BAQUE W.
IBARRA A.
VERA N.
VALDEZ D.

∑STADÍSTICA

APLICADA A AGRONOMIA, AGROINDUSTRIA Y CIENCIAS AFINES

Ejercicios resueltos en R, Infostat, Minitab, Statgraphic.

1era Edición

UNIVERSIDAD AGRARIA DEL ECUADOR

FACULTAD DE CIENCIAS AGRARIAS

WILMER BAQUE BUSTAMANTE

ALEX IBARRA VELASQUEZ

NESTOR VERA LUCIO

DANILO VALDEZ RIVERA

CAMPUS GUAYAQUIL

2019

LOS AUTORES

Ing. Wilmer Baque Bustamante, Universidad Agraria del Ecuador, profesor titular, Ingeniero en Estadística e Informática (ESPOL), Master en Investigación de Mercados (ESPOL). Profesor con experiencia por más de 15 años, dictando clases en la Universidad Católica de Guayaquil, Instituto Tecnológico Espíritu Santo, y Universidad de Guayaquil en la facultad de administración. Consultor en Investigación de Mercados, funcionario público desde 2010 hasta 2014.

Ing. Danilo Valdez Rivera, Universidad Agraria del Ecuador, profesor titular, Ingeniero Agrónomo (UAE), Master en Economía Agrícola (SIPUAE-UAE). Experiencia en Mango por más de 20 años en la empresa privada y en la docencia hace 5 años.

Ing. Néstor Vera Lucio, Universidad Agraria del Ecuador, profesor titular, Decano de la facultad de Economía Agrícola, Ingeniero en Estadística e Informática (ESPOL), Master en Docencia (SIPUAE-UAE). Consultor en Investigación de Mercados.

Ec. Alex Ibarra Velásquez, Universidad Agraria del Ecuador, profesor titular, Economista Agrícola (UAE), Master en Comercio Exterior y Marketing (ESPOL), Consultor económico y de mercados.

AGRADECIMIENTOS

Agradezco a dios por la oportunidad que me ha dado de tener a mis Padres
Y a hermanos junto a mí.
Por permitirme conocer otras culturas, personas alrededor del mundo,
dándome cuenta que somos todos iguales,
solamente separados por fronteras, religión, idioma, etc.

Al creador, fundador de la Universidad Agraria del Ecuador,
al Dr. Jacobo Bucaram por su apertura para la investigación,
a la Dra. Martha Bucaram Leverone de Jorgge,
rectora de la Universidad.

Un agradecimiento a los compañeros y estudiantes
quienes ayudaron con esta primera edición de este libro,
no se pudiera llegar a este objetivo sin la ayuda de los programas
estadísticos utilizados, como R Studio, Infostat, Statgraphic, Minitab.

DEDICATORIA

Dedico este Libro a Wilmer Baque Bustamante
hijo de Aurelio y Clementina, hermano de
Marcos y Byron, tío de Rafaella y
Padre de crianza de Yurén.

A mis enemigos, y fantoches que de alguna u otra
manera me motivaron con sus comentarios o
acciones a realizar esta obra.

A mis amigos los cuales son muy pocos que
creyeron en mí, los aprecio bastante, cada vez
que me consultaban algo, mi confianza se
Incrementaba.

Y por supuesto a Dios que me ha permitido
vivir esta vida, aceptándola
porque así son los caminos del señor,
me voy silbando bajito a la luz de la Luna.

Gracias totales
Wilmer Baque B.

CONTENIDO

Resumen

Siendo la estadística una ciencia importante en nuestros días, surgió la idea de contribuir a la sociedad con este libro que contiene ejercicios reales que servirán para el entendimiento de fenómenos, eventos de la vida cotidiana.

En el primer capítulo se plantea la definición de estadística, sus orígenes demás, de medidas de tendencia central y dispersión. En el siguiente capítulo, está la distribución Normal y la transformación de datos lo cual es muy importante para realizar pruebas estadísticas.

