$

SCE = SCT - SCTr

Ejemplo 6.23 La familia de las Aflatoxinas pertenecen a un grupo de micotoxinas, producidas por un amplio grupo de hongos micotoxigénicos. La aflatoxina M1, aunque es de menor toxicidad, puede tener un efecto carcinogénico, puede encontrarse en la producción de quesos y puede estar asociada al proceso de elaboración, según sea por coagulación enzimática o coagulación ácida. En un estudio realizado en una empresa productora de quesos, se determinó el contenido de AFM1en queso ricotta, elaborado a partir de leche contaminada con cinco concentraciones de hongos micotoxigénicos (T1 a T5) en niveles superiores a las establecidas en el código alimentario de Argentina, con el fin de establecer si existen diferencias. Los resultados en contenido de AFM1 en los quesos, en partes por billón (ppb) fueron los siguientes:

T1	T2	T3	T4	T5
1,53	3,15	3,89	8,18	5,86
1,61	3,96	3,86	5,64	5,46
3,75	3,59	5,70	7,36	5,69
2,89	1,89	5,62	5,33	6,49
3,26	1,45	5,79	8,82	7,81
	1,56	5,33	5,26	9.03
			7.10	7,49

T1	T2	T3	T4	T5
				8,98
$n_1 = 5$	$n_2 = 6$	$n_3 = 6$	$n_4 = 7$	$n_5 = 8$
T1 = 13,09	T2 = 15,50	T3 = 30,01	T4 = 47,69	T5 = 56,81

Con el fin de establecer las diferencias significativas, se aplica el procedimiento de Análisis de Varianza y probar la hipótesis. Para ello, se tiene:

T = T1 + T2 + T3 + T4 + T5 =163,15
$n = n_1 + n_2 + n_3 + n_4 + n_5 = 32$

Para los cálculos, a partir de las fórmulas (6.29) y (6.30), de la siguiente manera:

$$\sum_{i=1}^{5}\sum_{j=1}^{n_i} y_{ij}^2 = (1{,}53)^2 + (1{,}61)^2 + \cdots + (8{,}98)^2 = 994{,}35$$

$$SCT = \sum_{i=1}^{5}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{n} = 994{,}35 - \frac{(163{,}15)^2}{32} = 162{,}54$$

$$SCTr = \frac{\sum_{i=1}^{5} T_i^2}{n_i} - \frac{T^2}{n}$$

$$= \frac{(13{,}04)^2}{5} + \frac{(15{,}60)^2}{6} + \frac{(30{,}01)^2}{6} + \frac{(47{,}69)^2}{7} + \frac{(56{,}81)^2}{8} - \frac{(163{,}15)^2}{32}$$

$$= 952{,}995 - 831{,}81 = 121{,}185$$

$SCE = SCT - SCTr = 162{,}54 - 121{,}185 = 41{,}355$

$$CMTr = \frac{SCTr}{k-1} = \frac{121{,}185}{5-1} = 30{,}296$$

$$CME = \frac{SCE}{n-k} = \frac{41{,}355}{32-5} = 1{,}53$$

$$Fc = \frac{30{,}296}{1{,}531} = 19{,}78$$

Con estos valores se construye la Tabla de ANOVA:

Fuente de Variación	**Suma de Cuadrados**	**Grados de Libertad**	**Cuadrado Medio**	***Fc***
Tratamientos	121,185	4	30,269	19,78
Error	41,350	27	1,531	
Total	162,540	31		

La Regla de Decisión, teniendo un α = 0,05, se expresa de la siguiente manera:

Si Fc (19,78) > $f_{1-\alpha;\, k-1;\, n-k}$ se rechaza H_0

$f_{1-\alpha;\, k-1;\, n-k} = f_{0,95;\, 4;\, 27}$ = 2,73 (Tabla 8 del Apéndice, cuyo aparte se muestra a continuación)

1-α = 0,95 v_1 = ***grados de libertad del numerador***										
v_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	161,45	199,50	215,7	224,5	230,1	233,99	236,77	238,88	240,54	241,88
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	**2,73**	2,57	2,46	2,37	2,31	2,25	2,20

> Teniendo en cuenta que Fc (19,78) > $f_{0,95;\ 4;\ 27}$ (2,73), H_0 se rechaza, es decir, por lo menos una de las Medias presenta diferencias significativas, por lo cual, se concluye que, con el 95% de confiabilidad, existen diferencias significativas por lo menos en una de las concentraciones de Aflatoxina M1 del queso ricotta estudiado.

Como se ha observado, el Análisis de Varianza de un factor es un proceso sencillo, por lo cual, es ampliamente utilizado. Sin embargo, éste sólo puede ser aplicado, cuando las Varianzas de los tratamientos o grupos son iguales o no presentan diferencias significativas, de lo contrario, el experimentador, deberá utilizar otro diseño u otro análisis.

Pruebas de Comparación Múltiple

Cuando se rechaza la hipótesis nula, en consecuencia, se acepta la alterna, de que las Medias no son todas iguales; existen varias pruebas para analizar y definir cuál es el tratamiento más efectivo, uno de los más simples, es el Test de Diferencia Mínima Significativa LSD de Fisher.

Si las muestras de los tratamientos provienen de poblaciones normales e independientes, entonces, para los tratamientos q-ésimo y r-ésimo, es posible demostrar que:

$$\frac{\bar{y}_q - \bar{y}_r}{S\sqrt{\frac{1}{n_q}+\frac{1}{n_r}}} \sim t_{(n-k)} \qquad (6.34)$$

Donde: $S^2 = \frac{SCE}{n-k} = CME$

Las Medias μ_q y μ_r se consideran significativamente diferentes a un nivel de significancia de α si:

$$|\bar{y}_q - \bar{y}_r| > t_{(1-\frac{\alpha}{2};\, n-k)} \times S\sqrt{\frac{1}{n_q} + \frac{1}{n_r}} \qquad (6.35)$$

Ejemplo 6.24: Se aplicó cierto tratamiento con el producto químico Garlon, a uno conjunto de tocones de árboles, con la finalidad de estimular las raíces y propiciar el regeneramiento de brotes. Para ello se utilizó un aerosol, con tres niveles de concentración de Garlon. Al cabo de un tiempo determinado, se midió la altura de los retoños de los árboles. Con los datos que se presentan a continuación se pretende determinar diferencias significativas entres los tratamientos, utilizando un Análisis de Varianza de un factor y si existen diferencias, se utilizará el Test LSD de Fisher, para averiguar, cuál es la concentración de Garlon que mayor efecto tuvo sobre la altura de los retoños.

Altura de los retoños (m.)		
T1	**T2**	**T3**
1,52	1,63	2,56
1,38	1,82	3,32
1,29	1,35	2,76
1,48	1,03	2,63
1,63	2,30	2,12
	1,45	2,77
T1 =7,30	T2 = 9,50	T3 = 16,17
$n_1 = 5$	$n_2 = 6$	$n_3 = 6$
$\bar{y}_1 = 1{,}460$	$\bar{y}_2 = 1{,}597$	$\bar{y}_2 = 2{,}692$

Se procede con los pasos para el Análisis de Varianza, de la siguiente manera:

$\bar{y} = 1{,}944$

$T = 33{,}05$

$n = 17$

$$SCT = \sum_{i=1}^{3}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{n} = 71{,}3047 - \ 64{,}2531 = 7{,}0516$$

$$SCTr = \frac{\sum_{i=1}^{3} T_i^2}{n_i} - \frac{T^2}{n} = 69{,}5322 - 64{,}2531 = 5{,}2791$$

$SCE = SCT - SCTr = 1{,}7725$

De esta manera, la tabla de ANOVA quedaría así:

Fuente de Variación	**Suma de Cuadrados**	**Grados de Libertad**	**Cuadrado Medio**	***Fc***
Tratamientos	5,2791	2	2,640	20,8
Error	1,7725	14	0,127	
Total	7,0516	16		

Si $\alpha = 0{,}05$:

$f_{1-\alpha;\, k-1;\, n-k} = f_{0{,}95;\, 2;\, 4} = 3{,}74$ (Tabla 8 del Apéndice).

Dado que *Fc* (20,8) > $f_{0{,}95;\, 2;\, 4}$ (3,74), se rechaza H_0, es decir, por lo menos una Media presenta diferencias significativas, con el 95% de confiabilidad. En estas condiciones, es necesario utilizar el Test LSD de Fisher, para establecer las diferencias entre Medias y determinar el tratamiento que tiene mayor efecto.

De la Tabla de ANOVA se tiene que:
$S^2 = CME = 0{,}127$, entonces $S = 0{,}356$.

Con base en $\alpha = 0{,}05$:

$t_{(1-\frac{\alpha}{2};\, n-k)} = t_{0{,}975;\, 14} = 2{,}145$ (Tabla 8 del Apéndice).

Aplicando (6.35) se tiene para el primer par de Medias:

$|\bar{y}_2 - \bar{y}_1| = |1{,}597 - 1{,}460| = 0{,}137$

$$t_{(1-\frac{\alpha}{2};\, 14)} \times S\sqrt{\frac{1}{n_q}+\frac{1}{n_r}} = 2{,}145 \times 0{,}356 \times \sqrt{\frac{1}{6}+\frac{1}{5}}$$

$$= 0{,}7636 \times \sqrt{\frac{1}{6}+\frac{1}{5}} = 0{,}4623$$

Teniendo en cuenta que $|\bar{y}_2 - \bar{y}_1|$ (0,137) es menor que 0,4623, se concluye que μ_2, la diferencia no es significativa con respectos μ_1.

Para el segundo par de medias:

$|\bar{y}_3 - \bar{y}_2| = |2{,}695 - 1{,}597| = 1{,}098$

$$t_{(1-\frac{\alpha}{2};\, 14)} \times S\sqrt{\frac{1}{n_q}+\frac{1}{n_r}} = 0{,}7636 \times \sqrt{\frac{1}{6}+\frac{1}{5}} = 0{,}4623$$

Teniendo en cuenta que $|\bar{y}_3 - \bar{y}_2|$ (1,098) es mayor que 0,4623, se concluye que μ_3 tiene diferencia significativa con respectos μ_2.

Para el tercer par de medias:

$|\bar{y}_3 - \bar{y}_1| = |2{,}695 - 1{,}460| = 1{,}235$

$$t_{(1-\frac{\alpha}{2};\,14)} \times S\sqrt{\frac{1}{n_q}+\frac{1}{n_r}} = 0{,}7636 \times \sqrt{\frac{1}{6}+\frac{1}{5}} = 0{,}4623$$

Teniendo en cuenta que $|\bar{y}_3 - \bar{y}_2|$ (1,235) es mayor que 0,4623, se concluye que μ_3 tiene diferencia significativa con respectos μ_1.

De acuerdo con lo anterior, se puede concluir que, con el 95% de confiabilidad, el efecto de los tratamientos 1 y 2 no tienen diferencias significativas; sin embargo, el tratamiento 3 tiene un efecto significativamente diferente con respecto a los tratamientos 1 y 2, por lo cual se considera, el de mayor efectividad al tener la mayor Media (2,695 m de altura).

CAPÍTULO 7

Pruebas no paramétricas

La mayoría de los métodos de inferencia que se han estudiado en este texto, se basan en la suposición de que las observaciones provienen de poblaciones con Distribución Normal, sin embargo, con frecuencia, resulta difícil verificar dicho supuesto, sobre todo cuando los tamaños de muestra son pequeños o los datos son de tipo ordinal; es decir, la variable está constituida por datos en los que sólo el orden de los números es significativo y la distancia entre los números no se presenta a la interpretación práctica.

El término "No Paramétrico" se refiere a procedimientos de inferencia que, no requieren una Distribución Normal en la población, o alguna otra forma especificada en términos de parámetros. Los Procedimientos No Paramétricos continúan ganando popularidad, dado que tienen la posibilidad de ser aplicados a una variedad muy amplia de Distribuciones de Probabilidad en la población. Por lo general, utilizan aspectos simples de los datos de la muestra, como los signos de las mediciones, relaciones de orden o frecuencias de categorías.

Según Johnson (2012), la principal ventaja de tales Pruebas No Paramétricas es que se realizan inferencias exactas, cuando no pueden satisfacerse las suposiciones subyacentes a los llamados métodos estándar. Cuando se satisface la suposición normal, las pruebas estándar tendrán más poder. No obstante, la asimetría u otros alejamientos de la normalidad no surtirán efecto sobre la distribución muestral de algún estadístico no paramétrico, cuando la

hipótesis nula prevalezca. Más aún, este poder por lo general es satisfactorio, aun cuando las poblaciones se desvían de la normalidad. Además, las Pruebas No Paramétricas se aplican, así la elección de una escala numérica particular de medición, sea arbitraria. Sin embargo, su valor más fuerte, es el hecho de que el nivel de significancia es exacto, incluso cuando las poblaciones son bastante anormales".

Las Pruebas No Paramétricas más utilizadas son:

- Prueba de Signos.
- Prueba de Wilcoxon.
- Prueba de Rachas.
- Prueba de U de Mann-Whitney.
- Correlación de Rangos de Spearman.
- Prueba de Kruskal-Wallis.

La Tabla 7.1 muestra el propósito de las Pruebas No Paramétricas y la correspondencia con las Pruebas Paramétricas.

Tabla 7.1

Propósito de las principales Pruebas No Paramétricas y correspondencia con las Pruebas Paramétricas.

Prueba No Paramétrica	**Propósito**	**Prueba Paramétrica**
Signo de Wilcoxon	Datos pareados	Prueba *t* con datos pareados
Rachas	Prueba de aleatoriedad	Ninguna
U de Mann-Whitney	Compara dos muestras independientes	Prueba *t* para muestras independientes

Prueba No Paramétrica	Propósito	Prueba Paramétrica
Rangos de Spearman	Relación entre dos variables clasificadas ordinalmente	Coeficiente de correlación de Pearson
Kruskal-Wallis	Compara tres o más muestras	ANOVA (Prueba F)

PRUEBA DEL SIGNO

En algunas Pruebas No Paramétricas se utiliza, con frecuencia, la Mediana. Así, para una variable aleatoria $\mathcal{Y}$ se define la Mediana $\tilde{\mathcal{Y}}$, de tal manera que $P(\mathcal{Y} > \tilde{\mu}) \leq 0{,}5$ y $P(\mathcal{Y} < \tilde{\mu}) \leq 0{,}5$. Si la variable aleatoria es continua, se define como:

$$P(\mathcal{Y} > \tilde{\mu}) = P(\mathcal{Y} < \tilde{\mu}) = 0{,}5$$

Por otra parte, en el Capítulo 6, cuando se trató sobre Pruebas de Hipótesis para la Media μ, se definió H_0: $\mu = \mu_0$, contra cualquier hipótesis alternativa apropiada, es válida solo si la población de donde proviene la muestra es normal, o aproximadamente normal, o la muestra es grande $n \geq 30$. Si la población de donde proviene la muestra no es normal, o no se aproxima a una normal y la muestra es pequeña $n < 30$, se recurre a una Prueba No Paramétrica. Una de las viables para aplicar es la Prueba del Signo, diseñada para probar hipótesis relativas a la Mediana, en cualquier población en la cual se ha medido una variable continua, de manera que haya una Probabilidad igual a cero de obtener un valor igual a $\tilde{\mu}$ es posible probar la H_0: $\tilde{\mu} = \tilde{\mu}_0$ contra cualquier hipótesis alternativa adecuada, si la población muestreada es continua y simétrica; entonces, la Mediana y la Media son iguales $\mu = \tilde{\mu}_0$, en cuyo caso, es posible probar la hipótesis $H_0 = \mu = \tilde{\mu}_0$ contra cualquier hipótesis alternativa adecuada.

De la misma forma, si la población de donde se obtuvo la muestra es normal, entonces es simétrica y, por lo tanto, la Prueba del Signo puede aplicarse para probar la hipótesis sobre la Media de una población normal.

Entonces, sea y_1, y_2, …, y_n una muestra aleatoria de una población con Mediana $\tilde{\mu}$, suponiendo que es del interés probar la hipótesis nula $H_0 : \tilde{\mu} = \tilde{\mu}_0$ contra cualquier otra alternativa adecuada.

De la definición de Mediana se tiene que $P(y > \tilde{\mu}) = P(y < \tilde{\mu}) = 0{,}5$. Si la hipótesis nula es verdadera y la población es continua, esperamos que la mitad de los valores en la muestra son mayores que $\tilde{\mu}_0$ y la otra mitad corresponde a valores menores que $\tilde{\mu}_0$; si se reemplaza con un signo mas (+), cada valor de la muestra que exceda $\tilde{\mu}_0$ y cada valor de la muestra menor que $\tilde{\mu}_0$, se reemplaza con un signo menos (-), la suma de los signos mas debería ser igual a la suma de los signos menos. Cuando un signo aparece con más frecuencia de la que debería, ocurre solo por el azar, se rechaza la hipótesis nula de que $\tilde{\mu} = \tilde{\mu}_0$.

Dada la hipótesis nula $Ho: \tilde{\mu} = \tilde{\mu}_0$, contra cualquier alternativa adecuada, el estadístico de prueba para la Prueba del Signo, denotada por *S*, es el número de signos mas (+) entre los *n* valores.

Dado que, bajo H_0 cada punto constituye un ensayo independiente con una Probabilidad para el signo mas (+) de 0,5, el estadístico *S* tiene una distribución binomial con parámetros *n* y $p = 0{,}5$.

Para un *n* dado y $p = 0{,}5$, se emplea la Distribución Binomial para hallar regiones críticas con un nivel de significancia α. Para valores grandes de *n,* se utiliza la aproximación de la binomial para la Distribución Normal.

Para la Prueba Bilateral

H_0: $\tilde{\mu} = \tilde{\mu}_0$

H_1: $\tilde{\mu} \neq \tilde{\mu}_0$

Si S_1 es igual al número de valores y_i mayores que $\tilde{\mu}_0$ o si el número de valores y_i es menor que $\tilde{\mu}_0$. Si $+S_2 = n$, entonces, $S_1 = n - S_2$; en cualquier caso, el valor *p* de la Prueba Bilateral es dos veces el valor *p* de la Prueba Unilateral. Por otro lado, es posible utilizar una Distribución Normal con Media $\mu = np$ y $\sigma = \sqrt{np(1-p)}$ para aproximar la Distribución Binomial a una Distribución Normal, cuando *n* es grande ($n \geq 30$).

Si $p = 0{,}5$ la aproximación funciona si $n > 10$, el estadístico de prueba es:

$$Z = \frac{(S \pm 0{,}5) - 0{,}5n}{0{,}5\sqrt{n}} \qquad (7.1)$$

Se utiliza $S + 0{,}5$ cuando $S < n/2$
$S - 0{,}5$ cuando $S < n/2$

Para las hipótesis:
H_0: $\tilde{\mu} = \tilde{\mu}_0$
H_1: $\tilde{\mu} \neq \tilde{\mu}_0$
Ho se rechaza si $Z \leq z_{\alpha/2}$ o $Z \geq z_{1-\alpha/2}$

H_0: $\tilde{\mu} = \tilde{\mu}_0$
H_1: $\tilde{\mu} > \tilde{\mu}_0$
Ho se rechaza si $Z \geq z_{1-\alpha}$

H_0: $\tilde{\mu} = \tilde{\mu}_0$

$H_1: \tilde{\mu} < \tilde{\mu}_0$
Ho se rechaza si $Z \leq z_\alpha$

Ejemplo 7.1: El gerente de producción de una planta procesadora de camarón, afirma que su producto contiene en promedio más de 40 colas por libra. Con el fin de realizar una prueba de calidad, el técnico, encargado de otorgar la certificación, toma una muestra de 10 libras y hace el conteo del número de colas que contiene cada libra, con lo cual, obtiene los siguientes datos:

41 33 43 52 46 42 39 49 53 30

Para dar solución a este planteamiento es necesario probar, con un $\alpha = 0{,}05$, las hipótesis:

$H_0: \tilde{\mu} = 40$
$H_1: \tilde{\mu} > 40$

El estadístico de prueba es *S*, como se muestra en la fórmula (7.1), donde *S* es el número de datos que exceden a 40, es decir $S = 7$.

S tiene una Distribución Binomial, con parámetros $n = 10$ y $p = 0{,}5$; entonces, el valor *p* o la Probabilidad de encontrar un valor mayor o igual a 7, esto es:

Según la Función de Distribución Acumulativa Binomial (Tabla 2 del Apéndice) para $n = 10$, $x = 6$ y $p = 0{,}5$, el valor de $p(X \leq 6) = 0{,}8281$, entonces:

$$p(X > 7) = 1 - p(X \leq 6)$$
$$= 1 - 0{,}8281 = 0{,}1719$$

> Teniendo como valor de $p \geq 0{,}05$, se puede concluir, que no existen elementos suficientes para rechazar H_0, debiendo aceptarla, por cuanto, con el 95% de confiabilidad, no se muestran evidencias de que el número promedio de colas por libra sea mayor que 40.

Prueba del Signo para Datos pareados

En el Capítulo 6, sobre Pruebas de Hipótesis, se analizó las Pruebas para la Diferencia de Medias con datos pareados, con el propósito de eliminar los efectos causados por factores externos; por lo general, esto sucede cuando a los mismos individuos escogidos al azar, se comparan antes y después de haber recibido algún tratamiento.

En este caso, sean $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ n pares de datos de dos variables aleatorias X e Y, si para cada par de datos (x_i, y_i), en el que x_i sea mayor que y_i, se anota un signo mas (+) y, en cada caso contrario, se anota un signo menos (-). Sea p la Probabilidad de que x sea mayor que y, entonces, si la hipótesis nula es que X e Y tienen la misma distribución, el valor de p debe ser igual a 0,5; sin embargo, debe notarse que p puede ser 0,5 aunque las distribuciones de X e Y no sean idénticas, entonces la hipótesis nula para la Prueba de Signos es H_0: $p = 0{,}5$ y la alternativa puede ser H_1: $p > 0{,}5$ o H_1: $p < 0{,}5$ o H_1: $p \neq 0{,}5$.

Ejemplo 7.2: Se desea probar, si una dieta es efectiva para el incremento de peso de peces. Para ello, la dieta es suministrada a un grupo del Estanque A, el cual fue comparado con un grupo control cultivado en el Estanque B, que recibió una dieta comercial comúnmente utilizada en la estación acuícola; los dos grupos estuvieron

conformados por peces de la misma edad y peso. Se desea probar la efectividad de la nueva dieta, con un $\alpha = 0{,}05$.

Los resultados obtenidos del incremento de peso de una muestra de 10 animales, expresados en gramos, son los siguientes:

Grupos	**Incremento de peso en gramos**									
A	287	264	276	278	260	289	290	271	278	297
B	240	270	215	283	259	291	263	238	254	265

Se requiere probar las hipótesis:
H_0: $p = 0{,}5$
H_1: $p > 0{,}5$

Es importante aclarar que, debe rechazarse H_0 si el número de negativos (-) es muy pequeña. El proceso de asignación de signos se indica a continuación:

Pareja	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
A – B	+	-	+	-	+	-	+	+	+	+

Como se puede ver, existen tres signos negativos (-), por lo cual el valor del estadístico $S = 3$.

Teniendo en cuenta que, bajo H_0 S es binomial (Tabla 2 del Apéndice), con parámetros $p = 0{,}5$ y $n = 10$, el valor p o la Probabilidad de encontrar tres o menos signos negativos (-) es:

$P[S \leq 3, n = 10, \ = 0{,}5] = 0{,}1719$

> Dado que, el valor p = 0,1719 es mayor que α = 0,05, no existen elementos suficientes para rechazar Ho, debiendo aceptarla, por lo cual, no es posible afirmar la existencia de un efecto significativo de la dieta A, sobre el incremento de peso de los peces en los cuales se ha probado la dieta, en relación con la dieta control de tipo comercial.

Prueba de Rango con Signo de Wilcoxon

Esta prueba, además del signo, utiliza la magnitud de las diferencias y se utiliza, en el caso de tener una distribución de variable continua.

Se prueba la hipótesis nula H_0: $\tilde{\mu} = \tilde{\mu}_0$ enfrentada a cualquier hipótesis alternativa adecuada. Para ello, a cada valor de la muestra se resta $\tilde{\mu}_0$ y se descarta todas las diferencias iguales a cero. A continuación, se ordenan de menor a mayor los valores absolutos de las diferencias y se les asigna un rango de 1 a la mínima diferencia, 2 a la segunda y así sucesivamente; cuando el valor absoluto de dos o más diferencia es el mismo, se asigna a cada punto, el promedio de los rangos que se asignan, si los valores fueran diferentes.

Si Ho: $\tilde{\mu} = \tilde{\mu}_0$ es verdadera, el total de los rangos de las diferencias positivas T^+ debería se casi igual al total de los rangos de las diferencias negativas T^-.

Si $T = min\{T^+, T^-\}$, dependiendo de la hipótesis alternativa, se busca la prueba de rango con signo en T, T^+ o T^-, así:

H_0: $\tilde{\mu} = \tilde{\mu}_0$
H_0: $\tilde{\mu} \neq \tilde{\mu}_0$
H_0 se rechaza si $T \leq T_\alpha$.

$H_0: \tilde{\mu} = \tilde{\mu}_0$
$H_0: \tilde{\mu} > \tilde{\mu}_0$
H_0 se rechaza si $T^- \leq T_{2\alpha}$.

$H_0: \tilde{\mu} = \tilde{\mu}_0$
$H_0: \tilde{\mu} < \tilde{\mu}_0$
H_0 se rechaza si $T^+ \leq T_{2\alpha}$.

Donde T_α y $T_{2\alpha}$ se hallan en la Tabla de Valores Críticos para la Prueba de Rango con Signo (Tabla 9 del~~a~~ Apéndice). En dicha tabla, T_α y $T_{2\alpha}$ son los valores más grandes para los cuales los valores p correspondientes no exceden a α o 2α, respectivamente.

Para $n > 10$ la Distribución de Muestreo de T^+ se aproxima a la normal, donde:

$$E(T^+) = \frac{n(n+1)}{4} \qquad (7.2)$$

$$V(T^+) = \frac{n(n+1)(2n+1)}{2x4} \qquad (7.3)$$

Es decir, la variable $Z = \frac{T \pm E(T^+)}{\sqrt{Var(T^+)}}$ tiende a ser normal estándar. (7.4)

Ejemplo 7.3: Por experimentación, se sabe que, un tipo de ratón de laboratorio puede alcanzar una ganancia de peso de 65 g durante los tres primeros meses de vida. Para probar esta hipótesis, fueron alimentados 12 ratones con una nueva dieta, desde su nacimiento hasta los tres meses de vida, al final de los cuales, fueron observadas las siguientes ganancias de peso:

64,3 62,3 64,7 68,2 65,0 64,5 60,7 62,4 69,8 67,3 62,5 71,2

Para ello, se utiliza la Prueba de Rangos de Wilcoxon, con un nivel de significancia de $\alpha = 0{,}05$, con el fin de probar si la ganancia media del peso es 65,0 gramos.

De esta manera, las hipótesis a probar son:
H_0: $\mu = 65{,}0$ g
H_1: $\mu \neq 65{,}0$ g

En el procedimiento, se elimina el valor 65,0 y se resta este dato a cada valor, ordenando luego las diferencias, sin considerar el signo, como se muestra a continuación:

Ganancia de Peso	Diferencia	Rango
64,2	- 0,8	3
64,3	- 2,7	7
64,7	- 0,3	1
68,2	+ 3,2	8
64,5	- 0,5	2
60,7	- 4,3	9
62,4	- 2,6	6
69,8	+ 4,8	10
67,3	+ 2,3	4
62,5	- 2,5	5
71,2	+ 6,2	11

De esta manera, las sumas serían como sigue:

$T^+ = 8 + 4 + 10 + 11 = 33$
$T^- = 3 + 7 + 1 + 2 + 9 + 6 + 5 = 33$
$T = min\{33, 33\} = 33$

En la Tabla 9 del apéndice, para $n = 11$ y $\alpha = 0{,}05$ bilateral, $T'_{\alpha} = T'_{0{,}05} = 11$.

Dado que $T > T'_{0{,}05}$, no existen elementos suficientes para rechazar *H*o y, en consecuencia, con el 95% de confiabilidad, se puede afirmar que el tipo de ratón de laboratorio evaluado, alcanza una ganancia promedio de peso de 65,0 g a los tres meses de edad.

Para **datos pareados**, se prueba la hipótesis nula:

H_0: $\mu_1 = \mu_1$ contra cualquier alternativa adecuada
H_1: $\mu_1 \neq \mu_1$, H_1: $\mu_1 > \mu_1$, H_1: $\mu_1 < \mu_1$, y se emplea el mismo procedimiento descrito anteriormente.

Ejemplo 7.4: Algunos hongos del género *Saprolegnia,* son los causantes de la saprolegniosis o "enfermedad del algodón" en los peces de agua dulce. En una estación acuícola, dedicada a la producción de alevinos de trucha arco iris, donde frecuentemente se han presentado brotes de saprolegniosis, se pretende comprobar la efectividad del uso de ácido acético como tratamiento para el control de esta enfermedad en la etapa de alevinaje. Para ello, se tomó 12 muestras aleatorias de alevinos, en los cuales se contó, el número de animales afectados antes y después del tratamiento. Los resultados se muestran como sigue:

Antes	11	10	13	8	13	15	11	11	5	9	18	12
Después	11	7	9	3	8	11	10	9	6	11	12	5

Con $\alpha = 0{,}05$ se desea verificar si el tratamiento con ácido acético fue efectivo en la reducción de la incidencia de la saprolegniosis.

Si se designa: $\mu_1 = \mu$ antes
$\mu_2 = \mu$ después

Las hipótesis a probar serían:

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 > \mu_2$

Los resultados del procedimiento descrito anteriormente se plasman de la siguiente manera:

Antes	Después	Diferencia	Rango
10	7	+3	5
13	9	+4	6,5
8	3	+5	8,5
13	8	+5	8,5
15	11	+4	6,5
11	10	+1	1,5
11	9	+2	3,5
5	6	-1	1,5
9	11	-2	3,5
18	12	+6	10
12	5	+7	11

Se debe aclarar que, fue eliminado el dato 11 por empate. Teniendo en cuenta que los rangos para +1 y -1 serían 1 y 2, entonces el promedio es 1,5 que será asignado a los dos. De

manera similar, para +5 y +5 los rangos serían 8 y 9, el promedio es 8,5 que se asigna a los dos.

De esta manera:

$T^+ = 5 + 6{,}5 + 8{,}5 + 8{,}5 + 6{,}5 + 1{,}5 + 3{,}5 + 10 + 11 = 66$
$T^- = 1{,}5 + 3{,}5 = 5$

Para el caso de las hipótesis:

Ho: $\mu_1 = \mu_2$
H_1: $\mu_1 > \mu_2$

Se rechaza Ho si $T^- \leq T_{2\alpha}$, en la Tabla 9 del Apéndice, para $n = 11$ y $\alpha = 0{,}05$ unilateral:

$T'_{2\alpha} = T'_{0{,}10} = 14$

Dado que $T^- = 5$ es menor que $T'_{2\alpha} = T'_{0{,}10} = 14$, entonces, H_0 es rechazada y, en consecuencia, $\mu_1 > \mu_2$, es decir, el tratamiento con ácido acético ayuda a reducir la saprolegniosis en alevinos de trucha arcoíris.

Otra forma de resolver este ejemplo es aplicando la fórmula (7.4). Dado que $n = 11 > 10$, entonces, es posible aplicar dicha ecuación:

$$Z = \frac{T \pm E(T^+)}{\sqrt{Var(T^+)}}$$

$T^+ = 61$

$$E(T^+) = \frac{n(n+1)}{4} = \frac{11 \times 12}{4} = 33$$

$$V(T^+) = \frac{n(n+1)(2n+1)}{24} = \frac{11 \times 12 \times 23}{24} = 126{,}5$$

$$\sqrt{V(T^+)} = \sqrt{126{,}5} = 11{,}25$$

$$Z = \frac{66 - 33}{11{,}25} = 2{,}93$$

Teniendo como base las hipótesis:

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 > \mu_2$

Se rechaza H_0 si $Z \geq z_{1-\alpha}$

$z_{1-\alpha} = z_{1-0,05} = z_{0,95} = 1{,}64$

Dado que $Z = 2{,}93$ es mayor que $z_{0,95} = 1{,}64$, de igual manera que en el proceso anterior, se rechaza H_0.

PRUEBAS NO PARAMÉTRICAS PARA COMPARAR DOS POBLACIONES CON BASE EN MUESTRAS ALEATORIAS INDEPENDIENTES

Prueba de Mann-Whitney

Esta prueba puede considerarse el equivalente no paramétrico de la Prueba de *t* de Student para dos muestras independientes, estudiada en la Sección relacionada con "Contraste de Hipótesis para la Diferencia de Medias con muestras pequeñas".

Suponiendo que se tiene dos muestras aleatorias independientes de tamaño n_1 y n_2, de dos poblaciones con Distribuciones Continuas, si se juntan los datos de las dos muestras, se les adjudica los rangos y se representa la suma de los rangos de la primera población con R_1, el estadístico del contraste de Mann-Whitney es:

$$U = n_1 + n_2 + \frac{n_1(n_1+n_2)}{2} - R_1 \qquad (7.5)$$

Asumiendo que las dos Distribuciones Poblacionales son idénticas, excepto por posibles diferencias en el parámetro de centralización, se demuestra que:

$$E(U) = \mu_U = \frac{n_1\, n_2}{2} \qquad (7.6)$$

$$V(U) = \sigma_U^2 = \frac{n_1\, n_2(n_1+n_2+1)}{12} \qquad (7.7)$$

Si n_1 y n_2 son mayores que 10, la variable aleatoria $Z = \frac{U-\mu_U}{\sigma_U}$ es aproximadamente normal estándar.

Si el parámetro de centralidad es la Media μ, en las hipótesis:

Ho: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$ $\quad$ Ho se rechaza si $Z \geq z_{1-\alpha/2}$ o $Z \leq z_{\alpha/2}$

Ho: $\mu_1 = \mu_2$
H_1: $\mu_1 > \mu_2$ $\quad$ Ho se rechaza si $Z \geq z_{1-\alpha}$

Ho: $\mu_1 = \mu_2$
H_1: $\mu_1 < \mu_2$ $\quad$ Ho se rechaza si $Z \leq z_{\alpha}$

Ejemplo 7.5: Se desea estudiar el efecto de dos dietas diferentes, en las cuales se reemplaza el 50% de la harina de pescado por harina exoesqueleto de camarón (Dieta 1) y por harina de vísceras de pescado (Dieta 2), sobre el incremento de peso de carpa koi (*Cyprinus carpio*, var. Koi). Para este fin, se dispuso de 22 peces, con peso promedio de 7,5 gramos, a los cuales se les alimentó durante dos meses con el alimento diario requerido, a 12 con la Dieta 1 y a los otros 10 con la Dieta 2. Al término del tiempo de cultivo, los animales fueron pesados y se calculó la ganancia de peso en gramos por cada grupo, obteniendo los siguientes resultados:

Dieta	**Ganancia de Peso (gramos)**											
Dieta 1	134	146	104	119	161	108	83	113	129	97	123	124
Dieta 2	70	118	101	85	107	132	86	90	95	94		

Con $\alpha = 0{,}05$, se prueba la hipótesis nula que, las dos poblaciones muestreadas, tienen un crecimiento idéntico, contra la alternativa de que las muestras provenientes de las dos poblaciones tienen un crecimiento diferente.

Ho: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

μ_1 = Harina de exoesqueleto de camarón
μ_2 = Harina de vísceras de pescado

Tal como se mencionó, el procedimiento consiste en ordenar las 22 observaciones en forma ascendente, según la magnitud, teniendo en cuenta la población a la que pertenece y se le asigna el rango a cada observación, como sigue:

Población	2	1	2	2	2	2	2	1	2	1	2
Ganancia de Peso	70	83	85	86	90	94	95	97	101	104	107
Rango	1	2	3	4	5	6	7	8	9	10	11
Población	1	1	2	1	1	1	1	2	1	1	1
Ganancia de Peso	108	113	118	119	123	124	129	132	134	146	161
Rango	12	13	14	15	16	17	18	19	20	21	22

Para obtener la suma de los rangos, se procede a seleccionar la muestra de la Población 1 (Dieta 1) para asignar a R_1, de esta manera:

$R_1 = 2 + 8 + 10 + 12 + 13 + 15 + 16 + 17 + 18 + 20 + 21 + 22 = 174$

Con esto, el estadístico U de Mann-Whitney, aplicando (7.5) es:

$$U = n_1 + n_2 + \frac{n_1(n_1 + n_2)}{2} - R_1$$
$$= 12 \times 10 + \frac{12 \times 13}{2} - 174$$
$$= 120 + 78 - 178$$

Aplicando (7.6) se tiene:

$$E(U) = \mu_U = \frac{n_1\, n_2}{2} = \frac{12 \times 10}{2} = 60$$

Aplicando (7.2.3) se tiene:

$$V(U) = \sigma_U^2 = \frac{n_1\, n_2(n_1 + n_2 + 1)}{12} = 230$$

Por lo tanto:

$$Z = \frac{U - \mu_U}{\sqrt{\sigma_U^2}} = \frac{24 - 60}{\sqrt{230}} = -2{,}64$$

Para α = 0,05; 1 – $\alpha/2$ = 0,975, entonces, $Z_{0,975}$ =1,96 y $Z_{\alpha/2}$ = $Z_{0,025}$ = -1,96. Con base en esto, dado que -2,64 < -1,96, *H*o se rechaza y se concluye que, las dos poblaciones tienen diferencias significativas, es decir, el efecto de las dos dietas presenta diferencias significativas sobre el incremento de peso de la carpa Koi, con el 95% de confiabilidad.

Prueba de Kruskal-Wallis para k Muestras Independientes

En el Análisis de Varianza de una vía, el interés radica en probar la hipótesis nula *H*o: $\mu_1 = \mu_2 = \ldots = \mu_k$, donde las k muestras son independientes y provienen de poblaciones normales.

La Prueba No Paramétrica equivalente es la de Kruskal-Wallis, en la cual se supone que, los datos tienen por lo menos una escala ordinal y las poblaciones de donde provienen las muestras son continuas.

Como en la prueba de *U*, los datos se ordenan conjuntamente en forma ascendente, como si constituyeran una sola muestra; cuando ocurre un empate en la secuencia ordenada, se sugiere asignar el promedio de los rangos a las

observaciones, para las cuales ocurre el empate, de la misma manera como se realizó en las pruebas estudiadas en las secciones anteriores.

Para ello, sea R_i la suma de los rangos de los valores de la i-ésima muestra, la prueba se basa en el estadístico:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1) \qquad (7.8)$$

Donde $n = n_1 + n_2 + \ldots + n_k$ y k es el número de poblaciones muestreadas.

El estadístico de la prueba H se encuentra aproximadamente adecuado para una Distribución Chi-Cuadrada con n -1 grados de libertad.

De acuerdo con Canavos (1986), la aproximación Chi-Cuadrada es adecuada, excepto cuando k=3 y ninguno de los tamaños de la muestra es mayor que 5.

Así, dadas las hipótesis:

Ho: $\mu_1 = \mu_2 = \ldots = \mu_k$
H_1: por lo menos dos Medias presentan diferencias significativas,
El estadístico de prueba $H \sim \chi^2_{n-1}$ y H_0 se rechaza si $H \geq \chi^2_{1-\alpha;\, n-1}$.

Ejemplo 7.6: Se desea evaluar el efecto de tres cepas inoculantes de *Rizhobium leguminosarum* sobre la fijación biológica de nitrógeno de plantas de trébol blanco (*Trifolium repens*). Para ello, en un invernadero, se dispone de 18 macetas de esta especie vegetal, de las cuales, son asignadas al azar 6 macetas para cada una de las cepas y se procede a inocular. Los resultados del experimento se miden en miligramos de nitrógeno en las plantas, como sigue:

Cepa 1		Cepa 2		Cepa 3	
Rango	**N (mg)**	**Rango**	**N (mg)**	**Rango**	**N (mg)**
11	19,4	1	9,1	5	16,9
14	27,9	3	11,9	6	17,3
16	32,1	4	15,8	10	19,2
17	32,6	8	17,8	12	19,5
18	33,8	9	18,2	13	20,3
15	30,0	2	10,3	7	17,4

Para comparar el efecto, con α = 0,05, se utiliza la Prueba de Kruskal-Wallis, para probar la hipótesis:

H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: Por lo menos una de las Medias tiene un valor significativamente diferente.

Para los cálculos respectivos, se parte de los datos anteriores, donde se asignó le rango al frente de cada resultado. De esta manera se tiene que $n_1 = n_2 = n_3 = 6$, entonces $n = 18$.

La suma de los rangos de cada muestra es:

$R_1 = 11 + 14 + 16 + 17 + 18 + 15 = 91$

$R_2 = 1 + 3 + 4 + 8 + 9 + 2 = 27$

$R_3 = 5 + 6 + 10 + 12 + 13 + 17 = 53$

En estas condiciones, el estadístico H se calcula de la siguiente manera, según (7.8):

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$$

$$= \frac{12}{18 \times 19} \left[\frac{(91)^2}{6} + \frac{(27)^2}{6} + \frac{(53)^2}{6} \right] - 3 \times (19)$$

$$= 69{,}12 - 57 = 12{,}12$$

Ahora se procede a consultar la Tabla de Probabilidad de Chi-Cuadrado (Tabla 6 del Apéndice), para 1 - α = 0,95 y $k = \nu$ = 3-1 = 2 grados de libertad, donde el valor crítico es 5,99, es decir:

$$\chi^2_{1-\alpha;\, k-1} = \chi^2_{0{,}95;\, 2} = 5{,}99$$

Dado que H = 12,12 > 5,99, se rechaza H_0, o sea que existe por lo menos una Media con diferencia significativa, con el 95% de confiabilidad, donde el efecto de la cepa probada sobre el contenido de nitrógeno de las plantas es diferente.

CAPÍTULO 8

Principios básicos de las relaciones entre variables

ASPECTOS GENERALES

En la Producción y en la Salud Animal, establecer sí dos o más variables tienen algún tipo de relación, ayuda a tomar determinaciones técnicas importantes. Por ejemplo, al planear el manejo alimenticio de un hato debe establecerse sí el consumo de alimento, en materia seca o fresca, está determinado por el peso vivo de los animales, el contenido graso o de fibra de las raciones; en el Mejoramiento Genético de las especies productivas, es fundamental saber, si dos o más rasgos de importancia en la producción, están o no relacionados, es el caso de la grasa y la proteína de la leche; en Nutrición Animal, es importante determinar sí la calidad nutritiva de un forraje, depende de la edad al corte del pasto; en las Explotaciones Porcinas, es fundamental establecer sí la composición de la dieta afecta el espesor de la grasa dorsal de los animales destinados al sacrificio; en la Administración de Medicamentos, es imprescindible conocer si la dosis aplicada de un determinado fármaco, afecta la composición del producto final, como carne o leche e incluso el volumen de producción; en la Alimentación Animal debe comprobarse sí la conversión alimenticia está influenciada por la edad del animal; en el Manejo Reproductivo, el peso de la hembra al primer servicio, puede ser una variable que influya en el peso de las crías o el tamaño de la camada, lo mismo puede

suceder con la edad de los parentales; en Piscicultura, el peso y la talla son variables que guardan algún tipo de relación; como los anteriores, podrían citarse múltiples situaciones similares.

A continuación, se presentan los procedimientos estadísticos, que permiten confirmar o no la presencia de relación funcional entre variables y el tipo de dicha relación.

RELACIONES FUNCIONALES DE DEPENDENCIA

En este caso específico, se busca definir si una variable ***aleatoria,*** que en adelante se denominará ***Y,*** o dependiente, tiene una ***relación funcional*** de dependencia con una o más variables, la cual, en adelante se denominarán ***X***, o independiente. El objetivo de los análisis de esta naturaleza consiste en desarrollar una ecuación que permita estimar el ***valor medio*** de *Y*, conociendo los valores de *X*. Cuando el interés se centra en este tipo de problemas, se hace referencia a la ***Regresión***.

En los Análisis de Regresión pueden presentarse varias situaciones: la primera que se menciona en este texto, se refiere al cambio observado en ***una*** variable aleatoria ***Y*** por efecto o influencia de ***una*** variable ***X,*** pero además, el cambio de *Y* puede ser lineal, exponencial, potencial, etc. Cuando el cambio observado en *Y* se explica por el efecto de una sola variable *X*, se hablará de Regresión Simple.

REGRESIÓN LINEAL SIMPLE

Cuando el cambio observado en *Y*, por efecto del cambio en *X*, se explica de una forma tal que se puede representar en una línea recta, se habla de Regresión Lineal Simple. Para encontrar la ecuación que relacione las variables, por lo general, se inicia con la recolección de los datos que toma *Y*

(kg de proteína/kg de materia verde) cuando varían los valores de *X* (Edad de la planta, días) (Tabla 8.1), luego se marcan los puntos X_i, Y_i sobre un sistema de coordenadas y se obtiene, lo que en estadística, se denomina ***Nube de Puntos o Diagramas de Dispersión.*** A partir de esta figura es posible observar una curva denominada ***Aproximante,*** que sugiere el tipo de relación existente**.** Como estrategia didáctica, en la Figura 8.2 se presenta el Diagrama de Dispersión, correspondiente a la variación observada en el contenido proteico del pasto kikuyo (*Penisetum clandestinum*), por efecto de la variación en la edad de corte de la planta, expresada en días.

Los datos de la Tabla 8.1 representan la relación de asociación entre la edad de la planta y el contenido de proteína del *Penisetum clandestinum*.

Tabla 8.1
Edad de la planta y contenido de proteína del pasto kikuyo (*Penisetum clandestinum*)

***X*: Edad de la planta (días)**	***Y*: kg de proteína/kg de materia verde**
5	0,10
9	0,20
12	0,30
15	0,40
18	0,50
20	0,70
25	0,72
30	0,85
35	0,90
40	1,02

Como se aprecia, la nube de puntos (Figura 8.1) sugiere una relación lineal entre las variables *X, Y*; el trazado de una línea punteada confirma con mayor claridad este hecho.

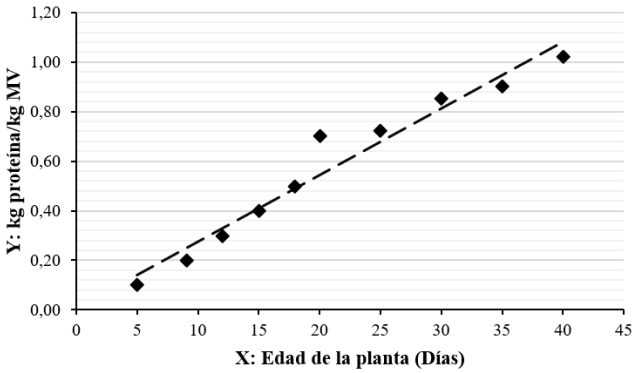


Figura 8.1 Diagrama de Dispersión de los datos de la Tabla 8.1.

Una vez establecida la existencia de dependencia lineal, se procede a encontrar una ecuación que describa el comportamiento de *Y* por efecto de la acción de *X*; para alcanzar este objetivo es necesario estimar dos parámetros denominados β_0 y β_1.

El primer término representa el intercepto, es decir, el punto de corte de la recta aproximante con el eje ***Y***; mientras que el segundo indica la pendiente de la recta. Para efectos de conclusiones y recomendaciones prácticas, β_0 indica el valor medio de *Y* cuando *X* tiene un valor igual a cero, valor que no siempre tendrá una explicación biológica real; β_1 deberá interpretarse como ***el cambio promedio de Y***, por cada incremento de *X* en una unidad. Definidos los dos parámetros anteriores, es posible describir el valor de *Y,* con la siguiente ecuación:

$$Y_i = \beta_0 + \beta_1 X_i \qquad (8.1)$$

Donde:

Y_i = Valor que toma la variable aleatoria en estudio o también denominada ***Variable Dependiente.***

β_0 y β_1 = Parámetros cuyo significado ya fue explicado.

X_i = Valor que toma X, ***variable no aleatoria***, denominada ***Variable Independiente, Predictora o Regresora.***

Con respecto a los valores de la variable independiente X, es conveniente advertir que, éstos pueden seleccionarse antes de desarrollar la ecuación, como en el caso donde la temperatura se fija en intervalos de 2Cº, para observar el consumo de alimento en un plantel avícola, de tal modo que, el primer valor de X es 20Cº y luego se incrementa a 22, 24, 26Cº, etc., hasta llegar al máximo biológico establecido por el investigador, de acuerdo con el comportamiento típico de la especie.

En este caso, es claro que el estudio de la regresión ***es controlado***, sin embargo, cuando los valores de X se escogen entre una gama de posibles valores, el estudio de la regresión es ***observacional***, como en el caso de los análisis bromatológicos, donde los valores que toma X pueden ser las cifras de humedad de una planta, cuando ésta se recolecta para llevar a cabo los análisis.

Así, en los estudios, bien sean controlados u observacionales, el propósito del Análisis de Regresión Lineal Simple será idéntico en los dos casos, es decir, encontrar la recta que brinde el ***mejor ajuste*** del comportamiento de las observaciones X, Y.

Para mayor claridad de los procedimientos, que se indicarán más adelante, conviene hacer las siguientes precisiones: Al observar los puntos trazados en la Figura 8.1, es fácil notar que, no todos los puntos se ubican exactamente sobre la

recta, algunos tienen un desvío positivo, otros negativos y otros cero, pero en todo caso, a cada punto (X_i, Y_i) le corresponderá un valor ε_i y, para todo el conjunto de datos, se tendrá valores ε_1, ε_2, ε_3, ε_4, ..., ε_n. En adelante, estas desviaciones se denominarán ***Error o Residuo***, aclarando que los ε_i son aleatorios.

Una medida, que podría indicar la calidad del ajuste de la curva, es la Sumatoria de los ε_i al cuadrado ($\sum \varepsilon_i^2$), ya que entre más pequeño sea este valor, mejor será el ajuste, así que, en la práctica, lo que realmente se busca es que esta expresión sea mínima para obtener los mejores estimadores de β_0 y β_1. Una recta que ajuste los datos, con la condición antes citada, se denomina ***Recta de Mínimos Cuadrados***. Para mayor claridad, se presentan a continuación, algunos aspectos importantes, para la estimación de los citados parámetros:

a) Considerar el modelo de la Regresión Lineal Simple:

$$Y_i = \beta_0 + \beta_1 X_i . \qquad (8.2)$$

b) Como ya se dijo, las Xi se conocen a priori; β_0 y β_1 son parámetros desconocidos. Como se mencionó antes, la ecuación (8.2.1), expresa el valor medio de Y, así que se podría expresar de la siguiente manera:

$$E(Y_i) = \beta_0 + \beta_1 X_i \qquad (8.3)$$

c) Si se consideran los residuos ε_i, como valores aleatorios con distribución $N(0, \sigma^2)$, es posible escribir esta expresión como:

$$Y_i = E(Yi) + \varepsilon_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (8.4)$$

De (8.2.3) es posible despejar $\varepsilon_i = Y_i - \beta_0 - \beta_1 X_i$ (8.5)

d) Se dijo que esta es una expresión válida para cada valor de *Y*; así que, si se consideran todos los valores de *Y*, incluidos en el análisis se tendrá:

$Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1$
$Y_2 = \beta_0 + \beta_1 X_2 + \varepsilon_2$
.
.
.
$Y_n = \beta_0 + \beta_1 X_n + \varepsilon_n$

De cada una de las anteriores expresiones se despeja ε_i

$$\varepsilon_i = Y_i - \beta_0 - \beta_1 X_i,\ i = 1, 2, \ldots, n$$

La suma de los cuadrados de las desviaciones de las observaciones con respecto a la recta es:

$$Q = \sum \varepsilon_i^2 = \sum (Y_i - \beta_o - \beta_1 X_1)^2 \qquad (8.6)$$

Esta es una función de los parámetros desconocidos β_0 y β_1. Los estimadores de mínimos cuadrados $\hat{\beta}_0$ y $\hat{\beta}_1$ de β_0 y β_1, respectivamente, se obtienen mediante la diferenciación de (8.6) con respecto a β_0 y β_1 y después al igualar cada derivada parcial a cero, así:

$$\frac{\partial Q}{\partial \beta_0} = 2 \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)(-1)$$

$$\frac{\partial Q}{\partial \beta_1} = 2 \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)(-X_i)$$

Después de simplificar se obtiene:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i = \sum Y_i \qquad (8.7)$$
$$\hat{\beta}_1 \sum X_i + \hat{\beta}_1 \sum X_i^2 = \sum X_i Y_i$$

Estas ecuaciones reciben el nombre de ***Ecuaciones Normales de Mínimos Cuadrados***.

Resolviendo el sistema de ecuaciones, se obtiene:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \frac{\sum Y_i}{n} - \hat{\beta}_1 \frac{\sum X_i}{n} \qquad (8.8)$$

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} \qquad (8.9)$$

Por lo tanto, la ecuación de regresión estimada para el modelo (8.4) es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad (8.10)$$

Ejemplo 8.1 Un investigador, en Nutrición Animal, está interesado en establecer sí el contenido de fibra de la dieta, expresado en kg/100 kg de alimento, afecta el consumo voluntario en conejos Nueva Zelanda. Para alcanzar su objetivo, prepara diferentes dietas con diversos contenidos de fibra y suministra las diferentes raciones a varios grupos de conejos, homogéneos en cuanto a edad, sexo, raza y peso, encontrando los siguientes resultados:

Fibra kg/100 kg de ración (*X*)	Consumo Voluntario (kg/día) (*Y*)
19,0	0,30
20,3	0,23
20,0	0,24
20,1	0,26
19,7	0,29
20,4	0,22
23,1	0,16
23,0	0,20
22,5	0,18
22,8	0,19
22,0	0,19
21,0	0,22

Para establecer la relación funcional de dependencia entre la fibra de la dieta (*X*) y el consumo voluntario (*Y*), en primer lugar, se construye el Diagrama de Dispersión, que se muestra en la Figura 8.2.

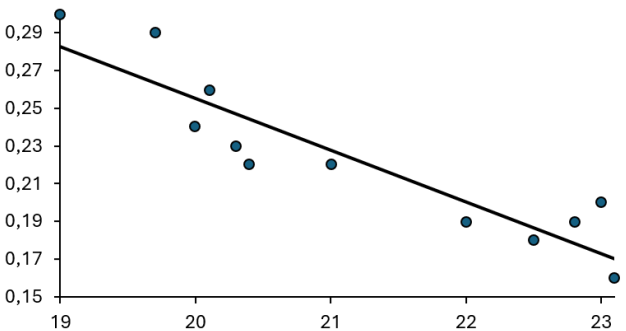


Figura 8.2 Diagrama de Dispersión de las variables Fibra en la Dieta (*X*) y Consumo Voluntario (*Y*)

Puesto que, la Figura 8.2, sugiere una relación lineal, se procede a establecer la recta de mejor ajuste y, para esto, se necesitará estimar los valores de β_0 y β_1, las cifras necesarias para su cálculo se obtienen de la siguiente manera:

$\bar{x} = 21{,}1583 \quad \bar{y} = 0{,}2233$
$\Sigma(x_i - \bar{x})^2 = 22{,}9492$
$\Sigma(y_i - \bar{y})^2 = 0{,}02067$
$\Sigma\ (x_i - \bar{x})(y_i - \bar{y}) = -0{,}63033$

Con las anteriores cantidades, es posible calcular los estimadores de los parámetros, aplicando (8.8) y (8.9), así:

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(x_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{-0{,}63033}{22{,}9492} = -0{,}02745$$

$$\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x} = 0{,}2233 - (-0{,}02745 \times 21{,}1583) = 0{,}804$$

Según esto, el modelo estimado de mejor ajuste de la recta de regresión, de acuerdo con (8.2.10) sería:

$$\hat{y} = 0{,}804 - 0{,}02745x$$

$\hat{\beta}_1$ en la recta representa la pendiente. La interpretación, aplicada de este valor, indica que el consumo de alimento en los conejos Nueva Zelanda, en promedio disminuye en 0,02745 kg, por cada incremento de un kilogramo de fibra en la dieta.

$\hat{\beta}_0$ representa el punto de corte o intercepto con el eje *Y*. Podría interpretarse como el valor del consumo, cuando la dieta no tiene

fibra, que en este caso. Tendría explicación lógica desde el punto de vista aplicado.

Conocidos los valores de los estimadores, ahora es posible hacer uso de una de las importantes Propiedades de la Regresión, que consiste en predecir el valor medio de *Y* para cualquier valor de *X*, siempre y cuando éste se encuentre en el rango de valores utilizados para calcular los estimadores de los parámetros. Así, si se desea estimar el valor medio de y cuando la dieta de la fibra se fije en 20,5 kg, basta con aplicar el siguiente procedimiento:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_i = 0{,}804 + (-0{,}02745 \times 20{,}5) = 0{,}242$$

Se puede concluir que, el consumo medio de alimento, será de 0,242 kg para la cantidad de fibra indicada. No se podrá utilizar la ecuación de predicción para valores de fibra menores de 19 o mayores de 23,1 kg.

El paso siguiente consiste en llevar a cabo la Prueba de Hipótesis para la Regresión. El procedimiento general de Prueba de Hipótesis no varía, con referencia al indicado para la Prueba de Medias o proporciones, con algunas particularidades específicas. El propósito de la prueba pretende generalizar el comportamiento de las variables deducido con los datos de la muestra, para la población de donde se extrajo esa muestra.

Para ello, es necesario hacer una consideración previa, antes de realizar el Análisis de Varianza. En la Figura 8.3, para cada punto observado y_i se verifica que:

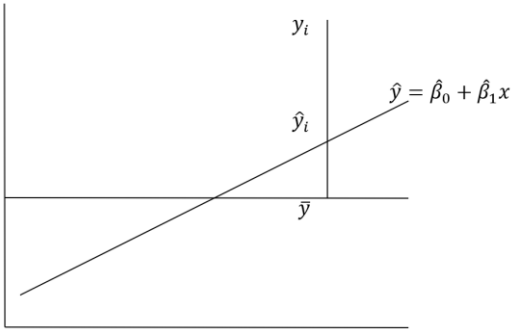


Figura 8.3 Ecuación de Regresión Lineal con verificación en los puntos de intersección.

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$

Siendo $\hat{y}_i$ el valor estimado de *Y* en dicha observación y $\bar{y}$ la media muestral de la variable *Y*, se demuestra que:

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2 \qquad (8.11)$$

Donde:

$\sum(y_i - \bar{y})^2$: Se denomina ***Suma de Cuadrados Total*** *(SCT)* y representa la variación total de las observaciones respecto a su media.

$\sum(\hat{y}i - \bar{y})^2$: Se denomina ***Suma de Cuadrados debido a la Regresión*** *(SCR)* y representa la variación de las observaciones que es atribuible al efecto de *X* sobre *Y*.

$\sum(y_i - \hat{y}_i)^2$: Se denomina ***Suma de Cuadrados debido al Error*** *(SCE)* y representa la variación de las observaciones con respecto a la recta de regresión estimada.

Reemplazando en (8.11), se tiene que:

$$SCT = SCR + SCE$$

$$SCT = \sum(y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$SCR = \sum(\hat{y}_i - \bar{y})^2 = (\hat{\beta}_1)^2[\sum(x_i - \bar{x})^2] = \hat{\beta}_1^2\left[\sum x_1^2 - \frac{(\sum x_i)^2}{n}\right]$$

$$SCE = SCT - SCR$$

Prueba de Hipótesis sobre la Significancia del Modelo de Regresión

Permite tomar una decisión global, con respecto a la aceptación o rechazo del modelo. Para esto puede utilizarse el Análisis de Varianza.

La hipótesis se plantea del siguiente modo:

H_0: $\beta_1 = 0$ X e Y no están relacionadas linealmente.

H_1: $\beta_1 \neq 0$ X e Y están relacionadas linealmente.

La tabla de ***Análisis de Varianza*** (*ANOVA*), para la Regresión Lineal Simple, se visualiiza en la Tabla 8.2.

Tabla 8.2

Tabla de ANOVA para la regresión lineal simple.

Fuentes de variación	**GL**	**Suma de cuadrados**	**Cuadrado medio**	***Fc***	***Ft***
Regresión	1	SCR	$CMR = \frac{SCR}{1}$	$\frac{CMR}{CME}$	
Error	n - 2	SCE	$CME = \frac{SCE}{n-2}$		
Total	n - 1	SCT			

Donde: CMR se denomina Cuadrado Medio de la Regresión.

CME se denomina Cuadrado Medio del Error.

$$\frac{CMR}{CME} = Fc$$

Tiene una distribución *F* con $n - 1$ y $n - 2$ grados de libertad (GL).
H_0 se rechaza si $Fc \geq Ft$, es decir, si $Fc \geq f_{1-\alpha,\ n-1,\ n-2}$

Ejemplo 8.2: A continuación, se presenta los resultados del Análisis de Varianza para los datos incluidos en el Ejemplo 8.1.

Fuentes de variación	**GL**	**SC**	**CM**	**Fc**	**Ft**
Regresión	1	0,01731	0,01731	51,51	10,04 *
Error	10	0,00336	0,000336		
Total	11	0,02067			

Los resultados que se presenta en la tabla anterior, se obtienen a partir de la aplicación de las fórmulas correspondientes, derivadas de (8.11), de la siguiente manera:

$SCR = (\hat{\beta}_1)^2 \sum(\hat{y}_i - \bar{y})^2 = (-0{,}02745)^2 \times 22{,}9492 = 0{,}01731$

$SCT = \sum(Y_i - \bar{Y})^2 = 0{,}02067$

$SCE = SCT - SCR = 0{,}00336$

$CMR = \frac{SCR}{1} = 0{,}01731$

$CME = \frac{SCE}{10} = 0{,}000336$

$$Fc = \frac{CMR}{CME} = \frac{0{,}01731}{0{,}000336} = 51{,}51$$

Si α = 0,01, entonces se obtiene los valores de la Tabla de Probabilidad de *F* de Fisher (Tabla 8 del Apéndice): $f_{1\text{-}\alpha,\ 1,\ n\text{-}2} = f_{0{,}99,\ 1,\ 10} = 10{,}04$.

Dado que $Fc = 51{,}51$ es mayor que $Ft = 10{,}04$, se rechaza *H*o y se concluye, con el 99% de confiabilidad, que existe relación lineal entre la fibra de la dieta y el consumo de alimento.

Coeficiente de Determinación

Otra medida que es necesario estimar, al realizar Análisis de Regresión, es el ***Coeficiente de Determinación*** R^2, que es una medida relativa del grado de asociación lineal entre las dos variables tanto para *X* como para *Y*, el cual viene dado por:

$$R^2 = \frac{Suma\ cuadrado\ de\ la\ regresión\ (SCR)}{Suma\ cuadrado\ total\ (SCT)} \qquad (8.12)$$

Este coeficiente mide la proporción de la variación observada en *Y*, que se explica por efecto de *X*. Los valores que toma están en el intervalo $0 \leq R^2 \leq 1$, entre más cercano de 1 se encuentre el valor mayor será el grado de asociación lineal entre *X* e *Y*. Es frecuente expresar a R^2 en porcentajes. En modelos biológicos, este valor se considera bueno cuando está por encima del 65%.

Ejemplo 8.3: Para el Ejemplo 8.2, se tiene que:

Aplicando la fórmula (8.12):

$$R^2 = \frac{SCR}{SCT} = \frac{0{,}01731}{0{,}02067} = 0{,}8374$$

Expresándolo como porcentaje: R^2 = 83,74%.

Esto significa que el 83,74% de la variación observada en el consumo, se explica por la variación en el contenido de fibra de la dieta.

Límites de Confianza

También es conveniente expresar el valor de β_1 como un límite de confianza, tal como se indicó en los casos de la Media y la Proporción. El procedimiento es básicamente igual, sólo que, en el caso de la Regresión, se debe calcular la expresión $S_{\hat{\beta}1}$ o error estándar de $\hat{\beta}_1$, cuyo cálculo se obtiene así:

$$S^2_{\hat{\beta}1} = \frac{CME}{\Sigma(Xi-\bar{X})^2} \qquad (8.13)$$

Un intervalo de confianza para β_1 al (1 - α)100% de confiabilidad viene dado por:

$$\hat{\beta}_1 \pm S_{\hat{\beta}_1} \times t_{1-\frac{\alpha}{2};\, n-2} \qquad (8.14)$$

Ejemplo 8.3: Para el Ejemplo 8.3, un intervalo de confianza para β_1 al 95% de confiabilidad, mediante la aplicación de (8.14), se obtendría de la siguiente manera:

$$\hat{\beta}_1 \pm S_{\hat{\beta}_1}\, t_{0,975,\,10}$$

$$S_{\hat{\beta}1} = \sqrt{\frac{CME}{\Sigma(Xi-\bar{X})^2}} = \sqrt{\frac{0,000336}{22,9492}} = 0,00302$$

$$t_{0,975;10} = 2,228$$

Reemplazando se obtiene: -0,02745 ± 0,00382 × 2,228

En consecuencia, se tiene que β_1 está comprendido entre -0,035 y -0,019, con una confiabilidad del 95%, lo cual significa que, en este experimento, el promedio del contenido de la dieta puede variar entre estos valores.

Es de interés construir un intervalo de confianza, a partir de la respuesta media de un valor específico de *X*, por ejemplo x_p.

Un intervalo de confianza al (1-α)100% de confiabilidad alrededor de la respuesta media $E(y_p)$ está dado por:

$$\hat{y}_p \pm t_{1-\frac{\alpha}{2};\, n-2} \times\ S(\hat{y}_p) \qquad (8.15)$$

Donde $\hat{y}_p$ se lo calcula a partir del modelo de Regresión ajustado $y_p = \hat{\beta}_0 + \hat{\beta}_1 x$, y la Varianza:

$$S^2(\hat{y}_p) = CME\left[\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right] \qquad (8.16)$$

Ejemplo 8.5: En el Ejemplo 8.3, que se viene analizando en este capítulo, se supone que, se desea construir un intervalo de confianza al 95% de confiabilidad, para la media de y_p en x_p = 20,5 kg.