En la siguiente parte, nos encontramos con correlación y regresión lineal simple, la cual ayuda de alguna manera a encontrar relaciones lineales directas o inversamente proporcionales de x con y. Adicionalmente, las técnicas de muestreo es un tema relevante y soporte para las ciencias ya que con las muestras se observan o realizan las pruebas estadísticas para validar científicamente una hipotesis.

Por último, surge la aplicación de funciones exponenciales y logarítmicas como una respuesta al crecimiento o decrecimiento de virus, insectos etc, donde es importante conocer características de la dinámica poblacional de alguna especie. Finalmente, las pruebas de hipotesis para la media poblacional y para dos poblaciones y estadísticas no paramétrica que nos ayuda a realizar análisis estadísticos de variables que no guardan una distribución normal.

Abstract

Being statistics an important science in our day, the idea of contributing to society with this book that contains real exercises that will serve to understand phenomena, events of daily life, emerged.

The definition of statistics, their remaining origins, measures of central tendency and dispersion are raised in the first chapter. In the next chapter, there is the Normal distribution and the data transformation which is very important for statistical tests.

In the next part, we find simple linear regression and correlation, which somehow helps to find direct or inversely proportional linear relationships of x with y. In addition, sampling techniques is a relevant issue and support for the sciences that with the samples are observed or performed statistical tests to scientifically validate a hypothesis.

Finally, the application of exponential and logarithmic functions arises as a response to the growth or decrease of the virus, insects, etc., where it is important to know the characteristics of the population dynamics of some species. Finally, the hypothesis tests for the population media and for two populations and non-parametric statistics that help us to perform statistical analyzes of variables that do not keep a normal distribution.

Introducción

La realización de este libro es para dar conocer las bondades de la estadística en las ciencias agrarias. El propósito es dar una herramienta dirigida a estudiantes, y/o profesionales para el uso de la estadística en la vida diaria.

Hoy el uso de la estadística se ha extendido más allá de sus orígenes como un servicio al estado o al gobierno. Personas y organizaciones usan estadística para entender datos y tomar decisiones en ciencias naturales y sociales, medicina, negocios y otras áreas. La estadística es pensada generalmente no como una subárea de las matemáticas sino como una ciencia diferente "aliada". Muchas universidades tienen departamentos en matemáticas y estadística separadamente. La estadística es enseñada en departamentos tan diversos como psicología, educación, agronomía, ciencias forestales, ingeniería, economía y salud pública.

Se utilizó casos reales del territorio, es decir cultivos tradicionales y no tradicionales que se ajusten a la realidad del país. Los ejercicios y demás temas han sido aterrizar realidades de los diferentes territorios aplicando la estadística en cuatro softwares estadísticos, Minitab, R, Statgraphic e Infostat donde estos son softwares libres y pueden ser descargados por estudiantes, investigadores, etc.

El uso de este libro básicamente es mayormente practico con ejercicios aplicados y reales para el mejor aprendizaje de los estudiantes que quieren sumergirse en el mundo de los datos y la estadística como instrumento en la toma de decisiones.

CAPITULO 1

1.1 ESTADISTICA, MEDIDAS DE TENDENCIA CENTRAL Y DISPERSION.

No se podría hablar de Estadística sin primero hablar de Probabilidad, el origen de las probabilidades se inicia en el año 1654 cuando el matemático francés Blaise Pascal hacia un viaje con el apasionado jugador de dados y cartas, conocido como El Caballero de Mere, quien era noble e ilustrado, es desde allí su utilización inicial para juegos de azar, consecuentemente después de algunos años se la utiliza como parte de la Estadística. Walker (1929) atribuye el primer uso del término “estadística” al profesor alemán, Gottfried Achenwall (1719 – 1772), quien utilizó la palabra alemana Statistik, que extrajo del término italiano statista (estadista). Creía, y con sobrada razón, que la nueva ciencia sería el aliado más eficaz del gobernante consciente, para la planificación de los recursos. La raíz de la palabra se halla, por otra parte, en el término latino status, que significa estado o situación. Indicando la importancia histórica de la recolección de datos por parte del gobierno de un país, relacionados principalmente a información demográfica que mide la estructura y dinámica de la población tales como fecundidad, natalidad, mortalidad, emigración, inmigración. (López & González, 2015)

El Dr. E. A. W. Zimmerman introdujo el término statistics (estadística) a Inglaterra. Su uso fue popularizado por Sir John Sinclair (1754 – 1835) en su obra Statistical Account of Scotland 1791 – 1799 (“Informe estadístico sobre Escocia 1791 – 1799”). Sin embargo, mucho antes del siglo XVII, la gente ya la utilizaba y registraba datos. He aquí algunas definiciones de Estadística:

a) Conjunto de métodos para planear estudios y experimentos, obtener datos y luego organizar, resumir, presentar, analizar, interpretar y llegar a conclusiones basadas en los datos. (Triola, 2009)

b) Ciencia derivada de la matemática que se ocupa de la recopilación de información contenida en datos provenientes de muestras y de su uso para hacer inferencias acerca de la población de donde fueron extraídos los mismos.

c) La Estadística estudia los métodos científicos para recolectar, organizar, resumir y analizar datos, así como para extraer conclusiones válidas y tomar decisiones razonables basadas con tal análisis. (Spiegel, Murray R.; Stephens, 2009)

1.2 Historia de la Estadística

Los comienzos de la estadística pueden ser hallados en el antiguo Egipto, cuyos faraones lograron recopilar, hacia el año 3050 antes de Cristo, datos relativos a la población y la riqueza del país. De acuerdo al historiador griego Heródoto, este registro de riqueza y de población se hizo con el objetivo de preparar la construcción de las pirámides. En el mismo Egipto, Ramsés II hizo un censo de las tierras con el objeto de verificar un nuevo reparto.

En el antiguo Israel, la Biblia da referencias en el libro de los Números, de los datos estadísticos obtenidos en dos recuentos de la población hebrea. El rey David por otra parte, ordenó a Joab, general del ejército hacer un censo de Israel con la finalidad de conocer el número de la población. Los chinos efectuaron censos hace más de cuarenta siglos y los griegos efectuaron censos periódicamente con fines tributarios, sociales (división de tierras) y militares (cálculo de recursos y hombres disponibles).

Pero fueron los romanos, maestros de la organización política, quienes mejor supieron emplear los recursos de la Estadística. Cada cinco años realizaban un censo de la población y sus funcionarios públicos tenían la obligación de anotar nacimientos, defunciones y matrimonios, sin olvidar los recuentos periódicos del ganado y de las riquezas contenidas en las tierras conquistadas. Para el nacimiento de Cristo sucedía uno de estos empadronamientos de la población bajo la autoridad del imperio.

Durante los mil años siguientes a la caída del imperio Romano se realizaron muy pocas operaciones estadísticas, con la notable excepción de las relaciones de tierras pertenecientes a la Iglesia, compiladas por Pipino el Breve en el 758 y por Carlomagno en el 762 DC. Durante el siglo IX se realizaron en Francia algunos censos parciales de siervos. En Inglaterra, Guillermo el Conquistador recopiló el Domesday Book o libro del Gran Catastro para el año 1086, un documento de la propiedad, extensión y valor de las tierras de Inglaterra. Esa obra fue el primer compendio estadístico de Inglaterra.

Aunque Carlomagno en Francia y Guillermo el Conquistador en Inglaterra, trataron de revivir la técnica romana, los métodos estadísticos permanecieron casi olvidados durante la Edad Media. Durante los siglos XV, XVI, y XVII, Leonardo de Vinci, Nicolás Copérnico, Galileo, Neper, William Harvey, Sir Francis Bacon y René Descartes, hicieron grandes contribuciones al método científico, de tal forma que cuando se crearon los Estados Nacionales y surgió como fuerza el comercio internacional, existía ya un método capaz de aplicarse a los datos económicos.