El valor estimado es $\hat{y}_p$= 0,804 + (-0,02745 × 20,5) = 0,242 y la desviación estándar estimada, según (8.2.14), es:

$$S^2(\hat{Y}_p) = \left\{0{,}000336\left[\frac{1}{12} + \frac{(20{,}5 - 21{,}158)^2}{22{,}9492}\right]\right\}^{1/2}$$
$$= 0{,}005859$$

y $t_{0{,}975;10} = 2{,}228$

Por lo tanto, un intervalo de confianza para la media de y_p, $E(y_p)$ es

$0{,}242 \pm 2{,}228 \times 0{,}005859 = (0{,}229 < y < 0{,}2550)$.

Esto indica, que la media del consumo de alimento, en función del contenido de fibra, puede variar entre 0,229 y 0,2550 kg por día.

Un intervalo de confianza de $E(y_p)$ se refiere a la verdadera respuesta promedio en $X = x_p$ (esto es, el parámetro de la población) y no a observaciones futuras.

Una aplicación importante, en un modelo de Regresión, es la predicción de observaciones nuevas o futura de Y, si x_0 es el valor que se quiere pronosticar, entonces:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

Es el estimador puntual del valor futuro de la respuesta y_0.

Un intervalo de predicción para una observación futura y_0 en el valor x_0 al (1-α)100% de confiabilidad, acorde con (8.16), está dado por:

$$\hat{y}_0 \pm t_{1-\frac{\alpha}{2};\, n-2} \times\ S(\hat{y}_0)$$

Donde: $S^2(\hat{y}_0) = CME\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right]$ tal como se indicó en (8.16).

Ejemplo 8.6: En el mismo Ejemplo 8.5, se quiere construir un intervalo de predicción al 95% de confiabilidad para la predicción y_0 en el valor x = 20,5 kg.

El valor estimado es $\hat{y}_0$ = 0,804 + (-0,02745 × 20,5) = 0,242, y la desviación estándar:

$$S^2(\hat{y}_0) = \left\{0{,}000336\left[1 + \frac{1}{12} + \frac{(20{,}5 - 21{,}158)^2}{22{,}9492}\right]\right\}^{1/2} = 0{,}0192$$

$t_{1\text{-}\alpha/2;\, n\text{-}2} = t_{0{,}975;\, 10} = 2{,}228$, por lo tanto, el intervalo es:

0,242 ± 2,228 × 0,0192

(0,199 < y < 0,285).

Con una confiabilidad del 95%, el pronóstico en y_x = 20,5 para x = 20,5 está comprendido entre 0,199 y 0,285.

CORRELACIÓN LINEAL

Cuando dos o más variables de interés son aleatorias y ninguna de ellas se considera dependiente de la otra, en este caso, se considera que, existe una relación de asociación, en cuyo caso, el Análisis de Correlación permite establecer sí las variables en cuestión están o no asociadas.

Si se analizan únicamente dos variables, la Correlación será Simple y, cuando se involucra a más de dos variables, el estudio de Correlación será Múltiple. El grado de asociación lineal entre dos o más variables está determinado por el Coeficiente de Correlación de Pearson o Correlación Producto Momento y su forma de cálculo se expresa así:

$$\rho = \frac{\frac{\Sigma(X_i-\mu_x)(Y_i-\mu_y)}{N}}{\sqrt{\frac{\sum(X_i-\mu_x)^2}{N}}\times\sqrt{\frac{\sum(X_i-\mu_x)^2}{N}}} = \frac{Covarianza\ (X,Y)}{\sqrt{Varianza\ X}\times\sqrt{Varianza\ Y}} \qquad (8.17)$$

$$\rho = \frac{\Sigma(X_i-\mu_x)(Y_i-\mu_y)}{\sqrt{\sum(X_i-\mu_x)^2}\times\sqrt{\sum(X_i-\mu_x)^2}} \qquad (8.18)$$

El estadígrafo de ρ se denota por la letra *r* y se calcula del mismo modo, sólo que en lugar de los parámetros μ y σ se utilizan los estadígrafos correspondientes.

El cálculo de la correlación muestral, esto es el estadígrafo, equivalentes con (8.17) se expresa de la siguiente manera:

$$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\times\sqrt{\sum(y_i-\bar{y})^2}} \qquad (8.19)$$

Una fórmula de trabajo, equivalente con (8.19) se expresa así:

$$r = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}\times\sqrt{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}}$$

El Coeficiente de Correlación puede tomar valores entre –1 y 1 incluyendo el cero. Un valor negativo indica asociación inversa, en un modo tal que, mientras una variable aumenta la otra disminuye; un valor de cero indica que, las dos variables son independientes, no están asociadas linealmente o no existe correlación lineal y, un valor positivo sugiere que, cuando una variable aumenta la otra también lo hace.

Establecer si dos o más variables están correlacionadas es un factor importante en muchos aspectos de la Producción Animal, especialmente en

los Programas de Mejoramiento Genético, puesto que, si se decide mejorar por selección una característica de importancia económica, se debe conocer si al mejorar un rasgo se desmejora otro, caso en el cual, la selección no tendría un efecto positivo. En cambio, cuando dos rasgos tienen correlación positiva, al mejorar uno también, de modo simultáneo, se mejora el otro, logrando así un impacto positivo en los planes de mejora.

Ejemplo 8.7: A continuación, se presenta un ejemplo con datos reales medidos en 1.999 y utilizados por Cerón et al (2.003), donde se presentan las cifras de producción de leche por lactancia ajustada a 305 días, expresada en kilogramos y el puntaje total por tipo (Tabla 8.2). El objetivo del análisis consiste en establecer sí estas dos variables aleatorias están correlacionadas.

Tabla 8.2

Producción de leche por lactancia, ajustada a 305 días, en kilogramos, de vacas en Colombia (2003).

Puntos (*X*)	Leche (*Y*)	Puntos (*X*)	Leche (*Y*)
81	16610	84	12310
84	15460	87	12310
82	15240	77	12310
80	15020	79	12250
92	14260	84	12240
90	14090	87	12240
81	14010	88	12240
82	13820	86	12210
81	13780	84	12210
89	13680	82	12140
78	13530	80	12130
77	13270	85	12130
78	13230	83	12130

Puntos (*X*)	**Leche (*Y*)**	**Puntos (*X*)**	**Leche (*Y*)**
81	13000	84	12100
86	12900	80	12100
90	12730	81	12080
84	12570	82	12040
80	12550	82	12030
88	12510	82	12020
79	12510	81	11970
85	12500	79	11970
84	12490	85	11900
89	12390	82	11890
80	12390	90	11870
84	12330	85	11870

En los planes de Mejoramiento de los Bovinos Lecheros, se ha discutido ampliamente, si la morfología de los animales es un aspecto relevante para alcanzar los objetivos de producción. Uno de los aspectos importantes, para dirimir la discusión, consiste en establecer matemáticamente si hay relación funcional de asociación entre el puntaje y la producción. Sí esta es positiva, podría diseñarse un plan de mejoramiento que busque mejorar simultáneamente los dos rasgos.

Las cantidades necesaria para calcular el Coeficiente de Correlación muestral *r* son las siguientes:

$$\sum (x_i - \bar{x})^2 = 666{,}08$$

$$\sum (y_i - \bar{y})^2 = 5{,}499003 \times 10^7$$

$$\sum(x_i - \bar{x})(y_2 - \bar{y}) = -6546{,}8$$

De acuerdo con los anteriores valores, el Coeficiente de Correlación se calcula de la siguiente manera:

$$r = \frac{-646{,}8}{\sqrt{666{,}08} \times \sqrt{5{,}4999003 \times 10^7}} = -0{,}00337$$

Una Prueba de Hipótesis especial, respecto al Coeficiente de Correlación Poblacional $\boldsymbol{\rho}$, para este ejemplo, se plantea como:

H_0: $\rho = 0$. El puntaje por tipo y la producción de leche no tienen correlación lineal.

H_1: $\rho \neq 0$. El puntaje por tipo y la producción de leche tienen correlación lineal.

El estadístico de prueba es:

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

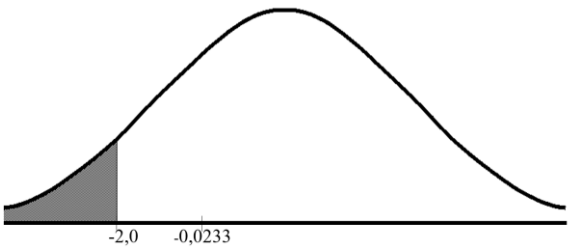
y tiene una distribución ***t*** con *n*-1 grados de libertad.

H_0 se rechaza si: $T > t_{1-\alpha/2}$ o $T < t_{\alpha/2;\, n-2}$

El cálculo del estadístico de prueba es:

$$Tc = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{-0{,}00337\sqrt{48}}{\sqrt{1-(-0{,}00337)^2}} = -0{,}0233$$

En la Tabla 7 del Apéndice, para α = 0,05, se tiene que $t_{\alpha/2;\, n-2}$ = $t_{0,025;\, 48}$ = -2,0. Debido a que -0,0233 no es menor que 2,0, entonces, no existen elementos suficientes para rechazar H_0, debiendo aceptarla. En el siguiente esquema se representan estos valores.



REGRESIÓN LINEAL SIMPLE CON NOTACIÓN Y OPERACIONES MATRICIALES

En primer lugar, es necesario re-escribir el modelo de la siguiente forma:

$$Y = X\beta + \varepsilon. \qquad (8.20)$$

Donde $\boldsymbol{Y}$ representa el vector de observaciones, $\boldsymbol{X}$ es una matriz de parámetros observables, $\boldsymbol{\beta}$ es un vector de parámetros desconocidos y $\boldsymbol{\varepsilon}$ es un vector de errores aleatorios.

El procedimiento para obtener los valores de las incógnitas, es decir, los valores de beta o estimadores de los parámetros, que ajustan la línea de regresión, es el de mínimos cuadrados, de tal manera que:

$$\boldsymbol{\beta = (X'X)^{-1} X'Y} \qquad (8.21)$$

$$\text{X'X} = \begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma {X_i}^2 \end{bmatrix};\ \text{X'Y} = \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix};\ \beta = \begin{bmatrix} \beta_o \\ \beta_1 \end{bmatrix} \qquad (8.22)$$

Ejemplo 8.8: Retomando los datos del Ejemplo 8.2, el problema puede resolverse, con notación matricial, con la metodología que se detalla a continuación.

Reemplazando las expresiones por los valores obtenidos con los datos de la tabla asociada al citado ejemplo, aplicando las fórmulas descritas en (8.4.3), se tiene:

$$X'X = \begin{bmatrix} 12 & 253{,}9 \\ 253{,}9 & 5395{,}05 \end{bmatrix};\ X'Y = \begin{bmatrix} 2{,}68 \\ 56{,}074 \end{bmatrix};\ \beta = \begin{bmatrix} \beta_o \\ \beta_1 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 19{,}59058063 & -0{,}9219652129 \\ -0{,}9219652129 & 0{,}043574566977 \end{bmatrix} \begin{bmatrix} 2{,}68 \\ 56{,}074 \end{bmatrix} = \begin{bmatrix} 0{,}8044787392 \\ -0{,}027466502065 \end{bmatrix}$$

Como se puede observar, para encontrar los estimadores de los Beta, basta con invertir $\boldsymbol{X'X}$ y luego multiplicar dicha inversa por el vector $\boldsymbol{X'Y}$.

Para llevar a cabo el Análisis de Varianza se construye una tabla, tal como se hizo en el ejemplo anterior, sólo que, para realizar los cálculos en forma matricial, se desarrolla de la siguiente manera:

a) Se debe calcular el Factor de Corrección así: $FC = \frac{(\Sigma y_i)^2}{n} = n\bar{y}^2$

b) Calcular la Suma Cuadrado Total: $SCT = \boldsymbol{Y'Y} - FC = \sum y_i^2 - FC$

c) Calcular la Suma Cuadrado de la Regresión: $SCR = \beta' X'Y - FC$

d) Calcular la Suma Cuadrado del Error:

SCE = SC total corregida – SC de la regresión corregida.

Siguiendo los pasos para la construcción de la tabla ANOVA, se tendría:

FV	***GL***	***SC***	***CM***	***Fc***	***Ft***
Regresión (X)	K	$\hat{\beta}' X'Y$ **-** FC	SCR/GLR	CMR/CME	
Error	n – 1- K	$(Y'Y - FC)$ - $(\hat{\beta}'X' Y - FC)$	SCE/GLE		
Total	n - 1	$Y'Y - FC$			

Al reemplazar las expresiones indicadas en la tabla por sus correspondientes valores, se tiene:

$FC = 0{,}5985$

$Y'Y = 0{,}6192 - FC = 0{,}0207$

$\hat{\beta}'X'Y = [0{,}8044787392 \quad -0{,}027466502] \begin{bmatrix} 2{,}68 \\ 56{,}074 \end{bmatrix} = 0{,}6158463851$ – $FC = 0{,}017346$

$SCE = 0{,}027 - 0{,}017346 = 0{,}0003354$

Estos valores coinciden con los datos del Ejemplo 8.2 y, lógicamente, se tiene las mismas conclusiones, ya que lo único que varía es el procedimiento para el cálculo.

Para llevar a cabo las Pruebas de Hipótesis, utilizando procedimientos matriciales, es necesario calcular la Matriz de Varianzas y Covarianzas; para esto basta con multiplicar la matriz $\boldsymbol{(X'X)^{-1}}$ por el cuadrado medio del error. Par este ejemplo se tiene:

$$\begin{bmatrix} 19{,}59058063 & -0{,}9219652129 \\ -0{,}9219652129 & 0{,}04357456697 \end{bmatrix} \times CME$$

Para facilitar los cálculos posteriores, la matriz de $\boldsymbol{(X'X)^{-1}}$, en adelante se denominará Matriz de Coeficientes y se denotará por $\boldsymbol{C}$, de la siguiente manera:

$$C = \begin{bmatrix} C_{00} & C_{01} \\ C_{10} & C_{11} \end{bmatrix} \qquad (8.23)$$

Para llevar a cabo la Prueba de Hipótesis para β_0 se procede así:

H_0: $\beta_0 = 0$
H_1: $\beta_0 \neq 0$

Como ya se mencionó, esta prueba no tiene mayor importancia en experimentos biológicos, ya que la hipótesis nula indica, que el punto de corte de la Recta de Regresión es cero, lo cual generalmente no tiene interpretación.

Para probar la hipótesis, se contrasta el valor *T* calculado (*Tc*) con el valor *T* de la Tabla de Probabilidad (*Tt*). *Tc* se calcula así:

$$TC = \frac{\hat{\beta}_0}{\sqrt{C_{00}.CME}} = \frac{0{,}804478}{\sqrt{(19{,}59058)(0{,}0003354)}} = \frac{0{,}804478}{0{,}081} = 9{,}93$$

El valor tabular para una confiabilidad del 95% y 10 grados de libertad en el error, es 2,306 (Tabla 7 del Apéndice), por lo que se rechaza la hipótesis nula, dado que *Tc* (9,93) es mayor que *Tt* (2,306), lo cual no tiene una explicación lógica en el ejemplo.

Para β_1 el valor de *Tc* será:

$$Tc = \frac{\hat{\beta}_1}{\sqrt{C_{11} \times CME}} = \frac{-0{,}02746}{\sqrt{(0{,}04357)(0{,}0003354)}} = -\frac{-0{,}02746}{0{,}003822} = -7{,}184$$

Tal como se había probado antes, en este caso, se rechaza la hipótesis nula y se concluye, que hay dependencia lineal entre consumo y contenido de fibra.

Los intervalos de confianza para los estimadores de los parámetros, se calculan de la siguiente manera:

$$IC\beta_i = \hat{\beta}_i \pm \sqrt{C_{ii}\ CME} \times T_{\frac{\alpha}{2};\ n-2} \qquad (8.24)$$

Para el ejemplo que se está desarrollando, el intervalo de confianza para β_1, según la fórmula (8.24), sería:

$IC\beta_1$ = -0,02746 ± (0,003822 × 2,228) = -0,02746 ± 0,0085

El Coeficiente de Determinación se calcula del siguiente modo:

$$R^2 = \frac{\beta' X'Y - FC}{Y'Y - FC} = \frac{0{,}017346}{0{,}207} = 0{,}837$$

La interpretación de esta cifra, se indicó en el desarrollo anterior del mismo ejemplo, sin uso de la notación matricial.

ANÁLISIS DE REGRESIÓN NO LINEAL

El Modelo de Regresión de la Línea Recta es el más simple y más comúnmente utilizado entre las expresiones matemáticas con las que se representa la relación de dependencia entre dos variables. En ocasiones, este modelo es inapropiado, porque la ecuación que se ajusta al conjunto de puntos, puede no ser lineal. La no linealidad se determina visualmente, a partir del Diagrama de Dispersión; en otras ocasiones, el conocimiento teórico del problema o la experiencia, sugieren que el Modelo no es Lineal. Existe un número ilimitado de modelos no lineales, que cumplen dicha función, en el caso en que la línea recta no sea la apropiada.

Aquí sólo se hace referencia a tres de los tipos más utilizados de estos modelos: la Recíproca, la Exponencial y la Parabólica.

Suponiendo que los estimadores $\hat{\beta}_1$, $\hat{\beta}_2$ y $\hat{\beta}_3$ sean iguales *a, b* y *c*, respectivamente.

El ***Modelo Recíproco*** se expresa como $Y = a + \frac{b}{x}$, cuyas características gráficas se describen en la Figura 8.4.

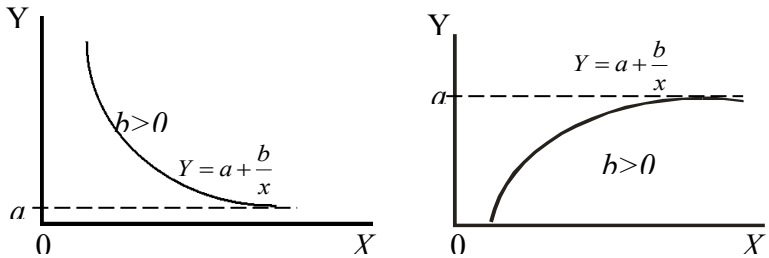


Figura 8.4 Representación gráfica del Modelo Recíproco.

Este modelo, mediante la transformación $Z = \frac{1}{X}$ de los datos, puede ser expresado como una línea recta $Y = a + bZ$.

El ***Modelo Exponencial*** tiene una expresión matemática $Y = ab^X$, cuya representación gráfica se muestra en la Figura 8.5.

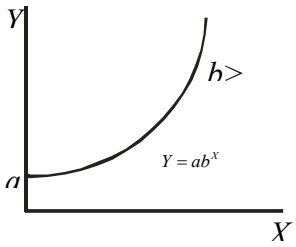


Figura 8.5 Representación gráfica del Modelo Exponencial.

En este caso, se expresa utilizando logaritmos:

$$Log\ Y = Log\ a + X\ (Log\ b) \qquad (8.25)$$

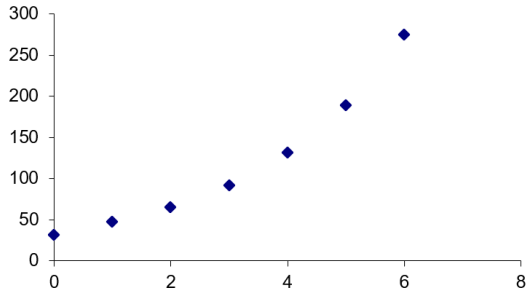
Haciendo la transformación $W = Log\ Y$, $g = Log\ a$ y $h = Log\ b$, el modelo se lo expresa como una recta:

$$W = g + hX \qquad (8.26)$$

Ejemplo 8.9: Los siguientes datos, corresponden al número de bacterias por unidad de volumen, presentes en un cultivo de *Salmonella sp*, durante seis horas.

Horas ***X***	**Número de bacterias por unidad de volumen** ***Y***
0	32
1	47
2	65
3	92
4	132
5	190
6	275

En el Diagrama de Dispersión, representado en el siguiente esquema, se puede apreciar el crecimiento exponencial de las bacterias.



Se ajusta a un modelo $Y = ab^X$

Tomando logaritmos, según la fórmula (8.25), se tiene: $log\ Y = log\ a + X\ (log\ b)$.

Si $W = log\ Y$, $g = log\ a$ y $h = log\ b$, el modelo se expresa como (8.26), mediante una línea recta:

$$W = g + h\,X$$

Ahora se puede resolver, de la misma manera que el Modelo de Línea Recta, teniendo en cuenta los siguientes datos:

X	***W = log y***
0	1,505
1	1,672
2	1,813
3	1,964
4	2,121
5	2,279
6	2,439

Los valores de los estimadores h y g se calculan como:

$$h = \frac{\sum XiWi - \frac{(\sum Xi)(\sum Wi)}{n}}{\sum X_i^2 - \frac{(\sum Xi)^2}{n}} = 0{,}154$$

$$g = \bar{W} - h\bar{X} = 1{,}507$$

Por lo tanto, el modelo se expresa como:

$$W = 1{,}507 + 0{,}154X$$

Debido a que los datos de la variable *Y* fueron transformados mediante logaritmos, no es posible interpretar este modelo calculado, por lo tanto, teniendo en cuenta las transformaciones, se tiene que:

Log a = *g,* entonces $a = 10^g = 10^{1{,}507} = 32{,}137$
Log b = *h*, entonces $b = 10^h = 10^{0{,}154} = 1{,}426$

Luego el Modelo Exponencial se expresa como:

$$Y = ab^X = 32{,}137\ (1{,}426)^X$$

Así, por ejemplo, para $X = 6$ el pronóstico bajo este modelo es:

$$Y = 32{,}137\ (1{,}426)^6 = 270.$$

Este valor está bastante cercano del valor observado:

$Y = 275$, para $X = 6$

El ***Modelo Parabólico*** se expresa mediante el modelo matemático:

$$Y = a + bX + cX^2 \qquad (8.27)$$

Este puede ser representado gráficamente, como aparece en la Figura 8.6.

Para hallar los valores de los estimadores *a, b* y *c* se resuelve el sistema:

$$\begin{aligned} \sum y_i &= na + b\sum x_i + c\sum x_i^2 \\ \sum x_i y_i &= a\sum x_i + b\sum x_i^2 + c\sum x_i^3 \\ \sum x_i^2 y_i &= a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 \end{aligned} \qquad (8.28)$$

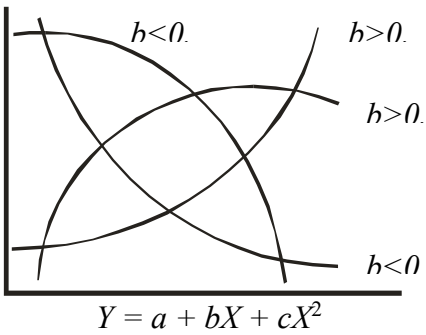


Figura 8.6 Representación gráfica del Modelo Parabólico

Ejemplo 8.10: En la Tabla 8.3 se presentan los datos de producción láctea de búfalas, en el Brasil, suministrados por Duarte et al (2001), donde *X* representa los días de lactancia, *Y* la producción en litros. Con esta información se pretende establecer un modelo que explique la producción en función de los días de lactancias.

Tabla 8.3

Datos de producción láctea de búfalas (*Bos buffalus)* en Brasil (2001)

(Duarte et al, 2001)

X	Y	XY	X^2	X^2Y	X^3	X^4
64	828	52.992	4.096	3.391.488	262.144	16.777.216
73	1.066	77.818	5.329	5.680.714	389.017	28.398.241
81	1.261	102.141	6.561	8.273.421	531.441	43.046.721
120	622	74.640	14.400	8.956.800	1.728.000	207.360.000
120	547	65.640	14.400	7.876.800	1.728.000	207.360.000
150	1.043	156.450	22.500	23.467.500	3.375.000	506.250.000
150	656	98.400	22.500	14.760.000	3.375.000	506.250.000
150	558	83.700	22.500	12.555.000	3.375.000	506.250.000
150	711	106.650	22.500	15.997.500	3.375.000	506.250.000
150	778	116.700	22.500	17.505.000	3.375.000	506.250.000
150	693	103.950	22.500	15.592.500	3.375.000	506.250.000
150	726	108.900	22.500	16.335.000	3.375.000	506.250.000
250	2.351	587.750	62.500	146.937.500	15.625.000	3.906.250.000
250	2.378	594.500	62.500	148.625.000	15.625.000	3.906.250.000
250	2.893	723.250	62.500	180.812.500	15.625.000	3.906.250.000
250	2.204	551.000	62.500	137.750.000	15.625.000	3.906.250.000
250	2.351	587.750	62.500	146.937.500	15.625.000	3.906.250.000
250	2.512	628.000	62.500	157.000.000	15.625.000	3.906.250.000
250	2.278	569.500	62.500	142.375.000	15.625.000	3.906.250.000
250	2.299	574.750	62.500	143.687.500	15.625.000	3.906.250.000
250	2.201	550.250	62.500	137.562.500	15.625.000	3.906.250.000
250	2.340	585.000	62.500	146.250.000	15.625.000	3.906.250.000
282	3.048	859.536	79.524	242.389.152	22.425.768	6.324.066.576
283	3.317	938.711	80.089	265.655.213	22.665.187	6.414.247.921
286	3.047	871.442	81.796	249.232.412	23.393.656	6.690.585.616
286	3.015	862.290	81.796	246.614.940	23.393.656	6.690.585.616
289	3.120	901.680	83.521	260.585.520	24.137.569	6.975.757.441
290	3.124	905.960	84.100	262.728.400	24.389.000	7.072.810.000
290	3.237	938.730	84.100	272.231.700	24.389.000	7.072.810.000
294	3.298	969.612	86.436	285.065.928	25.412.18b4	7.471.182.096
294	3.213	944.622	86.436	277.718.868	25.412.184	7.471.182.096

X	Y	XY	X^2	X^2Y	X^3	X^4
300	2.318	695.400	90.000	208.620.000	27.000.000	8.100.000.000
398	2.624	1.044.352	158.404	415.652.096	63.044.792	25.091.827.216
398	1.748	695.704	158.404	276.890.192	63.044.792	25.091.827.216
398	1.968	783.264	158.404	311.739.072	63.044.792	25.091.827.216
400	1.709	683.600	160.000	273.440.000	64.000.000	25.600.000.000
404	1.517	612.868	163.216	247.598.672	65.939.264	26.639.462.656
405	1.871	757.755	164.025	306.890.775	66.430.125	26.904.200.625
481	2.109	1.014.429	231.361	487.940.349	111.284.641	53.527.912.321
487	1.758	856.146	237.169	416.943.102	115.501.303	56.249.134.561
549	2.284	1.253.916	301.401	688.399.884	165.469.149	90.842.562.801
10.822	**81.621**	**23.689.748**	**3.397.468**	**7.634.665.498**	**1.204.890.664**	**468.431.174.152**

El Diagrama de Dispersión es el siguiente:

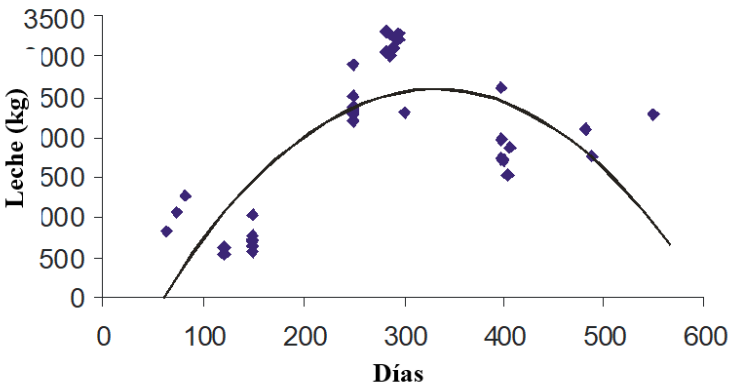


Figura 8.7 Diagrama de dispersión para los datos de producción láctea de búfalas (*Bos buffalus)* en Brasil (Duarte et al, 2001)

Se ajusta un Modelo Parabólico de segundo orden, tal como (8.27) $Y = a + bX + cX^2$

De acuerdo con (8.28), se resuelve el sistema:

$82.521 = 41a + 10.822b + 3.397.468c$

$23.959.748 = 10.822a + 3.397.468b + 1.204.890.664c$

$7.715.665.498 = 3.397.468a + 1.204.890.664b + 468.431.174.152c$

Al resolverlo se tiene:

$a = -1.198,11$ $b = 22,16$ $c = -0,032$

Por lo tanto, el modelo es:

$Y = -1.198,11 + 22,16X - 0,32X^2$

ANÁLISIS DE RESIDUALES. VALIDACIÓN DEL MODELO

Otra estrategia para evaluar cuánto se ajusta la Recta de Regresión de Mínimos Cuadrados, a los datos observados en la muestra utilizada, consiste en construir un gráfico de dispersión de los residuos, contra los valores predichos de la variable $\hat{Y}$ o contra X.

Se debe recordar que, el i-ésimo residuo e_i, es la diferencia entre el valor observado y_i y el correspondiente valor estimado $\hat{y}_i$.

Así, $e_i = y_i - \hat{y}_i$, para todo $i = 1, 2, 3, \ldots, n$

A las cantidades $d_i = \frac{e_i}{S} = \frac{e_i}{\sqrt{CME}}$, para todo $i = 1, 2, 3, \ldots, n$, se les denomina ***Residuales Estandarizados.***

Si el Modelo de Regresión Lineal Simple es correcto, los residuales estandarizados tienen una distribución normal estándar, por lo que los

residuales estarán aleatoriamente distribuidos (sin mostrar ningún patrón) alrededor del intervalo (-2, 2), como se muestra en la Figura 8.7, denominadas Gráficas de Residuales.

Las Gráficas de Residuales, que tienen un aspecto o tendencia similar al de la Figura 8.7, indican que el modelo es inadecuado, es decir, que es necesario añadir términos de orden superior (diferentes al lineal) para poder expresar la relación entre X e Y.

La Figura 8.7 corresponde a una Gráfica de Residuales que también muestra un patrón y, esto por lo general, indica que la Varianza de las observaciones puede aumentar la magnitud de X o Y. Para estabilizar la Varianza, se puede utilizar las transformaciones $\sqrt{Y}$, $Ln(y)$ o $1/y$.

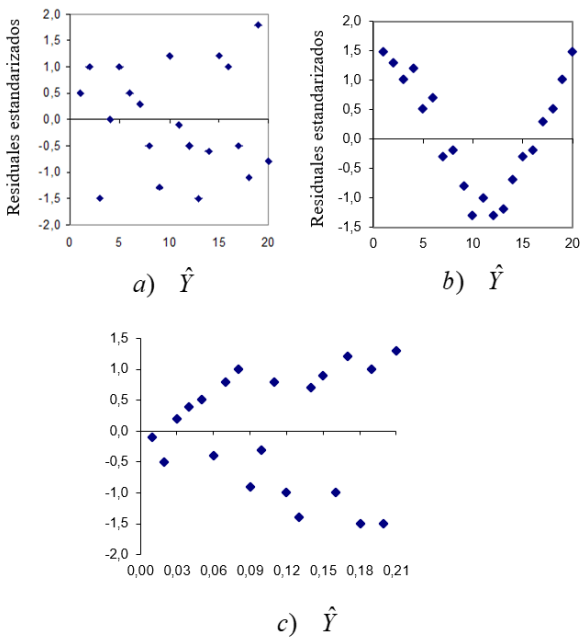


Figura 8.8 Patrones para Gráficos de Residuales

REGRESIÓN LINEAL MÚLTIPLE

La notación matricial, indicada para la Regresión Lineal Simple, se puede extender para este caso, donde el interés se centra en establecer la relación funcional de dependencia entre una variable *Y* con dos o más variables independientes *X*. Como se sabe, por ejemplo, el consumo de alimento no depende únicamente del contenido de fibra de la ración, sino también de otros factores, tales como, el peso vivo del animal, la temperatura del galpón, la densidad de alojamiento y muchas otras variables.

Las operaciones matriciales, indicadas en el numeral anterior, para la Regresión Lineal Simple, son de mucha utilidad para afrontar estudios de esta naturaleza. Las Pruebas de Hipótesis y los conceptos que deben manejarse, son básicamente extensiones del análisis simple.

Ejemplo 8.11: Para establecer la relación de dependencia lineal del consumo de alimento (Y) con respecto al peso corporal (X_1), la temperatura del galpón (X_2) y el diámetro de las partículas del alimento (X_3), en gallinas ponedoras Isabrown, se recopiló los datos consignados en la Tabla 8.4.

Tabla 8.4

Consumo de Alimento (kg) de gallinas Isabrown en función del Peso Corporal (kg), la Temperatura del Galpón (ºC) y el Diámetro de las Partículas del Alimento (mm).