Para el año 1532 empezaron a registrarse en Inglaterra las defunciones debido al temor que Enrique VII tenía por la peste. Más o menos por la misma época, en Francia la ley exigió a los clérigos registrar los bautismos, fallecimientos y matrimonios. Durante un brote de peste que apareció a fines de la década de 1500, el gobierno inglés comenzó a publicar estadísticas semanales de los decesos. Esa costumbre continuó muchos años, y en 1632 estos Bills of Mortality (Cuentas de Mortalidad) contenían los nacimientos y fallecimientos por sexo. En 1662, el capitán John Graunt usó documentos que abarcaban treinta años y efectuó predicciones sobre el número de personas que morirían de varias enfermedades y sobre las proporciones de nacimientos de varones y mujeres que cabría esperar.

El primer empleo de los datos estadísticos para fines ajenos a la política tuvo lugar en 1691 y estuvo a cargo de Gaspar Neumann, un profesor alemán que vivía en Breslau. Este investigador se propuso destruir la antigua creencia popular de que en los años terminados en siete moría más gente que en los restantes, y para lograrlo hurgó pacientemente en los archivos parroquiales de la ciudad. Después de revisar miles de partidas de defunción pudo demostrar que en tales años no fallecían más personas que en los demás. Los procedimientos de Neumann fueron conocidos por el astrónomo inglés Halley, descubridor del cometa que lleva su nombre, quien los aplicó al estudio de la vida humana. Sus cálculos sirvieron de base para las tablas de mortalidad que hoy utilizan todas las compañías de seguros.

Durante el siglo XVII y principios del XVIII, matemáticos como Bernoulli, Francis Maseres, Lagrange y Laplace desarrollaron la teoría de probabilidades. No obstante, durante cierto tiempo, la teoría de las probabilidades limitó su aplicación a los juegos de azar y hasta el siglo XVIII no comenzó a aplicarse a los grandes problemas científicos. Thomas Bayes (Londres, Inglaterra, 1702 - Tunbridge Wells, 1761), fue uno de los primeros en utilizar la probabilidad inductivamente y establecer una base matemática para la inferencia probabilística. Actualmente, con base en su obra, se ha desarrollado una poderosa teoría que ha conseguido notables aplicaciones en las más diversas áreas del conocimiento.

Godofredo Achenwall, profesor de la Universidad de Gotinga, acuñó en 1760 la palabra estadística. Jacques Quételect es quien aplica la Estadística a las ciencias sociales. Él interpretó la teoría de la probabilidad para su uso en las ciencias sociales y resolver la aplicación del principio de promedios y de la variabilidad a los fenómenos sociales. Entretanto, en el período del 1800 al 1820 se desarrollaron dos conceptos matemáticos fundamentales para la teoría estadística; la teoría de los errores de observación, aportada por Laplace y Gauss; y la teoría de los mínimos cuadrados desarrollada por Laplace, Gauss y Legendre. A finales del siglo XIX, Sir Francis Galton dio forma al método conocido como regresión. De aquí partió el desarrollo del coeficiente de correlación creado por Karl Pearson y otros cultivadores de la ciencia biométrica como J. Pease Norton, R. H. Hooker y G. Udny Yule, que efectuaron amplios estudios sobre la medida de las relaciones. Más adelante, a partir de 1919 la estadística experimental tuvo su desarrollo cuando Ronald A. Fisher asumió la dirección del departamento de Estadística de la Estación Experimental de Rothampstead en Londres, Inglaterra. La información sobre la historia de la Estadística es cortesía de (López & González, 2015).

En Ecuador, el Instituto de Censos y Estadísticas (INEC), es el ente encargado de elaborar las estadísticas y los censos poblacionales y de vivienda. “El primer Censo de Población en nuestro país se realizó en noviembre de 1950; el último censo, el séptimo, se realizó en noviembre del año pasado. En estos 60 años, la población del Ecuador pasó de 3’202.757 a 14’306.876 habitantes, lo que significa un crecimiento promedio anual de 2,5% anual” (Albornoz, 2011).