Y Consumo (kg)	X_1 Peso Corporal (kg)	X_2 Temperatura (ºC)	X_3 Granulometría (mm)
22,00	1,50	23,0	3,9
22,37	1,53	23,5	3,9

Y **Consumo (kg)**	X_1 **Peso Corporal (kg)**	X_2 **Temperatura (ºC)**	X_3 **Granulometría (mm)**
22,75	1,60	26,0	3,8
22,75	1,63	25,8	3,8
23,62	1,70	27,0	3,4
23,32	1,67	26,5	3,3
23,75	1,77	27,0	3,3
21,50	1,40	28,0	4,0
24,15	1,77	28,3	3,3
22,50	1,53	26,0	3,8
23,77	1,73	24,0	3,3
21,85	1,37	23,0	3,9
22,42	1,53	22,5	3,8
22,57	1,60	24,7	3,8
24,20	1,70	26,0	3,3
24,00	1,73	27,0	3,3
24,00	1,87	22,8	3,2
23,50	1,73	26,7	3,3
23,42	1,67	28,5	3,2
23,80	1,70	26,8	3,2
24,30	1,80	24,4	3,2
21,80	1,43	24,6	3,9
23,00	1,57	23,0	3,9
23,25	1,60	25,0	3,8
21,05	1,43	28,0	3,9
21,25	1,40	27,4	3,9
21,17	1,47	26,0	3,8
23,63	1,67	25,0	3,2
24,50	1,80	24,0	3,2
25,35	1,93	23,5	3,2
25,22	1,90	22,7	3,1
24,05	1,73	22,6	3,2
24,18	1,80	28,0	3,1

Y Consumo (kg)	X_1 Peso Corporal (kg)	X_2 Temperatura (ºC)	X_3 Granulometría (mm)
21,55	1,43	27,0	3,9
21,48	1,40	26,0	3,9
21,75	1,47	25,0	3,8
22,30	1,57	24,7	3,8
25,00	1,77	25,8	3,2
25,00	1,80	23,0	3,2
21,75	1,47	25,6	3,8
21,20	1,40	26,5	3,9
21,10	1,33	27,0	3,9
24,40	1,83	28,0	3,1

El modelo lineal, para estos datos, se puede expresar con la siguiente ecuación o modelo teórico:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon_i \qquad (8.28)$$

Este modelo se puede extender, a cualquier número de variables independientes:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon_i. \qquad (8.29)$$

En este caso se indica, que las X varían de 1 a k.

En forma matricial, el modelo es el mismo para cualquier número de variables independientes, sólo que el tamaño de las matrices que contienen las ecuaciones normales aumenta, a medida que se incluyen más variables en los análisis; así, para una variable independiente el tamaño de $\boldsymbol{X'X}$ es 2 × 2 y el de $\boldsymbol{X'Y}$ es 2 × 1; para dos variables independientes $\boldsymbol{X'X}$ tiene orden de 3 × 3 y $\boldsymbol{X'Y}$ orden de 3 × 1. En

general, las matrices anteriores se pueden identificar de la siguiente forma, para *k* variables independientes.

$$X'X = \begin{bmatrix} n & \Sigma X_{1i} & \Sigma X_{2i} & . & . & \Sigma X_{ki} \\ \Sigma X_{1i} & \Sigma X_{1i}^2 & \Sigma X_{1i}\Sigma X_{2i} & . & . & \Sigma X_{1i}\Sigma X_{ki} \\ \Sigma X_{2i} & \Sigma X_{2i}\Sigma X_{1i} & \Sigma X_{2i}^2 & . & . & \Sigma X_{2i}\Sigma X_{ki} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \Sigma X_{ki} & \Sigma X_{ki}\Sigma X_{1i} & \Sigma X_{ki}\Sigma X_{2i} & . & . & \Sigma X_{ki}^2 \end{bmatrix} . \; X'Y = \begin{bmatrix} \Sigma Y_i \\ \Sigma X_{1i}Y_i \\ \Sigma X_{2i}Y_i \\ . \\ . \\ \Sigma X_{ki}Y_i \end{bmatrix} \quad (8.30)$$

Los modelos de la Regresión Lineal Múltiple generan matrices de rango completo, por lo tanto, la matriz ***X'X*** tiene inversa que al multiplicarla por el vector ***X'Y*** dará como resultado la estimación de los parámetros.

La Tabla de Análisis de Varianza, las Pruebas de Hipótesis, el cálculo de los Intervalos de Confianza y del Coeficiente de Determinación, se hace del mismo modo indicado en la sección "Regresión lineal simple con notación y operaciones matriciales".

> Para ilustrar con mayor claridad el proceso de Análisis de Regresión Múltiple, a continuación, se desarrolla el procedimiento utilizando los datos de la Tabla 8.4, en el Ejemplo 8.11, propuesto anteriormente.
>
> En primer lugar, se organiza la información y se llevan a cabo los cálculos necesarios para facilitar la posterior construcción de las matrices. Las expresiones necesarias, para el Análisis de Varianza, son las indicadas en la Tabla 8.5.

Tabla 8.5

Expresiones matemáticas y cálculo de los valores del Análisis de Varianza

Variable	ΣY_i	ΣY_i^2	$\Sigma Y_i X_{1i}$	$\Sigma Y_i X_{2i}$	$\Sigma Y_i X_{3i}$
Consumo (*Y*)	990,52	22888,1	1641,57	25225,1	3501,86
Variable	ΣX_{1i}	ΣX_{1i}^2	$\Sigma X_{1i} X_{2i}$	$\Sigma X_{1i} X_{3i}$	
Peso corporal (X_1)	69,73	114,19	1775,24	245,588	
Variable	ΣX_{2i}	ΣX_{2i}^2	$\Sigma X_{2i} X_{1i}$	$\Sigma X_{2i} X_{3i}$	
Temperatura (X_2)	1095,9	28067,5	1775,24	3891,44	
Variable	ΣX_{3i}	ΣX_{3i}^2	$\Sigma X_{3i} X_{1i}$	$\Sigma X_{3i} X_{2i}$	
Granulometría (X_3)	152,7	546,73	245,588	3891,44	

Con las cifras anteriores, se construyen las matrices correspondientes a las ecuaciones normales indicadas en (8.30).

$$X'X = \begin{bmatrix} 43 & 69{,}73 & 1095{,}9 & 152{,}7 \\ 69{,}73 & 114{,}19 & 1775{,}24 & 245{,}588 \\ 1095{,}9 & 1775{,}24 & 28067{,}5 & 3891{,}44 \\ 152{,}7 & 245{,}5888 & 3891{,}44 & 546{,}73 \end{bmatrix} . \quad X'Y = \begin{bmatrix} 990{,}52 \\ 1614{,}57 \\ 25225{,}1 \\ 3501{,}86 \end{bmatrix} . \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$$

Para obtener los estimadores de los parámetros β se invierte ***X'X*** y se multiplica ese resultado por ***X'Y***

$$(X'X)^{-1} = \begin{bmatrix} 90{,}91042807 & -23{,}00531994 & -0{,}5271542517 & -11{,}30503669 \\ -23{,}00531994 & 6{,}35717012 & 0{,}0939392135 & 2{,}901082521 \\ -0{,}5271542517 & 0{,}0939392135 & 0{,}008670169059 & 0{,}04332425144 \\ -11{,}30503669 & 2{,}901082521 & 0{,}04332425144 & 1{,}547773715 \end{bmatrix}$$

$$(X'X)^{-1}.X'Y = \hat{\beta} = \begin{bmatrix} 18{,}72328477 \\ 5{,}677543238 \\ -0{,}06404879725 \\ -0{,}9186985828 \end{bmatrix}$$

De acuerdo con los anteriores resultados, los valores del vector solución $\hat{\beta}$ se interpretan del siguiente modo:

$\hat{\beta}_0$ = 18,72328477. Punto de corte del eje *Y* con los ejes X_1, X_2, X_3. Valor del consumo cuando el peso corporal, la temperatura del galpón y el tamaño de las partículas del alimento, valen cero.

$\hat{\beta}_1$ = 5,677543238. Incremento promedio en el consumo por cada aumento de un kg, en el peso vivo de las aves.

$\hat{\beta}_2$ = -0,06404879725. Disminución promedio del consumo por cada aumento de un grado centígrado, en la temperatura ambiente del galpón.

$\hat{\beta}_3$ = -0,9186985828. Disminución promedio del consumo por cada aumento de un mm, en el tamaño de las partículas del alimento.

De acuerdo con las anteriores cifras, se puede observar, que a mayor peso vivo mayor consumo, mientras que, a mayor temperatura y diámetro de las partículas, el consumo disminuye.

El paso siguiente consiste en realizar las Pruebas de Hipótesis para cada uno de los parámetros, así que deben realizarse los cálculos necesarios de las sumas de cuadrados, las que luego se descomponen en las diferentes fuentes de variación. Las cifras necesarias son las siguientes:

$$Fc = \frac{(\Sigma y_i)^2}{n} = \frac{(990{,}52)^2}{43} = 22816{,}97373 = n\bar{y}^2$$
$$= 43 \times (90{,}10232558)^2$$

SC total = $\boldsymbol{Y'Y} - FC = 22884{,}1 - 22816{,}97373 = 67{,}1627$

SC regresión = $\boldsymbol{\beta' X'Y} - FC = 2{,}287978788 \times 10^4 - 22816{,}97373 = 62{,}81415$

La *SC del error* se obtiene por diferencia de la *SC total corregida* menos *SC regresión corregida*.

Lo más práctico es construir la Tabla de Análisis de Varianza, donde se indican las respectivas sumas de cuadrados para cada fuente de variación, de la siguiente manera:

FV	***GL***	***SC***	***CM***	***Fc***	***Ft***
Regresión	3	62,81415	20,9380500	187,78	2,84*
Error	39	4,34855	0,1115012		
Total	42	67,16270			

Como el *Fc* es mayor que el *Ft,* se concluye, con 95% de confiabilidad, que por lo menos una de las variables independientes *X* influye sobre el consumo, es decir, que el consumo depende linealmente por lo menos de una de las variables independientes.

Para establecer si sólo es una o son más variables independientes las que presentan esta influencia, debe efectuarse las Pruebas de Hipótesis específicas, para cada una de las variables, utilizando la Distribución de Probabilidad *T* de Student.

La regla general para obtener el *T* calculado (*Tc*) es la siguiente:

$$Tc = \frac{\hat{B}_j}{\sqrt{C_{jj} CME}}.$$ Donde $j = 1, 2, \ldots\ldots, k.$

C_{jj} = j-ésimo coeficiente de la diagonal de la matriz de coeficientes ***C***.

Los coeficientes de la matriz ***C*** se obtienen de la matriz inversa ***X'X***.

El *Tc* para β_1 se obtiene al desarrollar la siguiente expresión:

$$Tc = \frac{\hat{\beta}_1}{\sqrt{C_{11}CME}} = \frac{5{,}677543238}{\sqrt{6{,}35717012 \times 0{,}1115012}} = \frac{5{,}677543238}{0{,}8419216691} = 6{,}74$$

El *Tt* para 39 GL con α del 5% es 2,021 (Consultar la Tabla 7 del Apéndice), así que se debe rechazar la hipótesis nula y concluir que, el consumo depende linealmente del peso vivo de las gallinas, puesto que |6,74| es mayor que el valor de la tabla.

Es importante recordar que, cuando el *Tc* es negativo, la comparación con el *Tt* se hace con el valor absoluto, tal como se indicó en la Prueba de Hipótesis para dos medias.

Para β_2 el *Tc* es el siguiente:

$$Tc = \frac{-0{,}06404879725}{\sqrt{0{,}008670169059 \times 0{,}1115012}} = \frac{-0{,}06404879725}{0{,}03109235041} = -2{,}06$$

Como |1 - 2,061| es mayor que 2,021, se rechaza la hipótesis nula y se concluye, que el consumo depende linealmente de la temperatura del galpón.

Para β_3 el *Tc* es el siguiente:

$$Tc = \frac{-0{,}9186985828}{\sqrt{1{,}547773715 \times 0{,}1115012}} = \frac{-0{,}9186985828}{0{,}4154258376} = -2{,}21$$

Ya que |1 – 2,21| es mayor que 2,021, se rechaza la hipótesis nula y, la conclusión, es que el consumo depende linealmente del tamaño de las partículas del alimento.

PREDICCIÓN DE RESPUESTAS FUTURAS

La predicción de la respuesta promedio $\hat{y}_p$ para el punto $X_1, X_2, \ldots, X_k$, viene dado por:

$$\hat{y}_p = X'p\ \hat{\beta}, \text{ Donde } X'p = [1\ \ X_1\ \ X_2 \ldots X_k]$$

Un intervalo de confianza para la respuesta promedio en el punto $X_1, X_2, \ldots X_k$ está dado por:

$$\hat{y}_p \pm\ t_{1-\frac{\alpha}{2};\, n-k-1} \times S_{(\hat{y}_p)}$$

Donde: $S^2_{(\hat{y}_p)} = CME[X'p(X'X)^{-1}]Xp$

Ejemplo 8.12: La predicción $\hat{y}_p$ para el punto $x_1 = 1{,}9$, $x_2 = 28$ y $x_3 = 3{,}3$, en el Ejemplo 8.11, viene dado por:

$$\hat{y}_p = [1 \quad 1{,}9 \quad 28 \quad 3{,}3] \begin{bmatrix} 18{,}72 \\ 5{,}68 \\ -0{,}064 \\ -0{,}92 \end{bmatrix}$$

$\hat{y}_p = 1 \times 18{,}72 + 1{,}9 \times 5{,}68 – 28 \times 0{,}64 – 3{,}3 \times 0{,}92 = 24{,}684$

Un intervalo de confianza, al 95% de confiabilidad, para la respuesta promedio $\hat{y}_p$ = 24,684 está dado por:

$$\hat{y}_p \ \pm \ t_{0,975;\,39} \ \times S_{(\hat{y}_p)}$$

$$S^2_{(\hat{y}_p)} = CME[X'p(X'X)^{-1}]Xp$$

$$[1 \quad 1{,}9 \quad 28 \quad 3{,}3] \begin{bmatrix} 90{,}910 & -23{,}005 & -0{,}527 & -11{,}305 \\ -23{,}005 & 6{,}357 & 0{,}094 & 2{,}901 \\ -0{,}527 & 0{,}094 & 0{,}009 & 0{,}043 \\ -11{,}305 & 2{,}901 & 0{,}043 & 1{,}548 \end{bmatrix} \begin{bmatrix} 1 \\ 1{,}9 \\ 28 \\ 3{,}3 \end{bmatrix}$$

Por lo tanto:

$$S_{(\hat{y}_p)} = \sqrt{0{,}1115012 \ \times \ 0{,}555} = 0{,}248763$$

$$\hat{y}_p \ \pm \ t_{0,975;\,39} \approx 2{,}021$$

Y el intervalo de confianza es: 24,684 ± 2,021 × 0,248763

= (24,18 25,19)

Un intervalo de predicción al (1– α)100% de confiabilidad para una observación futura $\hat{y}_0$ en el valor $X_1, X_2, \ldots, X_k$ se expresa por:

$$\hat{y}_0 \pm t_{1\,-\,\frac{\alpha}{2};\ n-k-1} \times S_{(\hat{y}_0)}$$

Donde: $\hat{y}_0 = X_0^{'}\,\hat{\beta}$

$$X_0 = [1 \quad X_1 \quad X_2 \ldots X_k]$$
$$S_{(\hat{y}_0)} = CME\left[1 + X_0^{'} \quad (X'X)^{-1}\ X_0\right]$$

En el ejemplo, un intervalo de predicción para el punto x_1 = 1,9,

x_2 = 28 y x_3 = 3,3 al 95% de confiabilidad, viene dado por:

$$24{,}684 \pm 2{,}021 \times \sqrt{0{,}1115012 \quad (1 + 0{,}555)}$$
$$24{,}684 \pm 2{,}021 \times 0{,}4164$$
$$24{,}684 \pm 0{,}8416$$
$$(23{,}84 \quad 25{,}53)$$

Referencias bibliográficas

ASOCIACIÓN BRASILEÑA DE CRIADORES DE CEBÚ (ABCZ). 2001. Reporte de datos de producción. Sao Paulo, Brasil.

BOLDMAN, K. G. and VAN VLECK, L. D. 1991. Derivative-Free Restricted Maximun Likelihood estimation in animal models with a sparse matrix solver. J. Dairy. Sci. 74: 4337-4343.

BOLDMAN, K. G.; KRIESE, L. A.; VAN VLECK, L. D. and KACHMAN, S. D. 1993. A manual for use of MTDFRML: A set of programs to obtain estimates of variances and covariances. U.S.A., Departament of Agriculture, Agricultural Research Service, Washington DC.

BUNGE, M. 1976. La ciencia su método y su filosofía. Buenos Aires: Ariel.

CANAVOS, G. 1986. Probabilidad y Estadística: Aplicaciones y métodos. México: McGraw Hill, 109 p.

CARVAJAL, A.; VALDERRAMA, C. y VÁSQUEZ, M. 1993. Biometría. Cali, Colombia: Universidad del Valle, Facultad de Salud, Departamento de Enfermería.

CERÓN, M.; TONHATI, H.; COSTA, C.; SOLARTE, C. y BENAVIDES, O. 2003. Factores de ajuste para producción de leche en bovinos Holstein colombiano. Universidad de Antioquia, Revista Colombiana de Ciencias Pecuarias. 16:1, pp 26-32.

COOPERATIVE RESOURCES INTERNATIONAL (CRI). 2000. Catálogo de sementales Holstein. WI, USA.

DARNELL, J.; LODISH, H, y BALTIMORE, D. 1986. Biología celular y herencia. Madrid: Labor.

DEKKERS, J. C. M.; JAMROZIK, J.; ten HAG, J. H.; SCHAEFFER, L. R. & WEERSINK, A. (1996). Genetic and economic evaluation of persistency in dairy cattle. *In Proceedings:* International Workshop on Genetic Improvement of Functional Traits in Cattle. Interbull Bulletin No. 12.

DE MOIVRE, A. (1733). Approximatio ad summam terminorum Binomii (a + b)^n in seriem expansi. (folleto privado; versión impresa incluida en De Moivre A., The Doctrine of Chances, 2nd ed., 1738, pp. 235–243).

DUARTE, JMC; TONHATI, H; CERON-MUNOZ, MF; BERROCAL, MM; CANES, TS. 2001. Efeitos ambientais sobre a produção no dia do controle e características físicoquímicas do leite em um rebanho bubalino no estado de São Paulo, Brasil. Revista do Instituto de Lacticínios "Candido Tostes". 56, pp 16-19.

GARCÍA, H. y SOLARTE, C. 2001. Clasificación de cuyes *Cavia porcellus* según el tamaño de la camada al nacimiento mediante análisis discriminante. Pasto, Colombia, Universidad de Nariño, Facultad de Ciencias Pecuarias, Revista Zootecnia, 4:7, pp 63-70.

GUERRERO, E.; ROMERO, A. y ZÚÑIGA, A. 2002. Evaluación de la harina de vísceras de pollo como fuente de proteína en la fase juvenil del Goldfish *Carassius auratus*. Trabajo de Grado de Ingeniería en Producción Acuícola. Pasto, Colombia, Universidad de Nariño, Facultad de Ciencias Pecuarias.

HENDERSON, C. R. 1975. Best linear Unbiased Prediction under a selection model. Biometrics, 31, pp 423-447.

HENDERSON, C. R. 1984. Aplication of linear models in animal breeding. Canada, Ontario, University of Guelph.

HENDERSON, C. R. 1986. Recent developments in variance and covariance estimation. J. Anim. Sci, 63, pp 208-216.

HEHDERSON, C. R. 1988. Theoretical basis and computational methods for a number of different models: Proc. of animal model Workshop. Alberta, USA, American Dairy Science.

HOFER, A. 1998. Variance components estimation in animal breeding: a review. J. Anim. Breed. Genet, 115, pp 247-265.

JOHNSON, R. A. 2012. Probabilidad y estadística para ingenieros. México, Pearson Educación.

LARROSA, J. R. y BOREAN, J. R. 1990. Las razas ovinas y su producción de leche. En: Leche ovina y caprina, una nueva alternativa agroindustrial. Montevideo, Uruguay, Universidad de la República, Facultad de Veterinaria, Editorial Agropecuaria Hemisferio Sur.

LUSH, J. 1945. Animal Breeding Plans. 3th ed. Ames, Iowa, USA, Iowa State University Press.

MARTÍNEZ, R y MARTÍNEZ, N. 1997. Diseño de Experimentos: Análisis de datos estándar y no estándar. Bogotá, Fondo Nacional Universitario.

MEYER, P. 1973. Probabilidad y aplicaciones estadísticas. México: Fondo Educativo Interamericano. 382 p.

MEYER, K. 1988. Programs to estimation variance components for individual animal models by restricted maximum likelihood (REML): User notes. Scotland: University of Edimburg, Institute of Animal Genetics.

MEYER, K. 1989. Restricted maximum likelihood to estimation variance components for animals models with several random effects using a derivative: free algorithm. Genet. Select. Evol. 21, pp 317-340.

MEYER, K. 1991. Estimating variance and covariances for multivariate animal model by restricted maximum Likelihood. Genet. Select. Evol., 23, pp 67-83.

MEYER, K. 1993. DFREML, version 2.1: User notes. Armidale, Australia: University of New England.

MEYER, K. 1998. DFREML version 3.0 β: Manual User guide. Armidale, Australia: University of New England.

MILTON, J. 1994. Estadística para Biología y Ciencias de la Salud. 2ª. Ed. Madrid: Interamericana McGraw-Hill. 220 p.

NEYMAN, J. & IWASZKIEWICZ, K. 1935. Statistical problems in agricultural experimentation. Supplement to the Journal of the Royal Statistical Society; 2: 2, pp. 107-180.

ORTEGA, A. y RODRÍGUEZ, J. 2004. Evaluación comparativa del efecto del extracto pituitario de carpa (EPC) y gonadotropina coriónica humana (HCG) en la reproducción inducida del bagre del Patía *Rhamdia quelem* en condiciones de cautiverio. Trabajo de grado Ingeniería en Producción Acuícola. Pasto, Colombia: Universidad de Nariño, Facultad de Ciencias Pecuarias 123 p.

PARZEN, E. 1992. Teoría moderna de probabilidades y sus aplicaciones. Edición en Español México: Limusa. 512 p.

POPPER, K. 1971. La lógica de la investigación científica. Madrid: Tecnos.

SCHAEFFER, L. R. 1983. Effectiveness of model for cow evaluation intraherd. Journal of Dairy Science, 66(4), 874–880. https://doi.org/10.3168/jds.S0022-0302(83)81870-0

SCHAEFFER, L. R. & HENDERSON, C. R. 1983. Best linear unbiased prediction when error vector is correlated with other random vectors in the model. Génétique, Sélection, Evolution, 15**,** 395–400. https://doi.org/10.1186/1297-9686-15-3-395

SEARLE, S. R. 1989. Variance components: some history and a summary account of estimation methods. J. Anim. Breed. Genet. 106: 1-29.

SIERRA, R. 1991. Diccionario práctico de estadística. Madrid: Paraninfo, 97 p.

SILVA, R.G. 1982. Métodos de Genética Quantitativa aplicados ao Melhoramento Genético Animal. Ribeirão Preto, Brasil: Sociedade Brasileira de Genética, 162 p.

SOLARTE, C. 2000. Modelo animal multicarácter para la estimación de parámetros genéticos y la selección de reproductores *Cavia porcellus* en Colombia. Tesis Ph.D. La Habana, Cuba: Universidad Agraria de La Habana.

SOLARTE, C. y GARCÍA, H. 2001. Análisis de las variables productivas del *Cavia porcellus* en la granja Botana. Revista Zootecnia. Pasto, Colombia: Universidad de Nariño, Facultad de Ciencias Pecuarias, 4: 7, pp 53-62.

SOLARTE, C. y VITERI, L. 2001. Índice de selección, prueba de progenie y prueba de comportamiento en cuyes *Cavia porcellus*. Revista Zootecnia. Pasto, Colombia: Universidad de Nariño, Facultad de Ciencias Pecuarias, 4: 7, pp 35-44.

SPIEGEL, M. 1910. Probabilidad y Estadística. Madrid: McGraw-Hill, 416 p.

STEEL, R. y TORRIE, H. 1985. Bioestadística: Principios y procedimientos. Trad. por Ricardo Martínez. Bogotá: McGraw-Hill, 613 p.

SUZUKI, D.; GRIFFITHS, A.; MILLAR, J. y LEWONTIN, R. 1993. Introducción al análisis genético. 4a ed. Madrid: McGraw-Hill, 328 p.

WALLACE, J. B.; CUFFNEY T. F.; WEBSTER J. R.; LUGTHART G. J.; CHUNG K. & GOLDOWITZ B. S. 1991. Export of fine organic particles from headwater streams: Effects of season, extreme discharges, and invertebrate manipulation. Limnology and Oceanography, 36 (4), pp. 670–682. https://doi.org/10.4319/lo.1991.36.4.0670

WALPOLE, R. E.; MEYERS, R. H.; MEYERS, S. L.; YE K. 2012. Probabilidad y Estadística para Ingeniería y Ciencias. 9a. Edición. México: Pearson Educación, 792 p.

YATES, F. 1935. The Design and Analysis of Factorial Experiments. Technical comunication Nr. 35. London: Imperial Bureau of Soil Science, Harpenden, 95 p.

Apéndice

EJERCICIOS PROPUESTOS

I. Estadística Descriptiva

1. Defina claramente los siguientes términos:

    1) Variable Aleatoria
    2) Variable Aleatoria Cuantitativa
    3) Variable Aleatoria Cualitativa
    4) Variable Aleatoria Discreta
    5) Variable Aleatoria Continua
    6) Inferencia Estadística
    7) Ciclo inferencial
    8) Parámetro y Estadígrafo
    9) Coeficiente de Curtosis
    10) Coeficiente de Asimetría

2. Calcule Media, Moda, Mediana, Rango, RIQ, Varianza y Desviación Estándar, con los siguientes datos correspondientes al Peso Vivo Individual y a la cuarta semana, de machos *Cavia porcellus*, expresados en gramos, obtenidos en el plantel cuyícola de la Granja Botana de la Universidad de Nariño, por Solarte y Viteri (2001)

452, 427, 432, 427, 418	330, 370, 410, 298, 335
410, 415, 422, 432, 457	340, 345, 340, 342, 298
458, 432, 298, 298, 268	430, 465, 470, 330, 315
370, 372, 375, 375, 365	318, 330, 330, 335, 335
298, 307, 315, 320, 320	330, 410, 415, 420, 420

418, 430, 330, 330, 335
356, 365, 356, 380, 268
356, 380, 268, 330, 370
410, 298, 335, 340, 345
340, 342.

3. Construya la Tabla de Frecuencias para los mismos datos y luego elabore el Histograma y las Ojivas mayor que y menor que.

4. En un experimento conducente a evaluar el comportamiento de los peces ornamentales Escalares, de los cuales, el 50% son machos y el 50% son hembras, se recogió información sobre las variables Longitud Estándar, Peso y Número de Bandas en el cuerpo, encontrando los siguientes resultados:

Longitud cm	**Peso g**	**No. bandas**	**Longitud cm**	**Peso g**	**No. bandas**	**Longitud cm**	**Peso g**	**No. bandas**
15	11	2	13	12	3	12	12	3
14	12	3	12	11	3	11	10	3
11	11	3	10	10	2	10	11	2
8	8	2	9	9	2	7	7	1
15	12	2	15	12	3	11	9	3
13	12	2	14	12	3	14	12	3
15	13	3	12	12	3	12	11	2
12	10	2	13	12	3	15	15	3
8	7	1	9	9	2	10	10	2
12	10	3	12	11	3	13	12	3
14	12	3	14	11	3	10	9	3
13	12	3	7	3	1	8	5	1
9	5	1	7	6	1	8	7	2
7	7	1	12	12	3	13	13	3
14	14	3	10	9	2	5	5	1
6	6	1	5	5	1	4	4	1
7	7	1	6	5	1	10	9	2
5	5	1	7	7	1	8	8	2
9	9	2	11	10	3	12	11	3

12	9	2	12	10	3	13	9	3
4	4	1	6	5	1	5	5	1
6	6	1	7	7	1	5	4	1
9	9	2	5	5	1	9	8	2
12	11	3	14	13	3	15	11	2
13	12	3	12	12	3	16	11	3

a) Construya la Tabla de Frecuencias para la Variable discreta y el Diagrama Circular correspondiente.

b) Calcule los Cuartiles 1, 2, 3 y 4 y los Percentiles 5, 10, 15, 25, 50, 75, 80, 85, 90 y 95 para la variable Peso con datos agrupados y sin agrupar. Interprete los resultados.

c) Calcule la Varianza, la Desviación Estándar, el Coeficiente de Variación, el Coeficientes Asimetría y la Curtosis para la Variable Longitud.

II. Probabilidades

5. En una urna se tienen 9 bolas idénticas, con diferente color, tres son blancas, tres rojas y tres negras. Se extrae tres bolas, una detrás de otra, sin devolverla al recipiente, y se anota el color. ¿Cuál es el Espacio Muestral para este experimento particular?

6. Si en un tanque se tiene 6 reproductores hembras de Trucha, debidamente identificados del 1 al 6, y se saca al azar 2 de ellos. ¿Cuál es el Espacio Muestral?

7. Una empresa camaronera cultiva tres especies diferentes: *P. vannamei, P. stylirostris* y *P. occidentalis*. Con el fin de realizar un estudio para

determinar los parásitos que los atacan, se tomó al azar muestras constituidas por 10 animales de cada especie.

a. Describa el Espacio Muestral correspondiente.
b. Describa el conjunto de camarones *P. vannamei* afectados con parásitos externos.
c. Describa el conjunto de camarones *P. stylirostris* con parásitos del tracto digestivo.
d. Describa el conjunto de camarones *P. occidentalis* con parásitos externos.
e. ¿Cuál será el conjunto B∪C, B∩C, B∪D, B∩D?
f. Si la Probabilidad de que B tenga parásitos externos es del 40%, C del 30% y D del 20%; calcular la P[B∪C∪D]?; calcular P[$\bar{A}$]?
g. Si del conjunto conformado por la especie *P. vannmei* se saca un animal, ¿cuál es la probabilidad de que no tenga parásitos externos?
h. Si se extrae tres camarones *P. stylirostris*, ¿cuál es la probabilidad de que los tres tengan parásitos?, ¿cuál es la probabilidad de que dos tengan parásitos?, ¿cuál es la probabilidad de que sólo uno tenga parásitos?
i. Si se extrae tres camarones, ¿cuál es la probabilidad de que todos sean *P. vannamei*? ¿Cuál es la probabilidad de que se obtenga uno de cada especie? ¿Cuál es la probabilidad de que dos sean *P. stylirostris* y uno *P. occidentalis*?
j. Si se saca una muestra al azar sin reemplazo, conformada por cuatro animales, ¿cuántas muestras se pueden formar?
k. Si el muestreo fuese con reemplazo, ¿cuántas muestras de cuatro animales se pueden formar?

8. En un proceso de incubación de ovas de carpa (*Cyprinus carpio*) se intenta probar un nuevo material para permitir una mejor adherencia de los huevos y disminuir la mortalidad. En total fueron sembrados 254.340 huevos, de los cuales murieron 28.350. ¿Cuál es la probabilidad de que un huevo que se incuba en este nuevo sistema se muera?

9. En un laboratorio dedicado a la reproducción de Bocachico (*Prochilodus reticultatus*) se utiliza extracto pituitario de carpa (EPC) para inducir a 12 reproductores, prymogonil para inducir a 15 reproductores y las dos hormonas combinadas para inducir a 9 reproductores. Si después de la inducción son colocados en un solo tanque, y se extrae uno al azar, ¿cuál es la probabilidad de que haya sido inducido con EPC o con prymogonil?

10. Si la inducción se hizo a un macho por cada dos hembras, ¿cuál es la probabilidad de que, al sacar un macho, haya sido inducido con hipófisis de carpa? ¿Cuál es la probabilidad de que, habiendo sido inducido con prymogonil, sea una hembra?

11. Considerando un parto no gemelar en *Bos taurus*, ¿cuál es la probabilidad de obtener un macho?, ¿cuál es la probabilidad de obtener una hembra?

    Si se trata de registrar el sexo en dos partos sucesivos, calcule las siguientes probabilidades:

    a) Obtener un macho.
    b) Obtener un macho y una hembra.
    c) Ningún macho.

12. Un Médico Veterinario está interesado en probar tres antibióticos, identificados con A, B y C, para el tratamiento de la mastitis clínica; para

esto dispone de 15 vacas enfermas. Se asigna al azar a cada vaca uno de los antibióticos; se pregunta:

a) ¿De cuántas maneras distintas se pueden asignar los tres tratamientos entre las 15 vacas?

b) Cuál es la probabilidad de que 10 vacas reciban el tratamiento A, 3 el B y 2 el C.

13. Se está planeando iniciar una explotación lechera, donde se utilizará el semen de tres reproductores. Los machos se seleccionan por catálogo, de un listado donde aparecen registrados 20 animales, de los cuales ocho son Holstein rojo, nueve Holstein negro y tres Pardo Suizo. Si los toros se eligen al azar, calcule las siguientes probabilidades:

a) Que los tres toros elegidos sean Holstein rojo.
b) Que dos sean Holstein rojo y uno Pardo Suizo.
c) Que se elija uno de cada raza.
d) Que por lo menos uno sea Pardo Suizo.
e) Que sean elegidos en el orden: Holstein rojo, Holstein negro, Pardo Suizo.

14. En la raza de ganado lechero Holstein, un alelo recesivo *r* produce pelo rojo y blanco; el alelo dominante *R* produce pelo blanco y negro. Si un toro portador es cruzado con vacas portadoras, determine las siguientes probabilidades:

a) El primer descendiente que nazca sea rojo y blanco.

b) Los primeros cuatro descendientes sean blanco y negro.

c) ¿Cuál es la proporción fenotípica esperada entre la progenie resultante de retrocruzar vacas F1 blanco y negro con el toro portador?

d) Si el toro portador se cruza con vacas blanco y negro homocigóticas, ¿qué proporción fenotípica puede esperarse entre la progenie resultante de retrocruzar las vacas de la F1 por el macho portador?

15. Si dos pares de alelos se transmiten independientemente, siendo *A* dominante sobre *a* y *B* sobre *b*, ¿cuál es la probabilidad de obtener:

a) ¿Un gameto *Ab* a partir de un individuo *AaBb*?

b) ¿Un cigoto *AABB* a partir de un cruzamiento *AaBB* × *AaBb*?

c) ¿Un gameto *Ab* a partir de un individuo *AABb*?

d) ¿Un cigoto *AABB* a partir de un cruzamiento *aabb* × *AABB*?

e) ¿Un fenotipo *AB* a partir de un cruzamiento *AaBb* × *AaBb*?

f) ¿Un fenotipo *AB* a partir de un cruzamiento *AaBb* × *AABB*?

g) ¿Un fenotipo *aB* a partir de un cruzamiento *AaBb* × *AaBB*?

III. Probabilidades con Variables Continuas, Tamaño de Muestra y Estimación por Intervalo.

16. Asuma que, en una muestra de 78 animales, de los 980 que se sacrifican mensualmente en una planta de beneficio, se calcula la Media del rendimiento en canal de los cerdos (*Sus scrofa*) en 67,5 kg y su Varianza es de 169 kg^2 , ¿cuáles son los límites de confianza para μ y σ, con α del 1%, 2,5% y 5%?

17. Considerando los mismos datos del Ejercicio 16, responda las siguientes preguntas:

a) ¿Cuál es la Probabilidad de encontrar animales con rendimiento superior a 70 kg?

b) ¿Cuál es la Probabilidad de encontrar animales con rendimiento inferior a 70 kg?

c) ¿Cuál es la Probabilidad de encontrar animales con rendimientos entre 60 y 80 kg?

d) ¿Cuál es la Probabilidad de encontrar animales con rendimientos entre 55 y 65 kg?

e) ¿Cuál es la Probabilidad de encontrar animales con rendimientos entre 65 y 95 kg?

f) ¿Cuál es la Probabilidad de encontrar animales con rendimientos menores de 60 kg y mayores de 90 kg?

g) ¿Cuál es la Probabilidad de encontrar animales con rendimientos superiores a 110 kg?

h) Sí los animales ubicados en el 5% superior se pagan a un mejor precio, ¿cuál será el mínimo peso para recibir bonificación?

i) Sí los animales ubicados en el 5% inferior no se comercializan en canal, sino que se destinan exclusivamente a salsamentaria, siendo pagados a un precio menor, ¿cuál será el mínimo peso para destinar un animal a salsamentaria?

j) Se pide determinar el tamaño de muestra para realizar el mismo estudio, considerando la Varianza antes citada, ¿cuántos animales deberá pesar, considerando valores α de 10%, 5%, 2,5% y 1% y márgenes de error del 1%, 5%, 10%, 15% y 20%?

19. En la zona lechera de Nariño, fueron sometidos a prueba 258 animales, de los cuales 58 resultaron positivas a Brucelosis, ¿cuál es el límite de confianza para la verdadera proporción de animales enfermos, con α del 1% y 5 %? ¿Cuántos animales debería someter a prueba para estimar la verdadera prevalencia en la zona, considerando valores α del 10%, 5%, 2,5% y 1% y errores del 1%, 5%, 10%, 15% y 20%, para cada alfa?

20. Un estudio pretende determinar la prevalencia de *Toxocora canis,* en una zona donde, por censo, se ha establecido que existen 1800 animales. Con base en esta información, determine el tamaño de muestra para realizar el estudio en los siguientes casos:

    a) Confiabilidad del 99%, margen de error del 1%, sin existencia de datos previos sobre la prevalencia.

    b) Confiabilidad del 99%, margen de error del 1% y prevalencia previa conocida del 40%.

    c) Confiabilidad del 99%, margen de error del 1%, prevalencia previa conocida del 40% y carencia de censo.

    c) Confiabilidad del 95%, margen de error del 5%, prevalencia previa conocida del 40% y carencia de censo.

21. Bacca y Portillo (2003), en su estudio, se propusieron como objetivo, determinar la prevalencia de anticuerpos de *Brucella sp* en trabajadores de derivados lácteos en un corregimiento del departamento de Nariño, Colombia. Para ello consiguieron un dato de Prevalencia Estimada del 12% y, un Censo Previo, determinó que en la zona 162 personas se dedicaban al procesamiento de leche. La confiabilidad pedida en el estudio fue del 97% y el margen de error del 3%. Con los datos indicados, ¿cuál debe ser el tamaño de muestra adecuado para realizar el estudio?

IV. Prueba de hipótesis

22. Un plan de vacunación auspiciado por la CEPANZOO, busca erradicar gradualmente la fiebre aftosa en Nariño, cuya prevalencia se ha calculado en 17%. Luego de cumplir con el programa de vacunación se toma al azar una muestra de 158 animales de los cuales 24 tenían aftosa. ¿Podría afirmar, con una confiabilidad del 95% que se está controlando la enfermedad? Justifique claramente su respuesta.

23. Un ensayo para comparar los coeficientes de digestibilidad del ensilaje de maíz en corderos y novillos (Steel y Torrie, 1985) arrojó los siguientes resultados:

Y1 (corderos)	**Y2 (novillos)**
57,8	64,2
56,2	58,7
61,9	63,1
54,4	62,5
53,6	59,8
56,4	59,2
53,2	

Con base en estos datos responda si hay o no diferencias en los coeficientes de digestibilidad en las dos especies y calcule un intervalo o límite de confianza con 95% de confiabilidad para la diferencia de los promedios del coeficiente de digestibilidad.

24. El mismo autor del problema inmediatamente anterior presenta otros datos en los que señala el porcentaje de grava fina en dos tipos de suelos así: Suelo A: 5,9, 3,8, 6,5, 18,3, 18,2, 16,1, 7,6. Suelo B: 7,6, 0,4, 1,1,

3,2, 6,5, 4,1, 4,7. ¿Considera que existen diferencias significativas en el porcentaje de grava fina en los dos suelos?

25. Un Zootecnista probó la eficacia de un concentrado comercial y una dieta balanceada por él para el levante de novillos cebú comercial. Al final del ensayo los resultados, expresados como la ganancia diaria en kg fueron los siguientes:

Concentrado comercial	**Dieta balanceada**
0,85	0,86
0,86	1,15
0,87	1,20
0,88	1,30
0,89	0,89
0,90	0,78
0,91	0,89
0,92	0,90
0,93	0,85
0,94	0,90
0,95	1,10
0,96	1,20
0,97	0,78
0,98	1,20
0,99	0,68

Con confiabilidad del 95%¿Cuál de las dos dietas recomendaría y por que?

26. Con la puesta en marcha de un programa de salud y prevención implantado por un equipo de Zootecnistas y Veterinarios se espera mejorar la producción de leche vaca día en una zona lechera. La media

de producción antes del citado plan era de 25,78 L. Para constar la eficacia del programa se tomaron al azar 25 vacas, cuya producción promedio fue de 26,14 L. con una varianza de 365 L^2. Con una confiabilidad del 95%, pruebe que el plan implantado logró incrementar la media de producción vaca / día.

27. Guerrero, et al (2.002) evaluaron la harina de vísceras de pollo (HVP) como fuente de proteína en la fase juvenil del Goldfish *Carassius auratus*. Los resultados para la variable conversión alimenticia aparente para dos raciones fueron los siguientes:

Dieta A (concentrado comercial)	**Dieta B (20% HVP)**
2,52	1,23
2,43	0,97
1,06	2,87
2,60	2,96

De acuerdo con las anteriores cifras, ¿podría asegurar que existen diferencias significativas entre las dos dietas? ¿Cuál tratamiento recomendaría?

V. Pruebas de independencia y Bondad de ajuste

28. En una población de ganado Holstein se aparearon individuos F1 para el color, del que se obtuvo 738 crías negras y 257 crías rojas, ¿se cumple la proporción Mendeliana 3:1? ¿por qué?

29. En esa misma población, del número 28, al cruzar dihíbridos para color de manto y presencia o ausencia de cuernos, se encontró los siguientes resultados: 559 crías negras mochas, 190 negras con cuernos, 199 rojas mochas y 52 rojas con cuernos ¿se cumple la proporción 9:3:3:1?

30. La pigmentación del pelo en los cuyes podría estar asociada con la resistencia a patógenos. Un experimento proporcionó los siguientes resultados:

Patógenos	**Color del pelo**			
	Blanco	**Castaño**	**Negro**	**Bayo**
Yersinia	354	198	192	303
Ácaros	80	89	72	86
Cándida	57	31	17	34
Nemátodos	62	60	34	41
Coccidia	78	61	60	57

De acuerdo con el número de casos positivos, ¿se puede afirmar que los dos atributos son independientes?

VI. REGRESIÓN Y CORRELACIÓN

31. Un investigador está interesado en establecer la relación de dependencia entre el consumo de alimento en cerdos y el contenido de la fibra de la ración. Al final del ensayo se encontró los siguientes datos:

Contenido de fibra de la dieta (kg) (X)	**Consumo en kg (Y)**
1	1,68
2	1,74
3	1,86
4	1,94
5	2,09
6	2,25

Con la información anterior, responda estas preguntas:

a. ¿Existe relación funcional de dependencia entre las variables? Explique su respuesta.
b. ¿Cuál es la Ecuación de Predicción?
c. ¿Cuál es el valor de β_1 y qué significa?
d. ¿Cuál es Límite de Confianza al 95% del verdadero β_1?
e. ¿Cuánto vale el Coeficiente de Determinación y qué significado tiene?

32. Con fines de selección genética, se realizó un estudio para establecer la relación de asociación entre grasa y proteína de la leche en cabras. Los resultados de las mediciones fueron los siguientes:

Grasa (kg/L)	**Proteína (kg/L)**
0,22	0,018
0,22	0,016
0,25	0,013
0,28	0,011
0,29	0,009
0,30	0,007
0,31	0,007
0,32	0,006
0,33	0,004
0,34	0,003
0,35	0,0022

De acuerdo con las cifras anteriores:

a. ¿Se puede afirmar que las variables están correlacionadas?
b. ¿Cuál es el valor del Coeficiente de Correlación?
c. ¿Qué recomendaría con fines de selección?

33. Un investigador en cuyicultura asevera que el peso individual al destete de los gazapos (PIDG) depende linealmente de la edad de las madres (en meses), de las madres al primer servicio (EMPS) y del peso de las mismas al primer servicio (PMPS). Al recolectar una muestra se encontró los siguientes resultados:

PIDG (g)	**EMPS**	**PMPS (g)**
280	4,3	840
260	5,0	860
270	4,6	850
280	4,0	840
284	3,0	900
302	3,0	870
300	3,2	610
310	3,0	750
285	5,0	840
250	4,0	800
230	5,0	880
280	3,0	800
310	4,0	960
350	4,0	940
360	3,5	950
280	3,0	900
250	4,0	850
260	4,0	840
270	3,0	830
300	3,2	950

a. ¿Tiene razón el investigador y por qué?
b. ¿Cuál es el cambio promedio en PIDG por efecto de EMPS y PMPS?
c. ¿Cuál es la Ecuación de Predicción?
d. ¿Cuánto vale el Coeficiente de Determinación y qué significa?
e. ¿Cuáles son los Límites de Confianza para cada β?
f. ¿Cuál sería el valor de PIDG si TCN vale 1 y PMPS 910?
g. ¿PIDG y PMPS son variables correlacionadas? ¿por qué?

Tabla 1.

Número de categorías sugerido según el tamaño de la muestra
(Tomada de Milton, 1994).

n	**Número de categorías**
Menos de 16	Datos insuficientes
16 -31	5
32 – 63	6
64 – 127	7
128 – 255	8
256 – 511	9
512 -1023	10
1024 -2047	11
2048 – 4095	12
4096 – 8190	13

Tabla 2.

Valores de la función de distribución acumulativa binomial

$$P[X \leq x] = F(x;\ n,\ p) = \sum_{i=0}^{x} \binom{n}{i} p^x (1-p)^{n-x}$$

		p										
n	***x***	**0,01**	**0,05**	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**0,35**	**0,40**	**0,45**	**0,50**
2	0	0,980	0,902	0,810	0,722	0,640	0,562	0,4900	0,422	0,360	0,3025	0,2500
	1	0,999	0,997	0,990	0,977	0,960	0,937	0,9100	0,877	0,840	0,7975	0,7500
	2	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,0000
3	0	0,970	0,857	0,729	0,614	0,512	0,421	0,3430	0,274	0,216	0,1664	0,1250
	1	0,999	0,992	0,972	0,939	0,896	0,843	0,7840	0,718	0,648	0,5748	0,5000
	2	1,000	0,999	0,999	0,996	0,992	0,984	0,9730	0,957	0,936	0,9089	0,8750
	3	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
4	0	0,960	0,814	0,656	0,522	0,409	0,316	0,2401	0,178	0,129	0,0915	0,0625
	1	0,999	0,986	0,947	0,890	0,819	0,738	0,6517	0,563	0,475	0,3910	0,3125

	2	1,000	0,999	0,996	0,988	0,972	0,949	0,9163	0,873	0,820	0,7585	0,6875
	3	1,000	1,000	0,999	0,999	0,998	0,996	0,9919	0,985	0,974	0,9590	0,9375
	4	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
5	0	0,951	0,773	0,590	0,443	0,327	0,237	0,1681	0,116	0,077	0,0503	0,0313
	1	0,999	0,977	0,918	0,835	0,737	0,632	0,5282	0,428	0,337	0,2562	0,1875
	2	1,000	0,998	0,991	0,973	0,942	0,896	0,8369	0,764	0,682	0,5931	0,5000
	3	1,000	1,000	0,999	0,997	0,993	0,984	0,9692	0,946	0,913	0,8688	0,8125
	4	1,000	1,000	1,000	0,999	0,999	0,999	0,9976	0,994	0,989	0,9815	0,9688
	5	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
n	***x***	**0,01**	**0,05**	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**0,35**	**0,40**	**0,45**	**0,50**
6	0	0,941	0,735	0,531	0,377	0,262	0,178	0,1176	0,075	0,046	0,0277	0,0156
	1	0,998	0,967	0,885	0,776	0,655	0,533	0,4202	0,319	0,233	0,1636	0,1094
	2	1,000	0,997	0,984	0,952	0,901	0,830	0,7443	0,647	0,544	0,4415	0,3438
	3	1,000	0,999	0,998	0,994	0,983	0,962	0,9295	0,882	0,820	0,7447	0,6563
	4	1,000	1,000	0,999	0,999	0,998	0,995	0,9891	0,977	0,959	0,9308	0,8906
	5	1,000	1,000	1,000	1,000	0,999	0,999	0,9993	0,998	0,995	0,9917	0,9844
	6	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
7	0	0,932	0,698	0,478	0,320	0,209	0,133	0,0824	0,049	0,028	0,0152	0,0078
	1	0,998	0,955	0,850	0,716	0,576	0,444	0,3294	0,233	0,158	0,1024	0,0625
	2	1,000	0,996	0,974	0,926	0,852	0,756	0,6471	0,532	0,419	0,3164	0,2266
	3	1,000	0,999	0,997	0,987	0,966	0,929	0,8740	0,800	0,710	0,6083	0,5000
	4	1,000	1,000	0,999	0,998	0,995	0,987	0,9712	0,944	0,903	0,8471	0,7734
	5	1,000	1,000	1,000	0,999	0,999	0,998	0,9962	0,991	0,981	0,9643	0,9375
	6	1,000	1,000	1,000	1,000	1,000	0,999	0,9998	0,999	0,998	0,9963	0,9922
	7	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
8	0	0,922	0,663	0,430	0,272	0,167	0,100	0,0576	0,031	0,016	0,0084	0,0039
	1	0,997	0,942	0,813	0,657	0,503	0,367	0,2553	0,169	0,106	0,0632	0,0352
	2	0,999	0,994	0,961	0,894	0,796	0,678	0,5518	0,427	0,315	0,2201	0,1445
	3	1,000	0,999	0,995	0,978	0,943	0,886	0,8059	0,706	0,594	0,4770	0,3633
	4	1,000	1,000	0,999	0,997	0,989	0,972	0,9420	0,893	0,826	0,7396	0,6367
	5	1,000	1,000	1,000	0,999	0,998	0,995	0,9887	0,974	0,950	0,9115	0,8555
	6	1,000	1,000	1,000	1,000	0,999	0,999	0,9987	0,996	0,991	0,9819	0,9648
	7	1,000	1,000	1,000	1,000	1,000	1,000	0,9999	0,999	0,999	0,9983	0,9961
	8	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
9	0	0,913	0,630	0,387	0,231	0,134	0,075	0,0404	0,020	0,010	0,0046	0,0020
	1	0,996	0,928	0,774	0,599	0,436	0,300	0,1960	0,121	0,070	0,0385	0,0195

	2	0,999	0,991	0,947	0,859	0,738	0,600	0,4628	0,337	0,231	0,1495	0,0898
	3	1,000	0,999	0,991	0,966	0,914	0,834	0,7297	0,608	0,482	03614	0,2539
	4	1,000	1,000	0,999	0,994	0,980	0,951	0,9012	0,828	0,733	0,6214	0,5000
	5	1,000	1,000	0,999	0,999	0,996	0,990	0,9747	0,946	0,900	0,8342	0,7461
	6	1,000	1,000	1,000	1,000	0,999	0,998	0,9957	0,988	0,975	0,9502	0,9102
	7	1,000	1,000	1,000	1,000	1,000	0,999	0,9996	0,998	0,996	0,9909	0,9805
	8	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	0,999	0,999	0,9992	0,9980
	9	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
1	0	0,904	0,598	0,348	0,196	0,107	0,056	0,0282	0,013	0,006	0,0025	0,0010
0	1	0,995	0,913	0,736	0,544	0,375	0,244	0,1493	0,086	0,046	0,0233	0,0107
	2	0,999	0,988	0,929	0,820	0,677	0,525	0,3828	0,261	0,167	0,0996	0,0547
	3	1,000	0,999	0,987	0,950	0,879	0,775	0,6496	0,513	0,382	0,2660	0,1719

n	***x***	**0,01**	**0,05**	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**0,35**	**0,40**	**0,45**	**0,50**
	4	1,000	0,999	0,998	0,990	0,967	0,921	0,8497	0,751	0,633	0,5044	0,3770
	5	1,000	1,000	0,999	0,998	0,993	0,980	0,9527	0,905	0,833	0,7384	0,6230
1	6	1,000	1,000	1,000	0,999	0,999	0,996	0,9894	0,974	0,945	0,8980	0,8281
0	7	1,000	1,000	1,000	1,000	0,999	0,999	0,9984	0,995	0,987	0,9726	0,9453
	8	1,000	1,000	1,000	1,000	1,000	1,000	0,9999	0,999	0,998	0,9955	0,9893
	9	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	0,999	0,9997	0,9990
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
		0,895	0,568	0,313	0,167	0,085	0,042	0,0198	0,008	0,003	0,0014	0,0005
	1	0,994	0,898	0,697	0,492	0,322	0,197	0,1130	0,060	0,030	0,0139	0,0059
	2	0,999	0,984	0,910	0,778	0,617	0,455	0,3127	0,200	0,118	0,0652	0,0327
	3	1,000	0,998	0,981	0,930	0,838	0,713	0,5696	0,425	0,296	0,1911	0,1133
1	4	1,000	0,999	0,997	0,984	0,949	0,885	0,7897	0,668	0,532	0,3971	0,2744
1	5	1,000	1,000	0,999	0,997	0,988	0,965	0,9218	0,851	0,753	0,6331	0,5000
	6	1,000	1,000	1,000	0,999	0,998	0,992	0,9784	0,949	0,900	0,8262	0,7256
	7	1,000	1,000	1,000	1,000	0,999	0,998	0,9957	0,987	0,970	0,9390	0,8867
	8	1,000	1,000	1,000	1,000	1,000	0,999	0,9994	0,998	0,994	0,9852	0,9673
	9	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	0,999	0,999	0,9978	0,9941
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	0,9998	0,9995
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
		0,886	0,540	0,282	0,142	0,068	0,031	0,0138	0,005	0,002	0,0008	0,0002

12	1	0,993	0,881	0,659	0,443	0,274	0,158	0,0850	0,042	0,019	0,0083	0,0032
	2	0,999	0,980	0,889	0,735	0,558	0,390	0,2528	0,151	0,083	0,0421	0,0193
	3	1,000	0,997	0,974	0,907	0,794	0,648	0,4925	0,346	0,225	0,1345	0,0730
	4	1,000	0,999	0,995	0,976	0,927	0,842	0,7237	0,583	0,438	0,3044	0,1938
	5	1,000	1,000	0,999	0,995	0,980	0,945	0,8822	0,787	0,665	0,5269	0,3872
	6	1,000	1,000	0,999	0,999	0,996	0,985	0,9614	0,915	0,841	0,7393	0,6128
	7	1,000	1,000	1,000	0,999	0,999	0,997	0,9905	0,974	0,942	0,8883	0,8062
	8	1,000	1,000	1,000	1,000	0,999	0,999	0,9983	0,994	0,984	0,9644	0,9270
	9	1,000	1,000	1,000	1,000	1,000	1,000	0,9998	0,999	0,997	0,9921	0,9807
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	0,999	0,999	0,9989	0,9968
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	0,9999	0,9998
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
13	0	0,877	0,513	0,254	0,120	0,055	0,023	0,0097	0,003	0,001	0,0004	0,0001
	1	0,992	0,864	0,621	0,398	0,233	0,126	0,0637	0,029	0,012	0,0049	0,0017
	2	0,999	0,975	0,866	0,692	0,501	0,332	0,2025	0,113	0,057	0,0269	0,0112
	3	1,000	0,996	0,965	0,882	0,747	0,584	0,4206	0,278	0,168	0,0929	0,0461
	4	1,000	0,999	0,993	0,965	0,900	0,794	0,6543	0,500	0,353	0,2279	0,1334
	5	1,000	1,000	0,999	0,992	0,970	0,919	0,8346	0,715	0,574	0,4268	0,2905
n	***x***	**0,01**	**0,05**	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**0,35**	**0,40**	**0,45**	**0,50**
13	6	1,000	1,000	0,999	0,998	0,993	0,975	0,9376	0,870	0,771	0,6437	0,5000
	7	1,000	1,000	1,000	0,999	0,998	0,994	0,9818	0,953	0,902	0,8212	0,7095
	8	1,000	1,000	1,000	1,000	0,999	0,999	0,9960	0,987	0,967	0,9302	0,8666
	9	1,000	1,000	1,000	1,000	1,000	0,999	0,9993	0,997	0,992	0,9797	0,9539
	1	1,000	1,000	1,000	1,000	1,000	1,000	0,9999	0,999	0,998	0,9959	0,9888
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	0,999	0,9995	0,9983
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	0,9999
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
14	0	0,868	0,487	0,228	0,102	0,044	0,017	0,0068	0,002	0,000	0,0002	0,0001
	1	0,991	0,847	0,584	0,356	0,197	0,101	0,0475	0,020	0,008	0,0029	0,0009
	2	0,999	0,969	0,841	0,647	0,448	0,281	0,1608	0,083	0,039	0,0170	0,0065
	3	1,000	0,995	0,955	0,853	0,698	0,521	0,3552	0,220	0,124	0,0632	0,0287
	4	1,000	0,999	0,990	0,953	0,870	0,741	0,5842	0,422	0,279	0,1672	0,0898
	5	1,000	1,000	0,998	0,988	0,956	0,888	0,7805	0,640	0,485	0,3373	0,2120
	6	1,000	1,000	0,999	0,997	0,988	0,961	0,9067	0,816	0,692	0,5461	0,3953
	7	1,000	1,000	1,000	0,999	0,997	0,989	0,9685	0,924	0,849	0,7414	0,6047
	8	1,000	1,000	1,000	1,000	0,999	0,997	0,9917	0,975	0,941	0,8811	0,7880

	9	1,000	1,000	1,000	1,000	1,000	0,999	0,9983	0,994	0,982	0,9574	0,9102
	1	1,000	1,000	1,000	1,000	1,000	1,000	0,9998	0,998	0,996	0,9886	0,9713
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	0,999	0,999	0,9978	0,9935
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	0,999	0,9997	0,9991
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	0,9999
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
	0	0,860	0,463	0,205	0,087	0,035	0,013	0,0047	0,001	0,000	0,0001	0,0000
	1	0,990	0,829	0,549	0,318	0,167	0,080	0,0353	0,014	0,005	0,0017	0,0005
	2	0,999	0,963	0,815	0,604	0,398	0,236	0,1268	0,061	0,027	0,0107	0,0037
	3	1,000	0,994	0,944	0,822	0,648	0,461	0,2969	0,172	0,090	0,0424	0,0176
	4	1,000	0,999	0,987	0,938	0,835	0,686	0,5155	0,351	0,217	0,1204	0,0592
	5	1,000	0,999	0,997	0,983	0,938	0,851	0,7216	0,564	0,403	0,2608	0,1509
	6	1,000	1,000	0,999	0,996	0,981	0,943	0,8689	0,754	0,609	0,4522	0,3036
15	7	1,000	1,000	1,000	0,999	0,995	0,982	0,9500	0,886	0,786	0,6535	0,5000
	8	1,000	1,000	1,000	0,999	0,999	0,995	0,9848	0,957	0,905	0,8182	0,6964
	9	1,000	1,000	1,000	1,000	0,999	0,999	0,9963	0,987	0,966	0,9231	0,8491
	1	1,000	1,000	1,000	1,000	1,000	0,999	0,9993	0,997	0,990	0,9745	0,9408
	1	1,000	1,000	1,000	1,000	1,000	1,000	0,9999	0,999	0,998	0,9937	0,9824
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	0,999	0,999	0,9989	0,9963
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	0,9999	0,9995
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
n	***x***	**0,01**	**0,05**	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**0,35**	**0,40**	**0,45**	**0,50**
		0,851	0,440	0,185	0,074	0,028	0,010	0,0033	0,001	0,000	0,0001	0,0000
	1	0,989	0,810	0,514	0,283	0,140	0,063	0,0261	0,009	0,003	0,0010	0,0003
	2	0,999	0,957	0,789	0,561	0,351	0,197	0,0994	0,045	0,018	0,0066	0,0021
	3	1,000	0,993	0,931	0,789	0,598	0,405	0,2459	0,133	0,065	0,0281	0,0106
	4	1,000	0,999	0,983	0,920	0,798	0,630	0,4499	0,289	0,166	0,0853	0,0384
16	5	1,000	0,999	0,996	0,976	0,918	0,810	0,6598	0,490	0,328	0,1976	0,1051
	6	1,000	1,000	0,999	0,994	0,973	0,920	0,8247	0,688	0,527	0,3660	0,2272
	7	1,000	1,000	0,999	0,998	0,993	0,972	0,9256	0,840	0,716	0,5629	0,4018
	8	1,000	1,000	1,000	0,999	0,998	0,992	0,9743	0,932	0,857	0,7441	0,5982
	9	1,000	1,000	1,000	1,000	0,999	0,998	0,9929	0,977	0,941	0,8759	0,7728
	1	1,000	1,000	1,000	1,000	1,000	0,999	0,9984	0,993	0,980	0,9514	0,8949
	1	1,000	1,000	1,000	1,000	1,000	1,000	0,9997	0,998	0,995	0,9851	0,9616
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	0,999	0,999	0,9965	0,9894

	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	0,999	0,9994	0,9979
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	0,9999	0,9997
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
17	0	0,842	0,418	0,166	0,063	0,022	0,007	0,0023	0,000	0,000	0,0000	0,0000
	1	0,987	0,792	0,481	0,252	0,118	0,050	0,0193	0,006	0,002	0,0006	0,0001
	2	0,999	0,949	0,761	0,519	0,309	0,163	0,0774	0,032	0,012	0,0041	0,0012
	3	1,000	0,991	0,917	0,755	0,548	0,353	0,2019	0,102	0,046	0,0184	0,0064
	4	1,000	0,998	0,977	0,901	0,758	0,573	0,3887	0,234	0,126	0,0596	0,0245
	5	1,000	0,999	0,995	0,968	0,894	0,765	0,5968	0,419	0,263	0,1471	0,0717
	6	1,000	1,000	0,999	0,991	0,962	0,892	0,7752	0,618	0,447	0,2902	0,1662
	7	1,000	1,000	0,999	0,998	0,989	0,959	0,8954	0,787	0,640	0,4743	0,3145
	8	1,000	1,000	1,000	0,999	0,997	0,987	0,9597	0,900	0,801	0,6626	0,5000
	9	1,000	1,000	1,000	1,000	0,999	0,996	0,9873	0,961	0,908	0,8166	0,6855
	1	1,000	1,000	1,000	1,000	0,999	0,999	0,9968	0,988	0,965	0,9174	0,8338
	1	1,000	1,000	1,000	1,000	1,000	0,999	0,9993	0,997	0,989	0,9699	0,9283
	1	1,000	1,000	1,000	1,000	1,000	1,000	0,9999	0,999	0,997	0,9914	0,9755
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	0,999	0,999	0,9981	0,9936
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	0,999	0,9997	0,9988
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	0,9999
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
18	0	0,834	0,397	0,150	0,053	0,018	0,005	0,0016	0,000	0,000	0,0000	0,0000
	1	0,986	0,773	0,450	0,224	0,099	0,039	0,0142	0,004	0,001	0,0003	0,0001
	2	0,999	0,941	0,733	0,479	0,271	0,135	0,0600	0,023	0,008	0,0025	0,0007

n	***x***	**0,01**	**0,05**	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**0,35**	**0,40**	**0,45**	**0,50**
18	3	1,000	0,989	0,901	0,720	0,501	0,305	0,1646	0,078	0,032	0,0120	0,0038
	4	1,000	0,998	0,971	0,879	0,716	0,518	0,3327	0,188	0,094	0,0411	0,0154
	5	1,000	0,999	0,993	0,958	0,867	0,717	0,5344	0,355	0,208	0,1077	0,0481
	6	1,000	1,000	0,998	0,988	0,948	0,861	0,7217	0,549	0,374	0,2258	0,1189
	7	1,000	1,000	0,999	0,997	0,983	0,943	0,8593	0,728	0,563	0,3915	0,2403
	8	1,000	1,000	1,000	0,999	0,995	0,980	0,9404	0,860	0,736	0,5778	0,4073
	9	1,000	1,000	1,000	0,999	0,999	0,994	0,9790	0,940	0,865	0,7473	0,5927
	1	1,000	1,000	1,000	1,000	0,999	0,998	0,9939	0,978	0,942	0,8720	0,7597

	1	1,000	1,000	1,000	1,000	1,000	0,999	0,9986	0,993	0,979	0,9463	0,8811
	1	1,000	1,000	1,000	1,000	1,000	1,000	0,9997	0,998	0,994	0,9817	0,9519
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	0,999	0,998	0,9951	0,9846
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	,0000	0,999	0,9990	0,9962
	1	1,000	1,000	1,000	1,000	1,000	,0000	1,0000	1,000	1,000	0,9999	0,9993
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	0,9999
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
	0	0,826	0,377	0,135	0,045	0,014	0,004	0,0011	0,000	0,0001	0,0000	0,0000
	1	0,984	0,754	0,420	0,198	0,082	0,031	0,0104	0,003	0,0008	0,0002	0,0000
	2	0,999	0,933	0,705	0,441	0,236	0,111	0,0462	0,017	0,0055	0,0015	0,0004
	3	1,000	0,986	0,885	0,684	0,455	0,263	0,1332	0,059	0,0230	0,0077	0,0022
	4	1,000	0,998	0,964	0,855	0,673	0,465	0,2822	0,150	0,0696	0,0280	0,0096
	5	1,000	0,999	0,991	0,946	0,836	0,667	0,4739	0,296	0,1629	0,0777	0,0318
	6	1,000	1,000	0,998	0,983	0,932	0,825	0,6655	0,481	0,3081	0,1727	0,0835
	7	1,000	1,000	0,999	0,995	0,976	0,922	0,8180	0,665	0,4878	0,3169	0,1796
	8	1,000	1,000	1,000	0,999	0,993	0,971	0,9161	0,814	0,6675	0,4940	0,3238
1	9	1,000	1,000	1,000	0,999	0,998	0,991	0,9674	0,912	0,8139	0,6710	0,5000
9	10	1,000	1,000	1,000	1,000	0,999	0,997	0,9895	0,965	0,9115	0,8159	0,6762
	11	1,000	1,000	1,000	1,000	1,000	0,999	0,9972	0,988	0,9648	0,9129	0,8204
	12	1,000	1,000	1,000	1,000	1,000	0,999	0,9994	0,996	0,9884	0,9658	0,9165
	13	1,000	1,000	1,000	1,000	1,000	1,000	0,9999	0,999	0,9969	0,9891	0,9682
	14	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	0,999	0,9994	0,9972	0,9904
	15	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	0,9999	0,9995	0,9978
	16	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,0000	0,9999	0,9996
	17	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,0000	1,0000	1,0000
	18	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,0000	1,0000	1,0000
	19	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,0000	1,0000	1,0000
2	0	0,817	0,358	0,121	0,038	0,011	0,003	0,0008	0,000	0,0000	0,0000	0,0000
0	1	0,983	0,735	0,391	0,175	0,069	0,024	0,0076	0,002	0,0005	0,0001	0,0000

n	***x***	**0,01**	**0,05**	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**0,35**	**0,40**	**0,45**	**0,50**
	2	0,999	0,924	0,676	0,404	0,206	0,091	0,0355	0,012	0,0036	0,0009	0,0002
	3	1,000	0,984	0,867	0,647	0,411	0,225	0,1071	0,044	0,0160	0,0049	0,0013
	4	1,000	0,997	0,956	0,829	0,629	0,414	0,2375	0,118	0,0510	0,0189	0,0059