Con los pocos datos disponibles a la fecha del Censo de 2010, son dos las tendencias que pueden resaltarse: la caída de la tasa de crecimiento y la concentración de la población en Pichincha y Guayas.

1.3 División de la Estadística

La Estadística para su mejor estudio se ha dividido en tres grandes ramas: Estadística Descriptiva, Probabilidades y la Estadística Inferencial.

1.3.1 La Estadística Descriptiva

Consiste en la presentación de datos en forma numérica, tablas y gráficas. Esta comprende cualquier actividad relacionada con los datos y está diseñada para resumir o describir los mismos, sin factores pertinentes adicionales; esto es, sin intentar inferir nada que vaya más allá de los datos, como tales. Es en general utilizada en la etapa inicial de los análisis, cuando se tiene contacto con los datos por primera vez. La Probabilidad puede ser pensada como la teoría matemática utilizada para estudiar la incertidumbre oriunda de fenómenos de carácter aleatorio, o sea, producto del azar.

1.3.2 La Estadística Inferencial.

Proviene de muestras, donde su análisis requiere de generalizaciones que van más allá de los datos. Como consecuencia, la característica más importante del reciente crecimiento de la estadística ha sido un cambio en el énfasis de los métodos estadísticos que son diseñados para contribuir al proceso de juicios científicos frente a la incertidumbre y variación. (Walpole Ronald, Myers Raymond, Myers Sharon, 2012)

La Estadística descriptiva y la inferencial comprenden la estadística aplicada. Hay también una disciplina llamada estadística matemática, la cual se refiere a las bases teóricas de la materia, e incluye el estudio de las probabilidades.

De acuerdo a la figura descrita arriba, están las medidas de tendencia central, como la media, la mediana y la moda, siendo parte importante en el análisis descriptivo de un conjunto de datos.

Adicionalmente, se encuentran las medidas de dispersión, que son las que miden la distancia con respecto al centro de la distribución de los datos en este caso con respecto a la media, mostrando así la varianza, desviación estándar y el rango.

Durante el desarrollo de este capítulo abordaremos cada una de estas medidas con ejemplos de datos obtenidos por instituciones como el MAG en el territorio ecuatoriano para dar un valor agregado a la información y los tipos de cultivos en el país.

Otra división de la estadística es:

1.3.4 Estadística Paramétrica

En la estadística paramétrica nuestro interés es hacer estimaciones y pruebas acerca de uno o más parámetros de la población. Además, en todas estas estimaciones y pruebas de hipótesis se establece como suposición general que la población o poblaciones de donde provienen las muestras deben estar distribuidas normalmente, aunque sea en forma aproximada.

1.3.5 Estadística No Paramétrica

Estudia las pruebas y modelos estadísticos cuya distribución subyacente no se ajusta a los llamados criterios paramétricos. Su distribución no puede ser definida a priori, pues son los datos observados los que la determinan. La utilización de estos métodos se hace recomendable cuando no se puede asumir que los datos se ajusten a una distribución normal o cuando el nivel de medida empleado no sea, como mínimo, de intervalo.

Otras ramas importantes de la Estadística son:

Geoestadística: comprende a un conjunto de herramientas y técnicas que sirven para analizar y predecir los valores de una variable que se muestra distribuida en el espacio o en el tiempo de una forma continua. Debido a su aplicación orientada a los Sistemas de Información Geográfica (SIG), también se podría definir como la estadística relacionada con los datos geográficos.

Inferencia Bayesiana: la metodología bayesiana está basada en la interpretación subjetiva de la probabilidad y tiene como punto central el Teorema de Bayes. Los modelos bayesianos primordialmente incorporan conocimiento previo para poder estimar modelos útiles dentro de un espacio muestral y de este modo poder estimar parámetros que provengan de la experiencia o de una teoría probabilística.