	5	1,000	0,999	0,988	0,932	0,804	0,617	0,4164	0,245	0,1256	0,0553	0,0207
	6	1,000	1,000	0,997	0,978	0,913	0,785	0,6080	0,416	0,2500	0,1299	0,0577
	7	1,000	1,000	0,999	0,994	0,967	0,898	0,7723	0,601	0,4159	0,2520	0,1316
	8	1,000	1,000	0,999	0,998	0,990	0,959	0,8867	0,762	0,5956	0,4143	0,2517
	9	1,000	1,000	1,000	0,999	0,997	0,986	0,9520	0,878	0,7553	0,5914	0,4119
	10	1,000	1,000	1,000	1,000	0,999	0,996	0,9829	0,946	0,8725	0,7507	0,5881
2	1	1,000	1,000	1,000	1,000	0,999	0,999	0,9949	0,980	0,943	0,8692	0,7483
0	1	1,000	1,000	1,000	1,000	1,000	0,999	0,9987	0,994	0,979	0,9420	0,8684
	1	1,000	1,000	1,000	1,000	1,000	1,000	0,9997	0,998	0,993	0,9786	0,9423
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	0,999	0,998	0,9936	0,9793
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	0,999	0,9985	0,9941
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	0,9997	0,9987
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	0,0998
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
	2	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
	0	0,777	0,277	0,071	0,017	0,003	0,000	0,0001	0,000	0,000	0,0000	0,0000
	1	0,974	0,642	0,271	1,093	0,027	0,007	0,0016	0,000	0,000	0,0000	0,0000
	2	0,998	0,872	0,537	0,253	0,098	0,032	0,0090	0,002	0,000	0,0001	0,0000
	3	0,999	0,965	07636	0,471	0,234	0,096	0,0332	0,009	0,002	0,0005	0,0001
	4	1,000	0,992	0,902	0,682	0,420	0,213	0,0905	0,032	0,009	0,0023	0,0005
	5	1,000	0,998	0,966	0,838	0,616	0,378	0,1935	0,082	0,029	0,0086	0,0020
	6	1,000	0,999	0,990	0,930	0,780	0,561	0,3407	0,173	0,073	0,0258	0,0073
	7	1,000	1,000	0,997	0,974	0,890	0,726	0,5118	0,306	0,153	0,0639	0,0216
2	8	1,000	1,000	0,999	0,992	0,953	0,850	0,6769	0,466	0,273	0,1340	0,0539
5	9	1,000	1,000	0,999	0,997	0,982	0,928	0,8106	0,630	0,424	0,2424	0,1148
	1	1,000	1,000	1,000	0,999	0,994	0,970	0,9022	0,771	0,585	0,3843	0,2122
	1	1,000	1,000	1,000	0,999	0,998	0,989	0,9558	0,874	0,732	0,5426	0,3450
	1	1,000	1,000	1,000	1,000	0,999	0,996	0,9825	0,939	0,846	0,6937	0,5000
	1	1,000	1,000	1,000	1,000	0,999	0,999	0,9940	0,974	0,922	0,8173	0,6550
	1	1,000	1,000	1,000	1,000	1,000	0,999	0,9982	0,990	0,965	0,9040	0,7878
	1	1,000	1,000	1,000	1,000	1,000	1,000	0,9995	0,997	0,986	0,9560	0,8852
	1	1,000	1,000	1,000	1,000	1,000	1,000	0,9999	0,999	0,995	0,9826	0,9461
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	0,999	0,998	0,9942	0,9784
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	0,999	0,9984	0,9927

n	***x***	**0,01**	**0,05**	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**0,35**	**0,40**	**0,45**	**0,50**
	1	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	0,999	0,9996	0,9980
	2	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	0,9999	0,9995
2	2	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	0,9999
5	2	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
	2	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
	2	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000
	2	1,000	1,000	1,000	1,000	1,000	1,000	1,0000	1,000	1,000	1,0000	1,0000

Tabla 3.

Valores de la función de distribución acumulativa de Poisson.

$$P(X \leq x) = F(x; \lambda) = \sum_{i=0}^{x} \frac{e^{-\lambda} \lambda^{i}}{i!}$$

	λs									
x	**0,1**	**0,2**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,8**	**0,9**	**1,0**
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	0,3679
1	0,9953	0,9825	0,9631	0,9384	0,9098	0,8781	0,8442	0,8088	0,7725	0,7358
2	0,9998	0,9989	0,9964	0,9921	0,9856	0,9769	0,9659	0,9526	0,9371	0,9197
3	1,0000	0,9999	0,9997	0,9992	0,9982	0,9966	0,9942	0,9909	0,9865	0,9810
4	1,0000	1,0000	1,0000	0,9999	0,9998	0,9996	0,9992	0,9986	0,9977	0,9963
5	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9997	0,9994
6	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999
7	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	λs									
x	**1,1**	**1,2**	**1,3**	**1,4**	**1,5**	**1,6**	**1,7**	**1,8**	**1,9**	**2,0**
0	0,3329	0,3012	0,2725	0,2466	0,2231	0,2019	0,1827	0,1653	0,1496	0,1353
1	0,6990	0,6626	0,6268	0,5918	0,5578	0,5249	0,4932	0,4628	0,4338	0,4060
2	0,9004	0,8795	0,8571	0,8335	0,8088	0,7834	0,7572	0,7306	0,7037	0,6767
3	0,9743	0,9662	0,9569	0,9463	0,9344	0,9212	0,9068	0,8913	0,8747	0,8571
4	0,9946	0,9923	0,9893	0,9857	0,9814	0,9763	0,9704	0,9636	0,9559	0,9473
5	0,9990	0,9985	0,9978	0,9968	0,9955	0,9940	0,9920	0,9896	0,9868	0,9834
6	0,9999	0,9997	0,9996	0,9994	0,9991	0,9987	0,9981	0,9974	0,9966	0,9955
7	1,0000	1,0000	0,9999	0,9999	0,9998	0,9997	0,9996	0,9994	0,9992	0,9989
8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9998	0,9998
9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	λs									
x	**2,1**	**2,2**	**2,3**	**2,4**	**2,5**	**2,6**	**2,7**	**2,8**	**2,9**	**3,0**
0	0,1225	0,1108	0,1003	0,0907	0,0821	0,0743	0,0672	0,0608	0,0550	0,0498
1	0,3796	0,3546	0,3309	0,3084	0,2873	0,2674	0,2487	0,2311	0,2146	0,1991
2	0,6496	0,6227	0,5960	0,5697	0,5438	0,5184	0,4936	0,4695	0,4460	0,4232
3	0,8386	0,8194	0,7993	0,7787	0,7576	0,7360	0,7141	0,6919	0,6696	0,6472
4	0,9379	0,9275	0,9163	0,9041	0,8912	0,8774	0,8629	0,8477	0,8318	0,8153
5	0,9796	0,9751	0,9700	0,9643	0,9580	0,9510	0,9433	0,9349	0,9258	0,9161
6	0,9941	0,9925	0,9906	0,9884	0,9858	0,9828	0,9794	0,9756	0,9713	0,9665
7	0,9985	0,9980	0,9974	0,9967	0,9958	0,9947	0,9934	0,9919	0,9901	0,9881
8	0,9997	0,9995	0,9994	0,9991	0,9989	0,9985	0,9981	0,9976	0,9969	0,9962
9	0,9999	0,9999	0,9999	0,9998	0,9997	0,9996	0,9995	0,9993	0,9991	0,9989
10	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9998	0,9998	0,9997
11	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999
	λs									

x	**3,1**	**3,2**	**3,3**	**3,4**	**3,5**	**3,6**	**3,7**	**3,8**	**3,9**	**4,0**
0	0,0450	0,0408	0,0369	0,0334	0,0302	0,0273	0,0247	0,0224	0,0202	0,0183
1	0,1847	0,1712	0,1586	0,1468	),1359O,	0,1257	0,1162	0,1074	0,0992	0,0916
2	0,4012	0,3799	0,3594	0,3397	0,3208	0,3027	0,2854	0,2689	0,2531	0,2381
3	0,6248	0,6025	0,5803	0,5584	0,5366	0,5152	0,4942	0,4735	0,4533	0,4335
4	0,7982	0,7806	0,7626	0,7442	0,7254	0,7064	0,6872	0,6678	0,6484	0,6288
5	0,9057	0,8946	0,8829	0,8705	0,8576	0,8441	0,8301	0,8156	0,8006	0,7851
6	0,9612	0,9554	0,9490	0,9421	0,9347	0,9267	0,9182	0,9091	0,8995	0,8893
7	0,9858	0,9832	0,9802	0,9769	0,9733	0,9692	0,9648	0,9599	0,9546	0,9489
8	0,9953	0,9943	0,9931	0,9917	0,9901	0,9883	0,9863	0,9840	0,9815	0,9786
9	0,9986	0,9982	0,9978	0,9973	0,9967	0,9960	0,9952	0,9942	0,9931	0,9919
10	0,9996	0,9995	0,9994	0,9992	0,9990	0,9987	0,9984	0,9981	0,9977	0,9972
11	0,9999	0,9999	0,9998	0,9998	0,9997	0,9996	0,9995	0,9994	0,9993	0,9991
12	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998	0,9997
13	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999
	λs									
x	**4,1**	**4,2**	**4,3**	**4,4**	**4,5**	**4,6**	**4,7**	**4,8**	**4,9**	**5,0**
0	0,0166	0,0150	0,0136	0,0123	0,0111	0,0101	0,0091	0,0082	0,0074	0,0067
1	0,0845	0,0780	0,0719	0,0663	0,0611	0,0563	0,0518	0,0477	0,0439	0,0404
2	0,2238	0,2102	0,1974	0,1851	0,1736	0,1626	0,1523	0,1425	0,1333	0,1247
3	0,4142	0,3954	0,3772	0,3595	0,3423	0,3257	0,3097	0,2942	0,2793	0,2650
4	0,6093	0,5898	0,5704	0,5512	0,5321	0,5132	0,4946	0,4763	0,4582	0,4405
5	0,7693	0,7531	0,7367	0,7199	0,7029	0,6858	0,6684	0,6510	0,6335	0,6160
6	0,8787	0,8675	0,8558	0,8436	0,8311	0,8180	0,8046	0,7908	0,7767	0,7622
7	0,9427	0,9361	0,9290	0,9214	0,9134	0,9050	0,8960	0,8867	0,8769	0,8666
8	0,9755	0,9721	0,9683	0,9642	0,9597	0,9549	0,9497	0,9442	0,9382	0,9319
9	0,9905	0,9889	0,9871	0,9851	0,9829	0,9805	0,9778	0,9749	0,9717	0,9682
10	0,9966	0,9959	0,9952	0,9943	0,9933	0,9922	0,9910	0,9896	0,9880	0,9863
11	0,9989	0,9986	0,9983	0,9980	0,9976	0,9971	0,9966	0,9960	0,9953	0,9945
12	0,9997	0,9996	0,9995	0,9993	0,9992	0,9990	0,9988	0,9986	0,9983	0,9980
13	0,9999	0,9999	0,9998	0,9998	0,9997	0,9997	0,9996	0,9995	0,9994	0,9993
14	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998
15	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999
	λs									
x	**5,1**	**5,2**	**5,3**	**5,4,**	**5,5**	**5,6**	**5,7**	**5,8**	**5,9**	**6,0**
0	0,0061	0,0055	0,0050	0,0045	0,0041	0,0037	0,0033	0,0030	0,0027	0,0025
1	0,0372	0,0342	0,0314	0,0289	0,0266	0,0244	0,0224	0,0206	0,0189	0,0174
2	0,1165	0,1088	0,1016	0,0948	0,0884	0,0824	0,0768	0,0715	0,0666	0,0620
3	0,2513	0,2381	0,2254	0,2133	0,2017	0,1906	0,1801	0,1700	0,1604	0,1512
4	0,4231	0,4061	0,3895	0,3733	0,3575	0,3422	0,3272	0,3127	0,2987	0,2851
5	0,5984	0,5809	0,5635	0,5461	0,5289	0,5119	0,4950	0,4783	0,4619	0,4457
6	0,7474	0,7324	0,7171	0,7017	0,6860	0,6703	0,6544	0,6384	0,6224	0,6063
7	0,8560	0,8449	0,8335	0,8217	0,8095	0,7970	0,7842	0,7710	0,7576	0,7440

8	0,9252	0,9181	0,9106	0,9027	0,8944	0,8857	0,8766	0,8672	0,8574	0,8472
9	0,9644	0,9603	0,9559	0,9512	0,9462	0,9409	0,9352	0,9292	0,9228	0,9161
10	0,9844	0,9823	0,9800	0,9775	0,9747	0,9718	0,9686	0,9651	0,9614	0,9574
11	0,9937	0,9927	0,9916	0,9904	0,9890	0,9875	0,9859	0,9841	0,9821	0,9799
12	0,9976	0,9972	0,9967	0,9962	0,9955	),9949 *	0,9941	0,9932	0,9922	0,9912
13	0,9992	0,9990	0,9988	0,9986	0,9983	0,9980	0,9977	0,9973	0,9969	0,9964
14	0,9997	0,9997	0,9996	0,9995	0,9994	0,9993	0,9991	0,9990	0,9988	0,9986
15	0,9999	0,9999	0,9999	0,9998	0,9998	0,9998	0,9997	0,9996	0,9996	0,9995
16	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9998
17	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999
	λs									
x	**6,1**	**6,2**	**6,3**	**6,4**	**6,5**	**6,6**	**6,7**	**6,8**	**6,9**	**7,0**
0	0,0022	0,0020	0,0018	0,0017	0,0015	0,0014	0,0012	0,0011	0,0010	0,0009
1	0,0159	0,0146	0,0134	0,0123	0,0113	0,0103	0,0095	0,0087	0,0080	0,0073
2	0,0577	0,0536	0,0498	0,0463	0,0430	0,0400	0,0371	0,0344	0,0320	0,0296
3	0,1425	0,1342	0,1264	0,1189	0,1119	0,1052	0,0988	0,0928	0,0871	0,0818
4	0,2719	0,2592	0,2469	0,2351	0,2237	0,2127	0,2022	0,1920	0,1823	0,1730
5	0,4298	0,4141	0,3988	0,3837	0,3690	0,3547	0,3407	0,3270	0,3137	0,3007
6	0,5902	0,5742	0,5582	0,5423	0,5265	0,5108	0,4953	0,4799	0,4647	0,4497
7	0,7301	0,7160	0,7018	0,6873	0,6728	0,6581	0,6433	0,6285	0,6136	0,5987
8	0,8367	0,8259	0,8148	0,8033	0,7916	0,7796	0,7673	0,7548	0,7420	0,7291
9	0,9090	0,9016	0,8939	0,8858	0,8774	0,8686	0,8596	0,8502	0,8405	0,8305
10	0,9531	0,9486	0,9437	0,9386	0,9332	0,9274	0,9214	0,9151	0,9084	0,9015
11	0,9776	0,9750	0,9723	0,9693	0,9661	0,9627	0,9591	0,9552	0,9510	0,9467
12	0,9900	0,9887	0,9873	0,9857	0,9840	0,9821	0,9801	0,9779	0,9755	0,9730
13	0,9958	0,9952	0,9945	0,9937	0,9929	0,9920	0,9909	0,9898	0,9885	0,9872
14	0,9984	0,9981	0,9978	0,9974	0,9970	0,9966	0,9961	0,9956	0,9950	0,9943
15	0,9994	0,9993	0,9992	0,9990	0,9988	0,9986	0,9984	0,9982	0,9979	0,9976
16	0,9998	0,9997	0,9997	0,9996	0,9996	0,9995	0,9994	0,9993	0,9992	0,9990
17	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998	0,9998	0,9997	0,9997	0,9996
18	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
19	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	λs									
x	**7,1**	**7,2**	**7,3**	**7,4**	**7,5**	**7,6**	**7,7**	**7,8**	**7,9**	**8,0**
0	0,0008	0,0007	0,0007	0,0006	0,0006	0,0005	0,0005	0,0004	0,0004	0,0003
1	0,0067	0,0061	0,0056	0,0051	0,0047	0,0043	0,0039	0,0036	0,0033	0,0030
2	0,0275	0,0255	0,0236	0,0219	0,0203	0,0188	0,0174	0,0161	0,0149	0,0138
3	0,0767	0,0719	0,0674	0,0632	0,0591	0,0554	0,0518	0,0485	0,0453	0,0424
4	0,1641	0,1555	0,1473	0,1395	0,1321	0,1249	0,1181	0,1117	0,1055	0,0996
5	0,2881	0,2759	0,2640	0,2526	0,2414	0,2307	0,2203	0,2103	0,2006	0,1912
6	0,4349	0,4204	0,4060	0,3920	0,3782	0,3646	0,3514	0,3384	0,3257	0,3134
7	0,5838	0,5689	0,5541	0,5393	0,5246	0,5100	0,4956	0,4812	0,4670	0,4530
8	0,7160	0,7027	0,6892	0,6757	0,6620	0,6482	0,6343	0,6204	0,6065	0,5926
9	0,8202	0,8097	0,7988	0,7877	0,7764	0,7649	0,7531	0,7411	0,7290	0,7166

10	0,8942	0,8867	0,8788	0,8707	0,8622	0,8535	0,8445	0,8352	0,8257	0,8159
11	0,9420	0,9371	0,9319	0,9265	0,9208	0,9148	0,9085	0,9020	0,8952	0,8881
12	0,9703	0,9673	0,9642	0,9609	0,9573	0,9536	0,9496	0,9454	0,9409	0,9362
13	0,9857	0,9841	0,9824	0,9805	0,9784	0,9762	0,9739	0,9714	0,9687	0,9658
14	0,9935	0,9927	0,9918	0,9908	0,9897	0,9886	0,9873	0,9859	0,9844	0,9827
15	0,9972	0,9969	0,9964	0,9959	0,9954	0,9948	0,9941	0,9934	0,9926	0,9918
16	0,9989	0,9987	0,9985	0,9983	0,9980	0,9978	0,9974	0,9971	0,9967	0,9963
17	0,9996	0,9995	0,9994	0,9993	0,9992	0,9991	0,9989	0,9988	0,9986	0,9984
18	0,9998	0,9998	0,9998	0,9997	0,9997	0,9996	0,9996	0,9995	0,9994	0,9993
19	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998	0,9998	0,9997
20	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999
21	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	λs									
x	**8,1**	**8,2**	**8,3**	**8,4**	**8,5**	**8,6**	**8,7**	**8,8**	**8,9**	**9,0**
0	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
1	0,0028	0,0025	0,0023	0,0021	0,0019	0,0018	0,0016	0,0015	0,0014	0,0012
2	0,0127	0,0118	0,0109	0,0100	0,0093	0,0086	0,0079	0,0073	0,0068	0,0062
3	0,0396	0,0370	0,0346	0,0323	0,0301	0,0281	0,0262	0,0244	0,0228	0,0212
4	0,0941	0,0887	0,0837	0,0789	0,0744	0,0701	0,0660	0,0621	0,0584	0,0550
5	0,1823	0,1736	0,1653	0,1573	0,1496	0,1422	0,1352	0,1284	0,1219	0,1157
6	0,3013	0,2896	0,2781	0,2670	0,2562	0,2457	0,2355	0,2256	0,2160	0,2068
7	0,4391	0,4254	0,4119	0,3987	0,3856	0,3728	0,3602	0,3478	0,3357	0,3239
8	0,5786	0,5647	0,5508	0,5369	0,5231	0,5094	0,4958	0,4823	0,4689	0,4557
9	0,7041	0,6915	0,6788	0,6659	0,6530	0,6400	0,6269	0,6137	0,6006	0,5874
10	0,8058	0,7956	0,7850	0,7743	0,7634	0,7522	0,7409	0,7294	0,7178	0,7060
11	0,8807	0,8731	0,8652	0,8571	0,8487	0,8400	0,8311	0,8220	0,8126	0,8030
12	0,9313	0,9261	0,9207	0,9150	0,9091	0,9029	0,8965	0,8898	0,8829	0,8758
13	0,9628	0,9595	0,9561	0,9524	0,9486	0,9445	0,9403	0,9358	0,9311	0,9262
14	0,9810	0,9791	0,9771	0,9749	0,9726	0,9701	0,9675	0,9647	0,9617	0,9585
15	0,9908	0,9898	0,9887	0,9875	0,9862	0,9848	0,9832	0,9816	0,9798	0,9780
16	0,9958	0,9953	0,9947	0,9941	0,9934	0,9926	0,9918	0,9909	0,9899	0,9889
17	0,9982	0,9979	0,9977	0,9973	0,9970	0,9966	0,9962	0,9957	0,9952	0,9947
18	0,9992	0,9991	0,9990	0,9989	0,9987	0,9985	0,9983	0,9981	0,9978	0,9976
19	0,9997	0,9997	0,9996	0,9995	0,9995	0,9994	0,9993	0,9992	0,9991	0,9989
20	0,9999	0,9999	0,9998	0,9998	0,9998	0,9998	0,9997	0,9997	0,9996	0,9996
21	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998
22	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999
23	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	λs									
x	**9,1**	**9,2**	**9,3**	**9,4**	**9,5**	**9,6**	**9,7**	**9,8**	**9,9**	**10,0**
0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0000
1	0,0011	0,0010	0,0009	0,0009	0,0008	0,0007	0,0007	0,0006	0,0005	0,0005
2	0,0058	0,0053	0,0049	0,0045	0,0042	0,0038	0,0035	0,0033	0,0030	0,0028

3	0,0198	0,0184	0,0172	0,0160	0,0149	0,0138	0,0129	0,0120	0,0111	0,0103
4	0,0517	0,0486	0,0456	0,0429	0,0403	0,0378	0,0355	0,0333	0,0312	0,0293
5	0,1098	0,1041	0,0987	0,0935	0,0885	0,0838	0,0793	0,0750	0,0710	0,0671
6	0,1978	0,1892	0,1808	0,1727	0,1650	0,1575	0,1502	0,1433	0,1366	0,1301
7	0,3123	0,3010	0,2900	0,2792	0,2687	0,2584	0,2485	0,2388	0,2294	0,2202
8	0,4426	0,4296	0,4168	0,4042	0,3918	0,3796	0,3676	0,3558	0,3442	0,3328
9	0,5742	0,5611	0,5480	0,5349	0,5218	0,5089	0,4960	0,4832	0,4705	0,4579
10	0,6941	0,6820	0,6699	0,6576	0,6453	0,6330	0,6205	0,6081	0,5956	0,5830
11	0,7932	0,7832	0,7730	0,7626	0,7520	0,7412	0,7303	0,7193	0,7081	0,6968
12	0,8684	0,8607	0,8529	0,8448	0,8364	0,8279	0,8191	0,8101	0,8009	0,7916
13	0,9210	0,9156	0,9100	0,9042	0,8981	0,8919	0,8853	0,8786	0,8716	0,8645
14	0,9552	0,9517	0,9480	0,9441	0,9400	0,9357	0,9312	0,9265	0,9216	0,9165
15	0,9760	0,9738	0,9715	0,9691	0,9665	0,9638	0,9609	0,9579	0,9546	0,9513
16	0,9878	0,9865	0,9852	0,9838	0,9823	0,9806	0,9789	0,9770	0,9751	0,9730
17	0,9941	0,9934	0,9927	0,9919	0,9911	0,9902	0,9892	0,9881	0,9870	0,9857
18	0,9973	0,9969	0,9966	0,9962	0,9957	0,9952	0,9947	0,9941	0,9935	0,9928
19	0,9988	0,9986	0,9985	0,9983	0,9980	0,9978	0,9975	0,9972	0,9969	0,9965
20	0,9995	0,9994	0,9993	0,9992	0,9991	0,9990	0,9989	0,9987	0,9986	0,9984
21	0,9998	0,9998	0,9997	0,9997	0,9996	0,9996	0,9995	0,9995	0,9994	0,9993
22	0,9999	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998	0,9998	0,9997	0,9997
23	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
24	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
25	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	λs									
x	**10,1**	**10,2**	**10,3**	**10,4**	**10,5**	**10,6**	**10,7**	**10,8**	**01,9**	**11,0**
0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
1	0,0005	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002
2	0,0026	0,0023	0,0022	0,0020	0,0018	0,0017	0,0016	0,0014	0,0013	0,0012
3	0,0096	0,0089	0,0083	0,0077	0,0071	0,0066	0,0062	0,0057	0,0053	0,0049
4	0,0274	0,0257	0,0241	0,0225	0,0211	0,0197	0,0185	0,0173	0,0162	0,0151
5	0,0634	0,0599	0,0566	0,0534	0,0504	0,0475	0,0448	0,0423	0,0398	0,0375
6	0,1240	0,1180	0,1123	0,1069	0,1016	0,0966	0,0918	0,0872	0,0828	0,0786
7	0,2113	0,2027	0,1944	0,1863	0,1785	0,1710	0,1636	0,1566	0,1498	0,1432
8	0,3217	0,3108	0,3001	0,2896	0,2794	0,2694	0,2597	0,2502	0,2410	0,2320
9	0,4455	0,4332	0,4210	0,4090	0,3971	0,3854	0,3739	0,3626	0,3515	0,3405
10	0,5705	0,5580	0,5456	0,5331	0,5207	0,5084	0,4961	0,4840	0,4719	0,4599
11	0,6853	0,6738	0,6622	0,6505	0,6387	0,6269	0,6150	0,6031	0,5912	0,5793
12	0,7820	0,7722	0,7623	0,7522	0,7420	0,7316	0,7210	0,7104	0,6996	0,6887
13	0,8571	0,8494	0,8416	0,8336	0,8253	0,8169	0,8083	0,7995	0,7905	0,7813
14	0,9112	0,9057	0,9000	0,8940	0,8879	0,8815	0,8750	0,8682	0,8612	0,8540
15	0,9477	0,9440	0,9400	0,9359	0,9317	0,9272	0,9225	0,9177	0,9126	0,9074
16	0,9707	0,9684	0,9658	0,9632	0,9604	0,9574	0,9543	0,9511	0,9477	0,9441
17	0,9844	0,9830	0,9815	0,9799	0,9781	0,9763	0,9744	0,9723	0,9701	0,9678
18	0,9921	0,9913	0,9904	0,9895	0,9885	0,9874	0,9863	0,9850	0,9837	0,9823
19	0,9962	0,9957	0,9953	0,9948	0,9942	0,9936	0,9930	0,9923	0,9915	0,9907

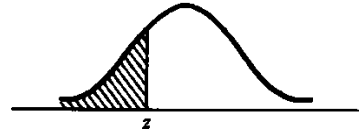
20	0,9982	0,9980	0,9978	0,9975	0,9972	0,9969	0,9966	0,9962	0,9958	0,9953
21	0,9992	0,9991	0,9990	0,9989	0,9987	0,9986	0,9984	0,9982	0,9980	0,9977
22	0,9997	0,9996	0,9996	0,9995	0,9994	0,9994	0,9993	0,9992	0,9991	0,9990
23	0,9999	0,9998	0,9998	0,9998	0,9998	0,9997	0,9997	0,9996	0,9996	0,9995
24	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998	0,9998
25	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999
26	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
27	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	λs									
x	**12,0**	**13,0**	**14,0**	**15,0**	**16,0**	**17,0**	**18,0**	**19,0**	**20,0**	**21,0**
0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
1	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
2	0,0005	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
3	0,0023	0,0011	0,0005	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000
4	0,0076	0,0037	0,0018	0,0009	0,0004	0,0002	0,0001	0,0000	0,0000	0,0000
5	0,0203	0,0107	0,0055	0,0028	0,0014	0,0007	0,0003	0,0002	0,0001	0,0000
6	0,0458	0,0259	0,0142	0,0076	0,0040	0,0021	0,0010	0,0005	0,0003	0,0001
7	0,0895	0,0540	0,0316	0,0180	0,0100	0,0054	0,0029	0,0015	0,0008	0,0004
8	0,1550	0,0998	0,0621	0,0374	0,0220	0,0126	0,0071	0,0039	0,0021	0,0011
9	0,2424	0,1658	0,1094	0,0699	0,0433	0,0261	0,0154	0,0089	0,0050	0,0028
10	0,3472	0,2517	0,1757	0,1185	0,0774	0,0491	0,0304	0,0183	0,0108	0,0063
11	0,4616	0,3532	0,2600	0,1848	0,1270	0,0847	0,0549	0,0347	0,0214	0,0129
12	0,5760	0,4631	0,3585	0,2676	0,1931	0,1350	0,0917	0,0606	0,0390	0,0245
13	0,6815	0,5730	0,4644	0,3632	0,2745	0,2009	0,1426	0,0984	0,0661	0,0434
14	0,7720	0,6751	0,5704	0,4657	0,3675	0,2808	0,2081	0,1497	0,1049	0,0716
15	0,8444	0,7636	0,6694	0,5681	0,4667	0,3715	0,2867	0,2148	0,1565	0,1111
16	0,8987	0,8355	0,7559	0,6641	0,5660	0,4677	0,3751	0,2920	0,2211	0,1629
17	0,9370	0,8905	0,8272	0,7489	0,6593	0,5640	0,4686	0,3784	0,2970	0,2270
18	0,9626	0,9302	0,8826	0,8195	0,7423	0,6550	0,5622	0,4695	0,3814	0,3017
19	0,9787	0,9573	0,9235	0,8752	0,8122	0,7363	0,6509	0,5606	0,4703	0,3843
20	0,9884	0,9750	0,9521	0,9170	0,8682	0,8055	0,7307	0,6472	0,5591	0,4710
21	0,9939	0,9859	0,9712	0,9469	0,9108	0,8615	0,7991	0,7255	0,6437	0,5577
22	0,9970	0,9924	0,9833	0,9673	0,9418	0,9047	0,8551	0,7931	0,7206	0,6405
23	0,9985	0,9960	0,9907	0,9805	0,9633	0,9367	0,8989	0,8490	0,7875	0,7160
24	0,9993	0,9980	0,9950	0,9888	0,9777	0,9594	0,9317	0,8933	0,8432	0,7822
25	0,9997	0,9990	0,9974	0,9938	0,9869	0,9748	0,9554	0,9269	0,8878	0,8377
26	0,9999	0,9995	0,9987	0,9967	0,9925	0,9848	0,9718	0,9514	0,9221	0,8826
27	0,9999	0,9998	0,9994	0,9983	0,9959	0,9912	0,9827	0,9687	0,9475	0,9175
28	1,0000	0,9999	0,9997	0,9991	0,9978	0,9950	0,9897	0,9805	0,9657	0,9436
29	1,0000	1,0000	0,9999	0,9996	0,9989	0,9973	0,9941	0,9882	0,9782	0,9626
30	1,0000	1,0000	0,9999	0,9998	0,9994	0,9986	0,9967	0,9930	0,9865	0,9758
31	1,0000	1,0000	1,0000	0,9999	0,9997	0,9993	0,9982	0,9960	0,9919	0,9848
32	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9990	0,9978	0,9953	0,9907
33	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9995	0,9988	0,9973	0,9945
34	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9994	0,9985	0,9968

35	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9992	0,9982
36	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9996	0,9990
37	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9995
38	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997
39	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999
40	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999
	λs									
x	**22,0**	**23,0**	**24,0**	**25,0**	**26,0**	**27,0**	**28,0**	**29,0**	**30,0**	
0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
2	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
6	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001
7	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002
8	0,0006	0,0003	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0006
9	0,0015	0,0008	0,0004	0,0002	0,0001	0,0001	0,0000	0,0000	0,0000	0,0015
10	0,0035	0,0020	0,0011	0,0006	0,0003	0,0002	0,0001	0,0000	0,0000	0,0035
11	0,0076	0,0044	0,0025	0,0014	0,0008	0,0004	0,0002	0,0001	0,0001	0,0076
12	0,0151	0,0091	0,0054	0,0031	0,0018	0,0010	0,0006	0,0003	0,0002	0,0151
13	0,0278	0,0174	0,0107	0,0065	0,0038	0,0022	0,0013	0,0007	0,0004	0,0278
14	0,0477	0,0311	0,0198	0,0124	0,0076	0,0046	0,0027	0,0016	0,0009	0,0477
15	0,0769	0,0520	0,0344	0,0223	0,0142	0,0088	0,0054	0,0033	0,0019	0,0769
16	0,1170	0,0821	0,0563	0,0377	0,0248	0,0160	0,0101	0,0063	0,0039	0,1170
17	0,1690	0,1228	0,0871	0,0605	0,0411	0,0274	0,0179	0,0115	0,0073	0,1690
18	0,2325	0,1748	0,1283	0,0920	0,0646	0,0445	0,0300	0,0199	0,0129	0,2325
19	0,3060	0,2377	0,1803	0,1336	0,0968	0,0687	0,0478	0,0326	0,0219	0,3060
20	0,3869	0,3101	0,2426	0,1855	0,1387	0,1015	0,0727	0,0511	0,0353	0,3869
21	0,4716	0,3894	0,3139	0,2473	0,1905	0,1436	0,1060	0,0767	0,0544	0,4716
22	0,5564	0,4723	0,3917	0,3175	0,2517	0,1952	0,1483	0,1104	0,0806	0,5564
23	0,6374	0,5551	0,4728	0,3939	0,3209	0,2559	0,1998	0,1529	0,1146	0,6374
24	0,7117	0,6346	0,5540	0,4734	0,3959	0,3242	0,2599	0,2042	0,1572	0,7117
25	0,7771	0,7077	0,6319	0,5529	0,4739	0,3979	0,3272	0,2637	0,2084	0,7771
26	0,8324	0,7723	0,7038	0,6294	0,5519	0,4744	0,3997	0,3301	0,2673	0,8324
27	0,8775	0,8274	0,7677	0,7002	0,6270	0,5509	0,4749	0,4014	0,3329	0,8775
28	0,9129	0,8726	0,8225	0,7634	0,6967	0,6247	0,5500	0,4753	0,4031	0,9129
29	0,9398	0,9085	0,8679	0,8179	0,7593	0,6935	0,6226	0,5492	0,4757	0,9398
30	0,9595	0,9360	0,9042	0,8633	0,8134	0,7553	0,6903	0,6206	0,5484	0,9595
31	0,9735	0,9564	0,9322	0,8999	0,8589	0,8092	0,7515	0,6874	0,6186	0,9735
32	0,9831	0,9711	0,9533	0,9285	0,8958	0,8546	0,8051	0,7479	0,6845	0,9831
33	0,9895	0,9813	0,9686	0,9502	0,9249	0,8918	0,8505	0,8011	0,7444	0,9895
34	0,9936	0,9882	0,9794	0,9662	0,9472	0,9213	0,8879	0,8465	0,7973	0,9936
35	0,9962	0,9927	0,9868	0,9775	0,9637	0,9441	0,9178	0,8841	0,8426	0,9962

36	0,9978	0,9956	0,9918	0,9854	0,9756	0,9612	0,9411	0,9144	0,8804	0,9978
37	0,9988	0,9974	0,9950	0,9908	0,9840	0,9737	0,9587	0,9381	0,9110	0,9988
38	0,9993	0,9985	0,9970	0,9943	0,9897	0,9825	0,9717	0,9562	0,9352	0,9993
39	0,9996	0,9992	0,9983	0,9966	0,9936	0,9887	0,9810	0,9697	0,9537	0,9996
40	0,9998	0,9996	0,9990	0,9980	0,9961	0,9928	0,9875	0,9795	0,9677	0,9998
41	0,9999	0,9998	0,9995	0,9988	0,9976	0,9955	0,9920	0,9864	0,9779	0,9999
42	1,0000	0,9999	0,9997	0,9993	0,9986	0,9973	0,9950	0,9911	0,9852	1,0000
43	1,0000	0,9999	0,9998	0,9996	0,9992	0,9984	0,9969	0,9944	0,9903	1,0000
44	1,0000	1,0000	0,9999	0,9998	0,9996	0,9991	0,9981	0,9965	0,9937	1,0000
45	1,0000	1,0000	1,0000	0,9999	0,9998	0,9995	0,9989	0,9978	0,9960	1,0000
46	1,0000	1,0000	1,0000	0,9999	0,9999	0,9997	0,9994	0,9987	0,9975	1,0000
47	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9996	0,9992	0,9985	1,0000
48	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9996	0,9991	1,0000
49	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9995	1,0000
50	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9997	1,0000
51	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	1,0000
52	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	1,0000
53	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	1,0000
54	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
55	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
56	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
57	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
58	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
59	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
60	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabla 4. Valores de la función de distribución acumulativa normal estándar.

$$P(Z \leq z) = F(z; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-t^2/2)dt$$



					α					
z	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
-4,0	0,00003	0,00003	0,00003	0,00003	0,00003	0,00003	0,00002	0,00002	0,00002	0,00002
-3,9	0,00005	0,00005	0,00004	0,00004	0,00004	0,00004	0,00004	0,00004	0,00003	0,00003
-3,8	0,00007	0,00007	0,00007	0,00006	0,00006	0,00006	0,00006	0,00005	0,00005	0,00005
-3,7	0,00011	0,00010	0,00010	0,00010	0,00009	0,00009	0,00008	0,00008	0,00008	0,00008
-3,6	0,00016	0,00015	0,00015	0,00014	0,00014	0,00013	0,00013	0,00012	0,00012	0,00011
-3,5	0,00023	0,00022	0,00022	0,00021	0,00020	0,00019	0,00019	0,00018	0,00017	0,00017
-3,4	0,00034	0,00032	0,00031	0,00030	0,00029	0,00028	0,00027	0,00026	0,00025	0,00024
-3,3	0,00048	0,00047	0,00045	0,00043	0,00042	0,00040	0,00039	0,00038	0,00036	0,00035
-3,2	0,00069	0,00066	0,00064	0,00062	0,00060	0,00058	0,00056	0,00054	0,00052	0,00050
-3,1	0,00097	0,00094	0,00090	0,00087	0,00084	0,00082	0,00079	0,00076	0,00074	0,00071
-3,0	0,00135	0,00131	0,00126	0,00122	0,00118	0,00114	0,00111	0,00107	0,00104	0,00100
-2,9	0,00187	0,00181	0,00175	0,00169	0,00164	0,00159	0,00154	0,00149	0,00144	0,00139
-2,8	0,00256	0,00248	0,00240	0,00233	0,00226	0,00219	0,00212	0,00205	0,00199	0,00193
-2,7	0,00347	0,00336	0,00326	0,00317	0,00307	0,00298	0,00289	0,00280	0,00272	0,00264

-2,6	0,00466	0,00453	0,00440	0,00427	0,00415	0,00402	0,00391	0,00379	0,00368	0,00357
-2,5	0,00621	0,00604	0,00587	0,00570	0,00554	0,00539	0,00523	0,00508	0,00494	0,00480
-2,4	0,00820	0,00798	0,00776	0,00755	0,00734	0,00714	0,00695	0,00676	0,00657	0,00639
-2,3	0,01072	0,01044	0,01017	0,00990	0,00964	0,00939	0,00914	0,00889	0,00866	0,00842
-2,2	0,01390	0,01355	0,01321	0,01287	0,01255	0,01222	0,01191	0,01160	0,01130	0,01101
-2,1	0,01786	0,01743	0,01700	0,01659	0,01618	0,01578	0,01539	0,01500	0,01463	0,01426
-2,0	0,02275	0,02222	0,02169	0,02118	0,02068	0,02018	0,01970	0,01923	0,01876	0,01831
-1,9	0,02872	0,02807	0,02743	0,02680	0,02619	0,02559	0,02500	0,02442	0,02385	0,02330
-1,8	0,03593	0,03515	0,03438	0,03362	0,03288	0,03216	0,03144	0,03074	0,03005	0,02938
-1,7	0,04457	0,04363	0,04272	0,04182	0,04093	0,04006	0,03920	0,03836	0,03754	0,03673
-1,6	0,05480	0,05370	0,05262	0,05155	0,05050	0,04947	0,04846	0,04746	0,04648	0,04551
-1,5	0,06681	0,06552	0,06426	0,06301	0,06178	0,06057	0,05938	0,05821	0,05705	0,05592
-1,4	0,08076	0,07927	0,07780	0,07636	0,07493	0,07353	0,07215	0,07078	0,06944	0,06811
-1,3	0,09680	0,09510	0,09342	0,09176	0,09012	0,08851	0,08691	0,08534	0,08379	0,08226
-1,2	0,11507	0,11314	0,11123	0,10935	0,10749	0,10565	0,10383	0,10204	0,10027	0,09853
-1,1	0,13567	0,13350	0,13136	0,12924	0,12714	0,12507	0,12302	0,12100	0,11900	0,11702
-1,0	0,15866	0,15625	0,15386	0,15151	0,14917	0,14686	0,14457	0,14231	0,14007	0,13786
-0,9	0,18406	0,18141	0,17879	0,17619	0,17361	0,17106	0,16853	0,16602	0,16354	0,16109
-0,8	0,21186	0,20897	0,20611	0,20327	0,20045	0,19766	0,19489	0,19215	0,18943	0,18673
-0,7	0,24196	0,23885	0,23576	0,23270	0,22965	0,22663	0,22363	0,22065	0,21770	0,21476
-0,6	0,27425	0,27093	0,26763	0,26435	0,26109	0,25785	0,25463	0,25143	0,24825	0,24510
-0,5	0,30854	0,30503	0,30153	0,29806	0,29460	0,29116	0,28774	0,28434	0,28096	0,27760
z	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**

-0,4	0,34458	0,34090	0,33724	0,33360	0,32997	0,32636	0,32276	0,31918	0,31561	0,31207
-0,3	0,38209	0,37828	0,37448	0,37070	0,36693	0,36317	0,35942	0,35569	0,35197	0,34827
-0,2	0,42074	0,41683	0,41294	0,40905	0,40517	0,40129	0,39743	0,39358	0,38974	0,38591
-0,1	0,46017	0,45620	0,45224	0,44828	0,44433	0,44038	0,43644	0,43251	0,42858	0,42465
-0,0	0,50000	0,49601	0,49202	0,48803	0,48405	0,48006	0,47608	0,47210	0,46812	0,46414
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327

1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

Tabla 5. Diez mil dígitos aleatorizados

	00-04	**05-09**	**10-14**	**15-19**	**20-24**	**25-29**	**30-34**	**35-39**	**40-44**	**45- 49**
00	88758	66605	33843	43623	62774	25517	09560	41880	85126	60755
01	35661	42832	16240	77410	20686	26656	59698	86241	13152	49187
02	26335	03771	46115	88133	40721	06787	95962	60841	91788	86386
03	60826	74718	56527	29508	91975	13695	25215	72237	06337	73439
04	95044	99896	13763	31764	93970	60987	14692	71039	34165	21297
05	83746	47694	06143	42741	38338	97694	69300	99864	19641	15083
06	27998	42562	63402	10056	81668	48744	08400	83124	19896	18805
07	82685	32323	74625	14510	85927	28017	80588	14756	54937	76379
08	18386	13862	10988	04197	18770	72757	71418	81133	69503	44037
09	21717	13141	22707	68165	58440	19187	08421	23872	03036	34208
10	18446	83052	31842	08634	15887	86070	08464	20565	74390	36541
11	66027	75177	47398	66423	70160	16232	67343	36205	50036	59411
12	51420	96779	54309	87456	78967	79638	68869	49062	02196	55109
13	27045	62626	73159	91149	96509	44204	92237	29969	49315	11804
14	13094	17725	14103	00067	68843	63565	93578	24756	10814	15185
15	92382	62518	17752	53163	63852	44840	02592	88572	03107	90169
16	16215	50809	49326	77232	90155	69955	93892	70445	00906	57002
17	09342	14528	64727	71403	84156	34083	35613	35670	10549	07468
18	38148	79001	03509	79424	39625	73315	18811	86230	99682	82896
19	23689	19997	72382	15247	80205	58090	43804	94548	82693	22799
20	25407	37726	73099	51057	68733	75768	77991	72641	95386	70138
21	25349	69456	19693	85568	93876	18661	69018	10332	83137	88257
22	02322	77491	56095	03055	37738	18216	81781	32245	84081	18436
23	15072	33261	99219	43307	39239	79712	94753	41450	30944	53912
24	27002	31036	85278	74547	84 809	36252	09373	69471	15606	77209
25	66181	83316	40386	54316	29505	86032	34563	93204	72973	90760
26	09779	01822	45537	13128	51128	82703	75350	25179	86104	40638
27	10791	07706	87481	26107	24857	27805	42710	63471	08804	23455
28	74833	55767	31312	76611	67389	04691	39687	13596	88730	86850
29	17583	24038	83701	28570	63561	00098	60784	76098	84217	34997
30	45601	46977	39325	09286	41133	34031	94867	11849	75171	57682
31	60683	33112	65995	64203	18070	65437	13624	90896	80945	71987
32	29956	81169	18877	15296	94368	16317	34239	03643	66081	12242
33	91713	84235	75296	69875	82414	05197	66596	13083	46278	73498
34	85704	86588	82837	67822	95963	83021	90732	32661	64751	83903
35	17921	26111	35373	86494	48266	01888	65735	05315	79328	13367
36	13929	71341	80488	89827	48277	07229	71953	16128	65074	28782
37	03248	18880	21667	01311	61806	80201	47889	83052	31029	06023
38	50583	17972	12690	00452	93766	16414	01212	27964	02766	28786
39	10636	46975	09449	45986	34672	46916	63881	83117	53947	95218
40	43896	41278	42205	10425	66560	59967	90139	73563	29875	79033
41	76714	80963	74907	16890	15492	27489	06067	22287	19760	13056

86	20514	11081	51131	56469	33947	77703	35679	45774	06776	67062
87	96763	56249	81243	62416	84451	14696	38195	70435	45948	67690
88	49439	61075	31558	59740	52759	55323	95226	01385	20158	54054
89	16294	50548	71317	32168	86071	47314	65393	56367	46910	51269
90	31381	94301	79273	32843	05862	36211	93960	00671	67631	23952
91	98032	87203	03227	66021	99666	98368	39222	36056	81992	20121
92	40700	31826	94774	11366	81391	33602	69608	84119	93204	26825
93	68692	66849	29366	77540	14978	06508	10824	65416	23629	63029
94	19047	10784	19607	20296	31804	72984	60060	50353	23260	58909
95	82867	69266	50733	62630	00956	61500	89913	30049	82321	62367
	00-04	**05-09**	**10-14**	**15-19**	**20-24**	**25-29**	**30-34**	**35-39**	**40-44**	**45- 49**
96	26528	28928	52600	72997	80943	04084	86662	90025	14360	64867
97	51166	00607	49962	30724	81707	14548	25844	47336	57492	02207
98	97245	15440	55182	15368	85136	98869	33712	95152	50973	98658
99	54998	88830	95639	45 i 04	72676	28220	82576	57381	34438	24565
	50-54	**55-59**	**60-64**	**65-69**	**70-74**	**75-79**	**80-84**	**85-89**	**90-94**	**95-99**
00	70896	44520	64720	49898	78088	76740	47460	83150	78905	59870
01	56809	42909	25853	47624	29486	14196	75841	00393	42390	24847
02	66109	84775	07515	49949	61482	91836	48126	80778	21302	24975
03	18071	36263	14053	52526	44347	04923	68100	57805	19521	15345
04	98732	15120	91754	12657	74675	78500	01247	49719	47635	55514
05	36075	83967	22268	77971	31169	68584	21336	72541	66959	39708
06	04110	45061	78062	18911	27855	09419	56459	00695	70323	04538
07	75658	58509	24479	10202	13150	95946	55087	38398	18718	95561
08	87403	19142	27208	35149	34889	27003	14181	44813	17784	41036
09	00005	52142	65021	64438	69610	12154	98422	65320	79996	01935
10	43674	47103	48614	70823	78252	82403	93424	05236	54588	27757
11	68597	68874	35567	98463	99671	05634	81533	47406	17228	44455
12	91874	70208	06308	40719	02772	69589	79936	07514	44950	35190
13	73854	19470	53014	29375	62256	77488	74388	53949	49607	19816
14	65926	34117	55344	68155	38099	56009	03513	05926	35584	42328
15	40005	35246	49440	40295	44390	83043	26090	80201	02934	49260
16	46686	29890	14821	69783	34733	11803	64845	32065	14527	38702
17	02717	61518	39583	72863	50707	96115	07416	05041	36756	61065
18	17048	22281	35573	28944	96889	51823	57268	03866	27658	91950
19	75304	53248	42151	93928	17343	88322	28683	11252	10355	65175
20	97844	62947	62230	30500	92816	85232	27222	91701	11057	83257
21	07611	71163	82212	20653	21499	51496	40715	78952	33029	64207
22	47744	04603	44522	62783	39347	72310	41460	31052	40814	94297
23	54293	43576	88116	67416	34908	15238	40561	73940	56850	31078
24	67556	93979	73363	00300	11217	74405	18937	79000	68834	48307
25	86581	73041	95809	73986	49408	53316	90841	73808	53421	82315
26	28020	86282	83365	76600	11261	74354	20968	60770	12141	09539
27	42578	32471	37840	30872	75074	79027	57813	62831	54715	26693
28	47290	15997	86163	10571	81911	92124	92971	80860	41012	58666

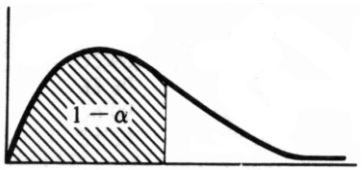
29	24856	63911	13221	77028	06573	33667	30732	47280	12926	67276
30	16352	24836	60799	76281	83402	44709	78930	82969	84468	36910
31	89060	79852	97854	28324	39638	86936	06702	74304	39873	19496
32	07637	30412	04921	26471	09605	07355	20466	49793	40539	21077
33	37711	47786	37468	31963	16908	50283	80884	08252	72655	58926
34	82994	53232	58202	73318	62471	49650	15888	73370	98748	69181
35	31722	67288	12110	04776	15168	68862	92347	90789	66961	04162
36	93819	78050	19364	38037	25706	90879	05215	00260	14426	88207
37	65557	24496	04713	23688	26623	41356	47049	60676	72236	01214
38	88001	91382	05129	36041	10257	55558	89979	58061	28957	10701
39	96648	70303	18191	62404	26558	92804	15415	02865	52449	78509
40	04118	51573	59356	02426	35010	37104	98316	44602	96478	08433
41	19317	27753	39431	26996	04465	69695	61374	06317	42225	62025
42	37182	91221	17307	68507	85725	81898	22588	22241	80337	89033
43	82990	03607	29560	60413	59743	75000	03806	13741	79671	25416
	50-54	**55-59**	**60-64**	**65-69**	**70-74**	**75-79**	**80-84**	**85-89**	**90-94**	**95-99**
44	97294	21991	11217	98087	79124	52275	31088	32085	23089	21498
45	86771	69504	13345	42544	59616	07867	78717	82840	74669	21515
46	26046	55559	12200	95106	56496	76662	44880	89457	84209	01332
47	39689	59999	92290	79024	70271	93352	90272	94495	26842	54477
48	83265	89573	01437	43786	52986	49041	17952	35035	88985	84671
49	15128	35791	11296	45319	06330	82027	90808	54351	43091	30387
50	58649	85086	16502	97541	76611	94229	34987	86718	87208	05426
51	97306	52449	55596	66739	36525	97563	29469	31235	79276	10831
52	09942	79344	78160	11015	55777	22047	57615	15717	86239	36578
53	83842	28631	74893	47911	92170	38181	30416	54860	44120	73031
54	73778	30395	20163	76111	13712	33449	99224	18206	51418	70006
55	88381	56550	47467	59663	61117	39716	32927	06168	06217	45477
56	31044	21404	15968	21357	30772	81482	38807	67231	84283	63552
57	00909	63837	91328	81106	11740	50193	86806	21931	18054	49601
58	69882	37028	41732	37425	80832	03320	20690	32653	90145	03029
59	26059	78324	22501	73825	16927	31545	15695	74216	98372	28547
60	38573	98078	38982	33078	93524	45606	53463	20391	81637	37269
61	70624	00063	81455	16924	12848	23801	55481	78978	26795	10553
62	49806	23976	05640	29804	38988	25024	76951	02341	63219	75864
63	05461	67523	48316	14613	08541	35231	38312	14969	67279	50502
64	76582	62153	53801	51219	30424	32599	49099	83959	68408	20147
65	16660	80470	75062	75588	24384	27874	20018	11428	32265	07692
66	60166	42424	97470	88451	81270	80070	72959	26220	59939	31127
67	28953	03272	31460	41691	57736	72052	22762	96323	27616	53123
68	47536	86439	95210	96386	38704	15484	07426	70675	06888	81203
69	73457	26657	36983	72410	30244	97711	25652	09373	66218	64077
70	11190	66193	66287	09116	48140	37669	02932	50799	17255	06181
71	57062	78964	44455	14036	36098	40773	11688	33150	07459	36127
72	99624	67254	67302	18991	97687	54099	94884	42283	63258	50651

73	97521	83669	85968	16135	30133	51312	17831	75016	80278	68953
74	40273	04838	13661	64757	17461	78085	60094	27010	80945	66439
75	57260	06176	49963	29760	69546	61336	39429	41985	18572	98128
76	03451	47098	63495	71227	79304	29753	99131	18419	71791	81515
77	62331	20492	15393	84270	24396	32962	21632	92965	38670	44923
78	32290	51079	06512	38806	93327	80086	19088	59887	98416	24918
79	28014	80428	92853	31333	32648	16734	43418	90124	15086	48444
80	18950	16091	29543	65817	07002	73115	94115	20271	50250	25061
81	17403	69503	01866	13049	07263	13039	83844	80143	39048	62654
82	27999	50489	66613	21843	71746	65868	16208	46781	93402	12323
83	87076	53174	12165	84495	47947	60706	64034	31635	65169	93070
84	89044	45974	14524	46906	26052	51851	84197	61694	57429	63395
85	98048	64400	24705	75711	36232	57624	41424	77366	52790	84705
86	09345	12956	49770	80311	32319	48238	16952	92088	51222	82865
87	07086	77628	76195	47584	62411	40397	71857	54823	26536	56792
88	93128	25657	46872	11206	06831	87944	97914	64670	45760	34353
89	85137	70964	29947	27795	25547	37682	96105	26848	09389	64326
90	32798	39024	13814	98546	46585	84108	74603	94812	73968	68766
91	62496	26371	89880	52078	47781	95260	83464	65942	91761	53727
92	62707	81825	40987	97656	89714	52177	23778	07482	91678	40128
	50-54	**55-59**	**60-64**	**65-69**	**70-74**	**75-79**	**80-84**	**85-89**	**90-94**	**95-99**
93	05500	28982	86124	19554	80818	94935	61924	31828	79369	23507
94	79476	31445	59498	85132	24582	26024	24002	63718	79164	43556
95	10653	29954	97568	91541	33139	84525	72271	02546	64818	14381
96	30524	06495	00886	40666	68574	49574	19705	16429	90981	08103
97	69050	22019	74066	14500	14506	06423	38332	34191	82663	85323
98	27908	78802	63446	07674	98871	63831	72449	42705	26513	19883
99	64520	16618	47409	19574	78136	46047	01277	79146	95759	36781

Tabla 6. Valores de cuantiles de la distribución de Chi-cuadrada.

$$P(X \leq x_{1-\alpha,\,\upsilon}) = F(x_{1-\alpha\,;\,\upsilon}) = \frac{1}{\Gamma(\upsilon/2)2^{\upsilon/2}} \int_0^{x_{1-\alpha,\upsilon}} t^{\upsilon/2-1}\, exp(-t/2)dt = 1-\alpha$$

$$P(X \leq x_{1-\alpha,\,\upsilon}) = F(x_{1-\alpha\,;\,\upsilon}) = \frac{1}{\Gamma(\upsilon/2)2^{\upsilon/2}} \int_0^{x_{1-\alpha,\upsilon}} t^{\upsilon/2-1}\, exp(-t/2)dt = 1-\alpha$$

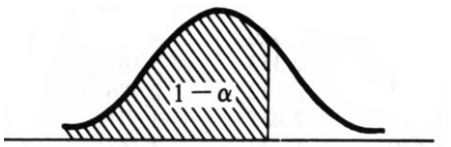


$x_{1\text{-}\alpha,\,\upsilon}$

υ	$x_{0,005}$	$x_{0,010}$	$x_{0,025}$	$x_{0,050}$	$x_{0,100}$	$x_{0,900}$	$x_{0,950}$	$x_{0,975}$	$x_{0,990}$	$x_{0,995}$
1	0,00	0,00	0,00	0,00	0,02	2,71	3,84	5,02	6,63	7,88
2	0,01	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	10,60
3	0,07	0,11	0,22	0,35	0,58	6,25	7,81	9,35	11,34	12,84
4	0,21	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	14,86
5	0,41	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	16,75
6	0,68	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81	18,55
7	0,99	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	21,95
9	1,73	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	25,19
11	2,60	3,05	3,82	4,57	5,58	17,28	19,68	21,92	24,72	26,76
12	3,07	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	28,30
13	3,57	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	29,82
14	4,07	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	31,32
15	4,60	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	32,80
16	5,14	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00	34,27
17	5,70	6,41	7,56	8,67	10,09	24,77	27,59	30,19	33,41	35,72
18	6,26	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,81	37,16
19	6,84	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	38,58
20	7,43	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	40,00
21	8,03	8,90	10,28	11,59	13,24	29,62	32,67	35,48	38,93	41,40
22	8,64	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29	42,80
23	9,26	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64	44,18
24	9,89	10,86	12,40	13,85	15,66	33,20	36,42	39,36	42,98	45,56
25	10,52	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31	46,93
26	11,16	12,20	13,84	15,38	17,29	35,56	38,89	41,92	45,64	48,29
27	11,81	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96	49,64

Tabla 7. Valores cuantiles de la distribución *t* de Student

$$P(T \leq t_{1-\alpha,\upsilon}) = \frac{\Gamma[(\upsilon+1)/2]}{\sqrt{\pi\upsilon}\Gamma(\upsilon/2)} \int_{-\infty}^{t_{1-\alpha,\upsilon}} [1 + (t^2/\upsilon)]^{-(\upsilon+1)/2} dt = 1 - \alpha$$



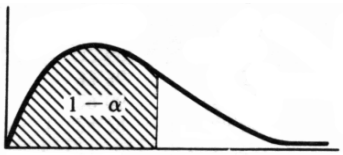
ν	$t_{0,001}$	$t_{0,005}$	$t_{0,010}$	$t_{0,025}$	$t_{0,050}$	$t_{0,100}$	$t_{0,200}$
1	-	-63,657	-31,821	-12,706	-6,314	-3,078	-1,376
2	-22,327	-9,925	-6,965	-4,303	-2,920	-1,886	-1,061
3	-10,215	-5,841	-4,541	-3,182	-2,353	-1,638	-0,978
4	-7,173	-4,604	-3,747	-2,776	-2,132	-1,533	-0,941
5	-5,893	-4,032	-3,365	-2,571	-2,015	-1,476	-0,920
6	-5,208	-3,707	-3,143	-2,447	-1,943	-1,440	-0,906
7	-4,785	-3,499	-2,998	-2,365	-1,895	-1,415	-0,896
8	-4,501	-3,355	-2,896	-2,306	-1,860	-1,397	-0,889
9	-4,297	-3,250	-2,821	-2,262	-1,833	-1,383	-0,883
10	-4,144	-3,169	-2,764	-2,228	-1,812	-1,372	-0,879
11	-4,025	-3,106	-2,718	-2,201	-1,796	-1,363	-0,876
12	-3,930	-3,055	-2,681	-2,179	-1,782	-1,356	-0,873
13	-3,852	-3,012	-2,650	-2,160	-1,771	-1,350	-0,870
14	-3,787	-2,977	-2,624	-2,145	-1,761	-1,345	-0,868
15	-3,733	-2,947	-2,602	-2,131	-1,753	-1,341	-0,866
16	-3,686	-2,921	-2,583	-2,120	-1,746	-1,337	-0,865
17	-3,646	-2,898	-2,567	-2,110	-1,740	-1,333	-0,863
18	-3,610	-2,878	-2,552	-2,101	-1,734	-1,330	-0,862
19	-3,579	-2,861	-2,539	-2,093	-1,729	-1,328	-0,861
20	-3,552	-2,845	-2,528	-2,086	-1,725	-1,325	-0,860
21	-3,527	-2,831	-2,518	-2,080	-1,721	-1,323	-0,859
22	-3,505	-2,819	-2,508	-2,074	-1,717	-1,321	-0,858
23	-3,485	-2,807	-2,500	-2,069	-1,714	-1,319	-0,858
24	-3,467	-2,797	-2,492	-2,064	-1,711	-1,318	-0,857
25	-3,450	-2,787	-2,485	-2,060	-1,708	-1,316	-0,856
26	-3,435	-2,779	-2,479	-2,056	-1,706	-1,315	-0,856
27	-3,421	-2,771	-2,473	-2,052	-1,703	-1,314	-0,855
28	-3,408	-2,763	-2,467	-2,048	-1,701	-1,313	-0,855
29	-3,396	-2,756	-2,462	-2,045	-1,699	-1,311	-0,854
30	-3,385	-2,750	-2,457	-2,042	-1,697	-1,310	-0,854
35	-3,340	-2,724	-2,438	-2,030	-1,690	-1,306	-0,852
40	-3,307	-2,704	-2,423	-2,021	-1,684	-1,303	-0,851

45	-3,281	-2,690	-2,412	-2,014	-1,679	-1,301	-0,850
50	-3,261	-2,678	-2,403	-2,009	-1,676	-1,299	-0,849
60	-3,232	-2,660	-2,390	-2,000	-1,671	-1,296	-0,848
70	-3,211	-2,648	-2,381	-1,994	-1,667	-1,294	-0,847
80	-3,195	-2,639	-2,374	-1,990	-1,664	-1,292	-0,846
90	-3,183	-2,632	-2,368	-1,987	-1,662	-1,291	-0,846
100	-3,174	-2,626	-2,364	-1,984	-1,660	-1,290	-0,845
200	-3,131	-2,601	-2,345	-1,972	-1,653	-1,286	-0,843
300	-3,118	-2,592	-2,339	-1,968	-1,650	-1,284	-0,843
400	-3,111	-2,588	-2,336	-1,966	-1,649	-1,284	-0,843
500	-3,107	-2,586	-2,334	-1,965	-1,648	-1,283	-0,842
600	-3,104	-2,584	-2,333	-1,964	-1,647	-1,283	-0,842
700	-3,102	-2,583	-2,332	-1,963	-1,647	-1,283	-0,842
800	-3,100	-2,582	-2,331	-1,963	-1,647	-1,283	-0,842
900	-3,099	-2,581	-2,330	-1,963	-1,647	-1,282	-0,842
1000	-3,098	-2,581	-2,330	-1,962	-1,646	-1,282	-0,842
v	$t_{0,800}$	$t_{0,900}$	$t_{0,950}$	$t_{0,975}$	$t_{0,990}$	$t_{0,995}$	$t_{0,999}$
1	1,376	3,078	6,314	12,706	31,821	63,657	318,309
2	1,061	1,886	2,920	4,303	6,965	9,925	22,327
3	0,978	1,638	2,353	3,182	4,541	5,841	10,215
4	0,941	1,533	2,132	2,776	3,747	4,604	7,173
5	0,920	1,476	2,015	2,571	3,365	4,032	5,893
6	0,906	1,440	1,943	2,447	3,143	3,707	5,208
7	0,896	1,415	1,895	2,365	2,998	3,499	4,785
8	0,889	1,397	1,860	2,306	2,896	3,355	4,501
9	0,883	1,383	1,833	2,262	2,821	3,250	4,297
10	0,879	1,372	1,812	2,228	2,764	3,169	4,144
11	0,876	1,363	1,796	2,201	2,718	3,106	4,025
12	0,873	1,356	1,782	2,179	2,681	3,055	3,930
13	0,870	1,350	1,771	2,160	2,650	3,012	3,852
14	0,868	1,345	1,761	2,145	2,624	2,977	3,787
15	0,866	1,341	1,753	2,131	2,602	2,947	3,733
16	0,865	1,337	1,746	2,120	2,583	2,921	3,686
17	0,863	1,333	1,740	2,110	2,567	2,898	3,646
18	0,862	1,330	1,734	2,101	2,552	2,878	3,610
19	0,861	1,328	1,729	2,093	2,539	2,861	3,579
20	0,860	1,325	1,725	2,086	2,528	2,845	3,552
21	0,859	1,323	1,721	2,080	2,518	2,831	3,527
22	0,858	1,321	1,717	2,074	2,508	2,819	3,505
23	0,858	1,319	1,714	2,069	2,500	2,807	3,485
24	0,857	1,318	1,711	2,064	2,492	2,797	3,467
25	0,856	1,316	1,708	2,060	2,485	2,787	3,450
26	0,856	1,315	1,706	2,056	2,479	2,779	3,435
27	0,855	1,314	1,703	2,052	2,473	2,771	3,421

28	0,855	1,313	1,701	2,048	2,467	2,763	3,408
29	0,854	1,311	1,699	2,045	2,462	2,756	3,396
30	0,854	1,310	1,697	2,042	2,457	2,750	3,385
35	0,852	1,306	1,690	2,030	2,438	2,724	3,340
40	0,851	1,303	1,684	2,021	2,423	2,704	3,307
45	0,850	1,301	1,679	2,014	2,412	2,690	3,281
50	0,849	1,299	1,676	2,009	2,403	2,678	3,261
60	0,848	1,296	1,671	2,000	2,390	2,660	3,232
70	0,847	1,294	1,667	1,994	2,381	2,648	3,211
80	0,846	1,292	1,664	1,990	2,374	2,639	3,195
90	0,846	1,291	1,662	1,987	2,368	2,632	3,183
100	0,845	1,290	1,660	1,984	2,364	2,626	3,174
200	0,843	1,286	1,653	1,972	2,345	2,601	3,131
300	0,843	1,284	1,650	1,968	2,339	2,592	3,118
400	0,843	1,284	1,649	1,966	2,336	2,588	3,111
500	0,842	1,283	1,648	1,965	2,334	2,586	3,107
600	0,842	1,283	1,647	1,964	2,333	2,584	3,104
700	0,842	1,283	1,647	1,963	2,332	2,583	3,102
800	0,842	1,283	1,647	1,963	2,331	2,582	3,100
900	0,842	1,282	1,647	1,963	2,330	2,581	3,099
1000	0,842	1,282	1,646	1,962	2,330	2,581	3,098

Tabla 8. Valores de cuantiles de la distribución *F*

$$P(F \leq f_{1-\alpha,v_1,v_2}) = \frac{\Gamma[(v_1+v_2)/2]{v_1}^{v_1/2}{v_2}^{v_2/2}}{\Gamma(v_1/2)\ \Gamma(v_2/2)} \int_{-\infty}^{t_{1-\alpha,v1,v2}} t^{(v_1-2)/2}(v_2+v_1t)^{(v_2+v_1)/2}dt = 1-\alpha$$



*f*1-*α*. *v*1.

1-α = 0,9

v_1 = grados de libertad del numerador

v2	1	2	3	4	5	6	7	8	9	10
1	39,86	49,50	53,59	55,83	57,24	58,20	8,91	59,44	59,86	60,19
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86

27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82
35	2,85	2,46	2,25	2,11	2,02	1,95	1,90	1,85	1,82	1,79
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76
45	2,82	2,42	2,21	2,07	1,98	1,91	1,85	1,81	1,77	1,74
50	2,81	2,41	2,20	2,06	1,97	1,90	1,84	1,80	1,76	1,73
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71
70	2,78	2,38	2,16	2,03	1,93	1,86	1,80	1,76	1,72	1,69
80	2,77	2,37	2,15	2,02	1,92	1,85	1,79	1,75	1,71	1,68
90	2,76	2,36	2,15	2,01	1,91	1,84	1,78	1,74	1,70	1,67
100	2,76	2,36	2,14	2,00	1,91	1,83	1,78	1,73	1,69	1,66
200	2,73	2,33	2,11	1,97	1,88	1,80	1,75	1,70	1,66	1,63
300	2,72	2,32	2,10	1,96	1,87	1,79	1,74	1,69	1,65	1,62
400	2,72	2,32	2,10	1,96	1,86	1,79	1,73	1,69	1,65	1,61
500	2,72	2,31	2,09	1,96	1,86	1,79	1,73	1,68	1,64	1,61
600	2,71	2,31	2,09	1,95	1,86	1,78	1,73	1,68	1,64	1,61
700	2,71	2,31	2,09	1,95	1,86	1,78	1,73	1,68	1,64	1,61
800	2,71	2,31	2,09	1,95	1,85	1,78	1,72	1,68	1,64	1,61
900	2,71	2,31	2,09	1,95	1,85	1,78	1,72	1,68	1,64	1,61
1000	2,71	2,31	2,09	1,95	1,85	1,78	1,72	1,68	1,64	1,61

1-α = 0,9

v_1 = grados de libertad del numerador

v_2	**11**	**12**	**15**	**20**	**25**	**30**	**40**	**50**	**100**	**1000**
1	60,47	60,71	61,22	61,74	62,05	62,26	62,53	62,69	63,01	63,30
2	9,40	9,41	9,42	9,44	9,45	9,46	9,47	9,47	9,48	9,49
3	5,22	5,22	5,20	5,18	5,17	5,17	5,16	5,15	5,14	5,13
4	3,91	3,90	3,87	3,84	3,83	3,82	3,80	3,80	3,78	3,76
5	3,28	3,27	3,24	3,21	3,19	3,17	3,16	3,15	3,13	3,11
6	2,92	2,90	2,87	2,84	2,81	2,80	2,78	2,77	2,75	2,72
7	2,68	2,67	2,63	2,59	2,57	2,56	2,54	2,52	2,50	2,47
8	2,52	2,50	2,46	2,42	2,40	2,38	2,36	2,35	2,32	2,30
9	2,40	2,38	2,34	2,30	2,27	2,25	2,23	2,22	2,19	2,16
10	2,30	2,28	2,24	2,20	2,17	2,16	2,13	2,12	2,09	2,06
11	2,23	2,21	2,17	2,12	2,10	2,08	2,05	2,04	2,01	1,98
12	2,17	2,15	2,10	2,06	2,03	2,01	1,99	1,97	1,94	1,91
13	2,12	2,10	2,05	2,01	1,98	1,96	1,93	1,92	1,88	1,85
14	2,07	2,05	2,01	1,96	1,93	1,91	1,89	1,87	1,83	1,80
15	2,04	2,02	1,97	1,92	1,89	1,87	1,85	1,83	1,79	1,76
16	2,01	1,99	1,94	1,89	1,86	1,84	1,81	1,79	1,76	1,72

17	1,98	1,96	1,91	1,86	1,83	1,81	1,78	1,76	1,73	1,69
18	1,95	1,93	1,89	1,84	1,80	1,78	1,75	1,74	1,70	1,66
19	1,93	1,91	1,86	1,81	1,78	1,76	1,73	1,71	1,67	1,64
20	1,91	1,89	1,84	1,79	1,76	1,74	1,71	1,69	1,65	1,61
21	1,90	1,87	1,83	1,78	1,74	1,72	1,69	1,67	1,63	1,59
22	1,88	1,86	1,81	1,76	1,73	1,70	1,67	1,65	1,61	1,57
23	1,87	1,84	1,80	1,74	1,71	1,69	1,66	1,64	1,59	1,55
24	1,85	1,83	1,78	1,73	1,70	1,67	1,64	1,62	1,58	1,54
25	1,84	1,82	1,77	1,72	1,68	1,66	1,63	1,61	1,56	1,52
26	1,83	1,81	1,76	1,71	1,67	1,65	1,61	1,59	1,55	1,51
27	1,82	1,80	1,75	1,70	1,66	1,64	1,60	1,58	1,54	1,50
28	1,81	1,79	1,74	1,69	1,65	1,63	1,59	1,57	1,53	1,48
29	1,80	1,78	1,73	1,68	1,64	1,62	1,58	1,56	1,52	1,47
30	1,79	1,77	1,72	1,67	1,63	1,61	1,57	1,55	1,51	1,46
35	1,76	1,74	1,69	1,63	1,60	1,57	1,53	1,51	1,47	1,42
40	1,74	1,71	1,66	1,61	1,57	1,54	1,51	1,48	1,43	1,38
45	1,72	1,70	1,64	1,58	1,55	1,52	1,48	1,46	1,41	1,36
50	1,70	1,68	1,63	1,57	1,53	1,50	1,46	1,44	1,39	1,33
60	1,68	1,66	1,60	1,54	1,50	1,48	1,44	1,41	1,36	1,30
70	1,66	1,64	1,59	1,53	1,49	1,46	1,42	1,39	1,34	1,27
80	1,65	1,63	1,57	1,51	1,47	1,44	1,40	1,38	1,32	1,25
90	1,64	1,62	1,56	1,50	1,46	1,43	1,39	1,36	1,30	1,24
100	1,64	1,61	1,56	1,49	1,45	1,42	1,38	1,35	1,29	1,22
200	1,60	1,58	1,52	1,46	1,41	1,38	1,34	1,31	1,24	1,16
300	1,59	1,57	1,51	1,45	1,40	1,37	1,32	1,29	1,22	1,13
400	1,59	1,56	1,50	1,44	1,39	1,36	1,32	1,29	1,21	1,12
500	1,58	1,56	1,50	1,44	1,39	1,36	1,31	1,28	1,21	1,11
600	1,58	1,56	1,50	1,43	1,39	1,36	1,31	1,28	1,20	1,10
700	1,58	1,56	1,50	1,43	1,39	1,35	1,31	1,28	1,20	1,09
800	1,58	1,55	1,50	1,43	1,39	1,35	1,31	1,28	1,20	1,09
900	1,58	1,55	1,49	1,43	1,38	1,35	1,31	1,27	1,20	1,09
1000	1,58	1,55	1,49	1,43	1,38	1,35	1,30	1,27	1,20	1,08

$1-\alpha = 0,95$

v_1 = grados de libertad del numerador

v_2	1	2	3	4	5	6	7	8	9	10
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96

5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
45	4,06	3,20	2,81	2,58	2,42	2,31	2,22	2,15	2,10	2,05
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88
300	3,87	3,03	2,63	2,40	2,24	2,13	2,04	1,97	1,91	1,86
400	3,86	3,02	2,63	2,39	2,24	2,12	2,03	1,96	1,90	1,85
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85
600	3,86	3,01	2,62	2,39	2,23	2,11	2,02	1,95	1,90	1,85
700	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84
800	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84
900	3,85	3,01	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84
1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84

$1-\alpha = 0,95$

v_1 = grados de libertad del numerador

v_2	11	12	15	20	25	30	40	50	100	1000
1	242,98	243,91	245,95	248,01	249,26	250,10	251,14	251,77	253,04	254,19
2	19,40	19,41	19,43	19,45	19,46	19,46	19,47	19,48	19,49	19,49
3	8,76	8,74	8,70	8,66	8,63	8,62	8,59	8,58	8,55	8,53
4	5,94	5,91	5,86	5,80	5,77	5,75	5,72	5,70	5,66	5,63
5	4,70	4,68	4,62	4,56	4,52	4,50	4,46	4,44	4,41	4,37
6	4,03	4,00	3,94	3,87	3,83	3,81	3,77	3,75	3,71	3,67
7	3,60	3,57	3,51	3,44	3,40	3,38	3,34	3,32	3,27	3,23
8	3,31	3,28	3,22	3,15	3,11	3,08	3,04	3,02	2,97	2,93
9	3,10	3,07	3,01	2,94	2,89	2,86	2,83	2,80	2,76	2,71
10	2,94	2,91	2,85	2,77	2,73	2,70	2,66	2,64	2,59	2,54
11	2,82	2,79	2,72	2,65	2,60	2,57	2,53	2,51	2,46	2,41
12	2,72	2,69	2,62	2,54	2,50	2,47	2,43	2,40	2,35	2,30
13	2,63	2,60	2,53	2,46	2,41	2,38	2,34	2,31	2,26	2,21
14	2,57	2,53	2,46	2,39	2,34	2,31	2,27	2,24	2,19	2,14
15	2,51	2,48	2,40	2,33	2,28	2,25	2,20	2,18	2,12	2,07
16	2,46	2,42	2,35	2,28	2,23	2,19	2,15	2,12	2,07	2,02
17	2,41	2,38	2,31	2,23	2,18	2,15	2,10	2,08	2,02	1,97
18	2,37	2,34	2,27	2,19	2,14	2,11	2,06	2,04	1,98	1,92
19	2,34	2,31	2,23	2,16	2,11	2,07	2,03	2,00	1,94	1,88
20	2,31	2,28	2,20	2,12	2,07	2,04	1,99	1,97	1,91	1,85
21	2,28	2,25	2,18	2,10	2,05	2,01	1,96	1,94	1,88	1,82
22	2,26	2,23	2,15	2,07	2,02	1,98	1,94	1,91	1,85	1,79
23	2,24	2,20	2,13	2,05	2,00	1,96	1,91	1,88	1,82	1,76
24	2,22	2,18	2,11	2,03	1,97	1,94	1,89	1,86	1,80	1,74
25	2,20	2,16	2,09	2,01	1,96	1,92	1,87	1,84	1,78	1,72
26	2,18	2,15	2,07	1,99	1,94	1,90	1,85	1,82	1,76	1,70
27	2,17	2,13	2,06	1,97	1,92	1,88	1,84	1,81	1,74	1,68
28	2,15	2,12	2,04	1,96	1,91	1,87	1,82	1,79	1,73	1,66
29	2,14	2,10	2,03	1,94	1,89	1,85	1,81	1,77	1,71	1,65
30	2,13	2,09	2,01	1,93	1,88	1,84	1,79	1,76	1,70	1,63
35	2,07	2,04	1,96	1,88	1,82	1,79	1,74	1,70	1,63	1,57
40	2,04	2,00	1,92	1,84	1,78	1,74	1,69	1,66	1,59	1,52
45	2,01	1,97	1,89	1,81	1,75	1,71	1,66	1,63	1,55	1,48
50	1,99	1,95	1,87	1,78	1,73	1,69	1,63	1,60	1,52	1,45
60	1,95	1,92	1,84	1,75	1,69	1,65	1,59	1,56	1,48	1,40
70	1,93	1,89	1,81	1,72	1,66	1,62	1,57	1,53	1,45	1,36
80	1,91	1,88	1,79	1,70	1,64	1,60	1,54	1,51	1,43	1,34
90	1,90	1,86	1,78	1,69	1,63	1,59	1,53	1,49	1,41	1,31
100	1,89	1,85	1,77	1,68	1,62	1,57	1,52	1,48	1,39	1,30

200	1,84	1,80	1,72	1,62	1,56	1,52	1,46	1,41	1,32	1,21
300	1,82	1,78	1,70	1,61	1,54	1,50	1,43	1,39	1,30	1,17
400	1,81	1,78	1,69	1,60	1,53	1,49	1,42	1,38	1,28	1,15
500	1,81	1,77	1,69	1,59	1,53	1,48	1,42	1,38	1,28	1,14
600	1,80	1,77	1,68	1,59	1,52	1,48	1,41	1,37	1,27	1,13
700	1,80	1,77	1,68	1,59	1,52	1,48	1,41	1,37	1,27	1,12
800	1,80	1,76	1,68	1,58	1,52	1,47	1,41	1,37	1,26	1,12
900	1,80	1,76	1,68	1,58	1,52	1,47	1,41	1,36	1,26	1,11
1000	1,80	1,76	1,68	1,58	1,52	1,47	1,41	1,36	1,26	1,11

1-α = 0,99

v_1 = grados de libertad del numerador

v_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	647,79	799,50	864,16	899,58	921,85	937,11	948,22	956,66	963,28	968,63
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	2,57
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	2,53
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51

35	5,48	4,11	3,52	3,18	2,96	2,80	2,68	2,58	2,50	2,44
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39
45	5,38	4,01	3,42	3,09	2,86	2,70	2,58	2,49	2,41	2,35
50	5,34	3,97	3,39	3,05	2,83	2,67	2,55	2,46	2,38	2,32
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27
70	5,25	3,89	3,31	2,97	2,75	2,59	2,47	2,38	2,30	2,24
80	5,22	3,86	3,28	2,95	2,73	2,57	2,45	2,35	2,28	2,21
90	5,20	3,84	3,26	2,93	2,71	2,55	2,43	2,34	2,26	2,19
100	5,18	3,83	3,25	2,92	2,70	2,54	2,42	2,32	2,24	2,18
200	5,10	3,76	3,18	2,85	2,63	2,47	2,35	2,26	2,18	2,11
300	5,07	3,73	3,16	2,83	2,61	2,45	2,33	2,23	2,16	2,09
400	5,06	3,72	3,15	2,82	2,60	2,44	2,32	2,22	2,15	2,08
500	5,05	3,72	3,14	2,81	2,59	2,43	2,31	2,22	2,14	2,07
600	5,05	3,71	3,14	2,81	2,59	2,43	2,31	2,21	2,13	2,07
700	5,05	3,71	3,13	2,80	2,58	2,43	2,31	2,21	2,13	2,07
800	5,04	3,71	3,13	2,80	2,58	2,42	2,30	2,21	2,13	2,06
900	5,04	3,70	3,13	2,80	2,58	2,42	2,30	2,21	2,13	2,06
1000	5,04	3,70	3,13	2,80	2,58	2,42	2,30	2,20	2,13	2,06

1-α = 0,99

v_1 = grados de libertad del numerador

v_1	**11**	**12**	**15**	**20**	**25**	**30**	**40**	**50**	**100**	**1000**
1	973,03	976,71	984,87	993,10	998,08	1001,4	1005,6	1008,1	1013,1	1017,7
2	39,41	39,41	39,43	39,45	39,46	39,46	39,47	39,48	39,49	39,50
3	14,37	14,34	14,25	14,17	14,12	14,08	14,04	14,01	13,96	13,91
4	8,79	8,75	8,66	8,56	8,50	8,46	8,41	8,38	8,32	8,26
5	6,57	6,52	6,43	6,33	6,27	6,23	6,18	6,14	6,08	6,02
6	5,41	5,37	5,27	5,17	5,11	5,07	5,01	4,98	4,92	4,86
7	4,71	4,67	4,57	4,47	4,40	4,36	4,31	4,28	4,21	4,15
8	4,24	4,20	4,10	4,00	3,94	3,89	3,84	3,81	3,74	3,68
9	3,91	3,87	3,77	3,67	3,60	3,56	3,51	3,47	3,40	3,34
10	3,66	3,62	3,52	3,42	3,35	3,31	3,26	3,22	3,15	3,09
11	3,47	3,43	3,33	3,23	3,16	3,12	3,06	3,03	2,96	2,89
12	3,32	3,28	3,18	3,07	3,01	2,96	2,91	2,87	2,80	2,73
13	3,20	3,15	3,05	2,95	2,88	2,84	2,78	2,74	2,67	2,60
14	3,09	3,05	2,95	2,84	2,78	2,73	2,67	2,64	2,56	2,50
15	3,01	2,96	2,86	2,76	2,69	2,64	2,59	2,55	2,47	2,40
16	2,93	2,89	2,79	2,68	2,61	2,57	2,51	2,47	2,40	2,32
17	2,87	2,82	2,72	2,62	2,55	2,50	2,44	2,41	2,33	2,26
18	2,81	2,77	2,67	2,56	2,49	2,44	2,38	2,35	2,27	2,20
19	2,76	2,72	2,62	2,51	2,44	2,39	2,33	2,30	2,22	2,14
20	2,72	2,68	2,57	2,46	2,40	2,35	2,29	2,25	2,17	2,09
21	2,68	2,64	2,53	2,42	2,36	2,31	2,25	2,21	2,13	2,05

22	2,65	2,60	2,50	2,39	2,32	2,27	2,21	2,17	2,09	2,01
23	2,62	2,57	2,47	2,36	2,29	2,24	2,18	2,14	2,06	1,98
24	2,59	2,54	2,44	2,33	2,26	2,21	2,15	2,11	2,02	1,94
25	2,56	2,51	2,41	2,30	2,23	2,18	2,12	2,08	2,00	1,91
26	2,54	2,49	2,39	2,28	2,21	2,16	2,09	2,05	1,97	1,89
27	2,51	2,47	2,36	2,25	2,18	2,13	2,07	2,03	1,94	1,86
28	2,49	2,45	2,34	2,23	2,16	2,11	2,05	2,01	1,92	1,84
29	2,48	2,43	2,32	2,21	2,14	2,09	2,03	1,99	1,90	1,82
30	2,46	2,41	2,31	2,20	2,12	2,07	2,01	1,97	1,88	1,80
35	2,39	2,34	2,23	2,12	2,05	2,00	1,93	1,89	1,80	1,71
40	2,33	2,29	2,18	2,07	1,99	1,94	1,88	1,83	1,74	1,65
45	2,29	2,25	2,14	2,03	1,95	1,90	1,83	1,79	1,69	1,60
50	2,26	2,22	2,11	1,99	1,92	1,87	1,80	1,75	1,66	1,56
60	2,22	2,17	2,06	1,94	1,87	1,82	1,74	1,70	1,60	1,49
70	2,18	2,14	2,03	1,91	1,83	1,78	1,71	1,66	1,56	1,45
80	2,16	2,11	2,00	1,88	1,81	1,75	1,68	1,63	1,53	1,41
90	2,14	2,09	1,98	1,86	1,79	1,73	1,66	1,61	1,50	1,39
100	2,12	2,08	1,97	1,85	1,77	1,71	1,64	1,59	1,48	1,36
200	2,06	2,01	1,90	1,78	1,70	1,64	1,56	1,51	1,39	1,25
300	2,04	1,99	1,88	1,75	1,67	1,62	1,54	1,48	1,36	1,21
400	2,03	1,98	1,87	1,74	1,66	1,60	1,52	1,47	1,35	1,18
500	2,02	1,97	1,86	1,74	1,65	1,60	1,52	1,46	1,34	1,17
600	2,01	1,97	1,85	1,73	1,65	1,59	1,51	1,46	1,33	1,16
700	2,01	1,96	1,85	1,73	1,65	1,59	1,51	1,45	1,32	1,15
800	2,01	1,96	1,85	1,73	1,64	1,58	1,50	1,45	1,32	1,14
900	2,01	1,96	1,85	1,72	1,64	1,58	1,50	1,45	1,32	1,14
1000	2,01	1,96	1,85	1,72	1,64	1,58	1,50	1,45	1,32	1,13

1-α = 0,99

v_1 = grados de libertad del numerador

v_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	4052,18	4999,50	5403,35	624,58	5763,65	858,99	5928,36	5981,07	6022,47	6055,85
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30

13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98
35	7,42	5,27	4,40	3,91	3,59	3,37	3,20	3,07	2,96	2,88
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80
45	7,23	5,11	4,25	3,77	3,45	3,23	3,07	2,94	2,83	2,74
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41
300	6,72	4,68	3,85	3,38	3,08	2,86	2,70	2,57	2,47	2,38
400	6,70	4,66	3,83	3,37	3,06	2,85	2,68	2,56	2,45	2,37
500	6,69	4,65	3,82	3,36	3,05	2,84	2,68	2,55	2,44	2,36
600	6,68	4,64	3,81	3,35	3,05	2,83	2,67	2,54	2,44	2,35
700	6,67	4,64	3,81	3,35	3,04	2,83	2,66	2,54	2,43	2,35
800	6,67	4,63	3,81	3,34	3,04	2,82	2,66	2,53	2,43	2,34
900	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34

1-α = 0,99

v_1 = grados de libertad del numerador

v_1	**11**	**12**	**15**	**20**	**25**	**30**	**40**	**50**	**100**	**1000**
1	6083,32	106,32	6157,28	6208,73	6239,83	6260,65	6286,78	6302,52	6334,11	6362,68
2	99,41	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49	99,50
3	27,13	27,05	26,87	26,69	26,58	26,50	26,41	26,35	26,24	26,14

4	14,45	14,37	14,20	14,02	13,91	13,84	13,75	13,69	13,58	13,47
5	9,96	9,89	9,72	9,55	9,45	9,38	9,29	9,24	9,13	9,03
6	7,79	7,72	7,56	7,40	7,30	7,23	7,14	7,09	6,99	6,89
7	6,54	6,47	6,31	6,16	6,06	5,99	5,91	5,86	5,75	5,66
8	5,73	5,67	5,52	5,36	5,26	5,20	5,12	5,07	4,96	4,87
9	5,18	5,11	4,96	4,81	4,71	4,65	4,57	4,52	4,41	4,32
10	4,77	4,71	4,56	4,41	4,31	4,25	4,17	4,12	4,01	3,92
11	4,46	4,40	4,25	4,10	4,01	3,94	3,86	3,81	3,71	3,61
12	4,22	4,16	4,01	3,86	3,76	3,70	3,62	3,57	3,47	3,37
13	4,02	3,96	3,82	3,66	3,57	3,51	3,43	3,38	3,27	3,18
14	3,86	3,80	3,66	3,51	3,41	3,35	3,27	3,22	3,11	3,02
15	3,73	3,67	3,52	3,37	3,28	3,21	3,13	3,08	2,98	2,88
16	3,62	3,55	3,41	3,26	3,16	3,10	3,02	2,97	2,86	2,76
17	3,52	3,46	3,31	3,16	3,07	3,00	2,92	2,87	2,76	2,66
18	3,43	3,37	3,23	3,08	2,98	2,92	2,84	2,78	2,68	2,58
19	3,36	3,30	3,15	3,00	2,91	2,84	2,76	2,71	2,60	2,50
20	3,29	3,23	3,09	2,94	2,84	2,78	2,69	2,64	2,54	2,43
21	3,24	3,17	3,03	2,88	2,79	2,72	2,64	2,58	2,48	2,37
22	3,18	3,12	2,98	2,83	2,73	2,67	2,58	2,53	2,42	2,32
23	3,14	3,07	2,93	2,78	2,69	2,62	2,54	2,48	2,37	2,27
24	3,09	3,03	2,89	2,74	2,64	2,58	2,49	2,44	2,33	2,22
25	3,06	2,99	2,85	2,70	2,60	2,54	2,45	2,40	2,29	2,18
26	3,02	2,96	2,81	2,66	2,57	2,50	2,42	2,36	2,25	2,14
27	2,99	2,93	2,78	2,63	2,54	2,47	2,38	2,33	2,22	2,11
28	2,96	2,90	2,75	2,60	2,51	2,44	2,35	2,30	2,19	2,08
29	2,93	2,87	2,73	2,57	2,48	2,41	2,33	2,27	2,16	2,05
30	2,91	2,84	2,70	2,55	2,45	2,39	2,30	2,25	2,13	2,02
35	2,80	2,74	2,60	2,44	2,35	2,28	2,19	2,14	2,02	1,90
40	2,73	2,66	2,52	2,37	2,27	2,20	2,11	2,06	1,94	1,82
45	2,67	2,61	2,46	2,31	2,21	2,14	2,05	2,00	1,88	1,75
50	2,63	2,56	2,42	2,27	2,17	2,10	2,01	1,95	1,82	1,70
60	2,56	2,50	2,35	2,20	2,10	2,03	1,94	1,88	1,75	1,62
70	2,51	2,45	2,31	2,15	2,05	1,98	1,89	1,83	1,70	1,56
80	2,48	2,42	2,27	2,12	2,01	1,94	1,85	1,79	1,65	1,51
90	2,45	2,39	2,24	2,09	1,99	1,92	1,82	1,76	1,62	1,48
100	2,43	2,37	2,22	2,07	1,97	1,89	1,80	1,74	1,60	1,45
200	2,34	2,27	2,13	1,97	1,87	1,79	1,69	1,63	1,48	1,30
300	2,31	2,24	2,10	1,94	1,84	1,76	1,66	1,59	1,44	1,25
400	2,29	2,23	2,08	1,92	1,82	1,75	1,64	1,58	1,42	1,22
500	2,28	2,22	2,07	1,92	1,81	1,74	1,63	1,57	1,41	1,20
600	2,28	2,21	2,07	1,91	1,80	1,73	1,63	1,56	1,40	1,19
700	2,27	2,21	2,06	1,90	1,80	1,72	1,62	1,55	1,39	1,18
800	2,27	2,21	2,06	1,90	1,80	1,72	1,62	1,55	1,39	1,17
900	2,27	2,20	2,06	1,90	1,79	1,72	1,61	1,55	1,39	1,16
1000	2,27	2,20	2,06	1,90	1,79	1,72	1,61	1,54	1,38	1,16

Tabla 9. Valores críticos para la prueba de rangos con signo.

n	Unilateral $\alpha = 0.01$ Bilateral $\alpha = 0.02$	Unilateral $\alpha = 0.025$ Bilateral $\alpha = 0.05$	Unilateral $\alpha = 0.05$ Bilateral $\alpha = 0.1$
5			1
6		1	2
7	0	2	4
8	2	4	6
9	3	6	8
10	5	8	11
11	7	11	14
12	10	14	17
13	13	17	21
14	16	21	26
15	20	25	30
16	24	30	36
17	28	35	41
18	33	40	47
19	38	46	54
20	43	52	60
21	49	59	68
22	56	66	75
23	62	73	83
24	69	81	92
25	77	90	101
26	85	98	110
27	93	107	120
28	102	117	130
29	111	127	141
30	120	137	152

Tomado de Walpole et al (2012). Probabilidad y Estadística para Ingeniería y Ciencias.

Acerca de

los autores

Carlos Solarte Portilla

Dr. Sc., Universidad de la Habana, Cuba. MSc en Producción Animal, Universidad Nacional de Colombia. Zootecnista, Universidad de Nariño. Profesor Tiempo Completo adscrito al Departamento de Producción y Procesamiento Animal de la Universidad de Nariño. Ex-Rector de la Universidad de Nariño.

Correo electrónico: solartecar@gmail.com

Hernán Abdón García

Magíster en Estadística. Universidad Nacional de Colombia, Sede Bogotá; Especialista en Gestión de Proyectos, CINDE - Universidad de Nariño; Licenciado en Matemáticas y Física, Universidad de Nariño. Profesor de Tiempo Completo adscrito al Departamento de Matemáticas y Estadística de la Universidad de Nariño, Categoría Asociado en el escalafón docente. Director del Centro de Estudios y Asesorías en Estadística-CEASE, Universidad de Nariño. Coordinador de la Maestría en Estadística Aplicada, Universidad de Nariño.

Correo electrónico: habgarcia@gmail.com; hgarcia@udenar.edu.co

Marco Antonio Imués Figueroa

Máster en Acuicultura, Universidad de Barcelona, España. Especialista en Estadística, Universidad Nacional de Colombia. Especialista en Computación Aplicada, Universidad Mariana, Pasto. Zootecnista, Universidad de Nariño. Profesor de Tiempo Completo adscrito al Departamento de Recursos Hidrobiológicos de la Universidad de Nariño. Director de Proyecto de Investigación aprobado y financiado por el Sistema General de Regalías de Colombia. Evaluador de Proyectos del Ministerio de Ciencia y Tecnología.

Correo electrónico: marcoi@udenar.edu.co; maimuez@yahoo.com

ëditorial
Universidad de Nariño

Año de publicación: 2026
San Juan de Pasto, Nariño, Colombia

Durante décadas se consideró que las teorías y técnicas estadísticas tenían un papel secundario en la formación de biólogos y profesionales de las ciencias basadas en la Biología. Hoy esta visión ha cambiado por completo, pues el progreso de las Ciencias Biológicas y de las áreas aplicadas al sector pecuario depende estrechamente del desarrollo estadístico. Gracias a estos métodos ha sido posible comprender mejor el comportamiento animal y resolver numerosos problemas prácticos. Asimismo, avances como nuevos fármacos, suplementos, la valoración genética de poblaciones y la caracterización de su variabilidad, así como el diseño de estrategias para controlar epizootias y pandemias, se sustentan en procedimientos estadísticos apoyados por el desarrollo de software y hardware modernos.

Esta obra constituye una segunda edición, en un compendio riguroso de los principios de la Bioestadística orientados a la Producción y la Salud Animal, destacando su utilidad en la comprensión y análisis de fenómenos biológicos complejos. Desde las bases conceptuales presentadas en el Capítulo 1 y los métodos descriptivos del Capítulo 2, el texto establece una estructura sólida para el tratamiento adecuado de datos experimentales. Su enfoque favorece la formación de criterios analíticos indispensables en el quehacer académico y profesional.

Los Capítulos 3 al 6 desarrollan de manera sistemática los elementos fundamentales de la probabilidad, las distribuciones estadísticas, la estimación de parámetros y el contraste de hipótesis, integrando ejemplos aplicados a sistemas de producción animal. Estos contenidos permiten al lector comprender con mayor profundidad la variabilidad inherente a los organismos vivos y las implicaciones inferenciales de los análisis. La secuencia temática propicia una comprensión coherente de los métodos estadísticos y su pertinencia en la investigación científica.

Esta segunda edición, corregida, mejorada y aumentada, incorpora avances relevantes en el análisis de varianza y los métodos no paramétricos desarrollados en los Capítulos 7 y 8, ampliando el espectro de herramientas disponibles para el estudio de problemas propios del sector pecuario. Estas actualizaciones fortalecen la capacidad del lector para abordar situaciones experimentales más exigentes y ajustadas a las necesidades actuales del campo. En conjunto, el libro se consolida como un recurso académico de referencia para quienes poseen conocimientos intermedios en Bioestadística y requieren aplicarlos con rigor en la investigación en producción y salud animal.

ISBN: 978-628-7864-96-2

9 786287 864962

Universidad de Nariño
FUNDADA EN 1904

Universidad de Nariño
ACREDITADA EN ALTA CALIDAD
RESOLUCIÓN MEN 000022 - ENERO 11 DE 2023

êditorial
Universidad de Nariño