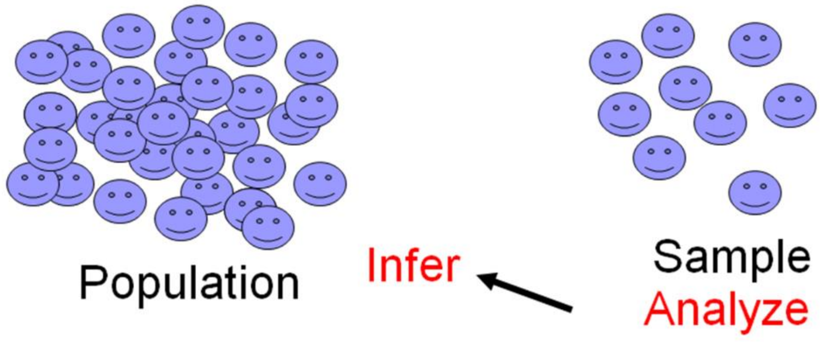
Estadística Multivariada: las técnicas estadísticas multivariadas permiten establecer, a partir de numerosos datos y variables ciertas relaciones, investigar estructuras latentes y ensayar diversas maneras de organizar dichos datos, bien transformándolos y presentándolos bajo una forma nueva más asequible, bien reduciéndolos, sin perder demasiada información inicial.

1.4 Población.

Es el conjunto de todos los elementos de interés en un estudio determinad, para distinguir una población de una muestra se denotará como **N**. Según las características de la población objetivo obtendremos subpoblaciones, por ejemplo, la variable peso, la variable altura, la variable sexo, la variable estado civil, etc.

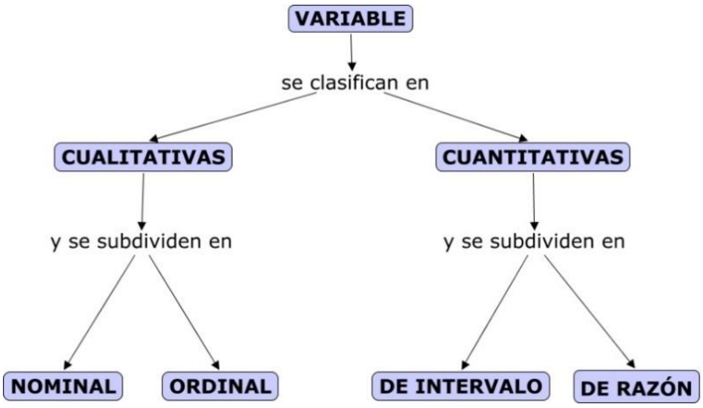
1.5 Muestra.

A través de una población obtendremos elementos que serán parte de una muestra aleatoria para inferir o describir características de la población en estudio. Se describe a una muestra como **n**.



1.6 Tipos de Variables.

Una variable es una característica, propiedad o atributo, con respecto a la cual los elementos de una población difieren de alguna forma.



Variable cualitativa es aquella que mide una cualidad.

Variable nominal es aquella cuyos valores son nombres o códigos sin una relación de orden intrínseco entre ellos. Ejemplos son: vigor de la planta; nivel de fitotoxicidad; o color del fruto.

Variable ordinal corresponde a aquella cuyos valores son nombres o códigos, pero con una relación de orden intrínseco entre ellos, es decir, sus valores conllevan un ordenamiento de mejor a peor o de mayor a menor. Por ejemplo: la calificación (excelente, bueno, regular, malo); la calidad del fruto (extra, primera, segunda, ...) o nivel de infestación (sana, leve, moderada).

Variable cuantitativa es aquella que mide una cantidad.

Variable discreta usualmente es aquella que solo toma valores enteros, finitos o numerables. Por ejemplo: número de hijos por familia; número de elementos defectuosos en una partida de repuestos o número de insectos por hoja, Numero de Mazorcas en la planta

Variable continua son las de mayor jerarquía matemática, y corresponden a aquellas que pueden asumir cualquier valor dentro de un cierto real rango. Por ejemplo: altura de planta; peso; rendimiento de un cultivo o el tiempo que demora un corredor en los 100 m.

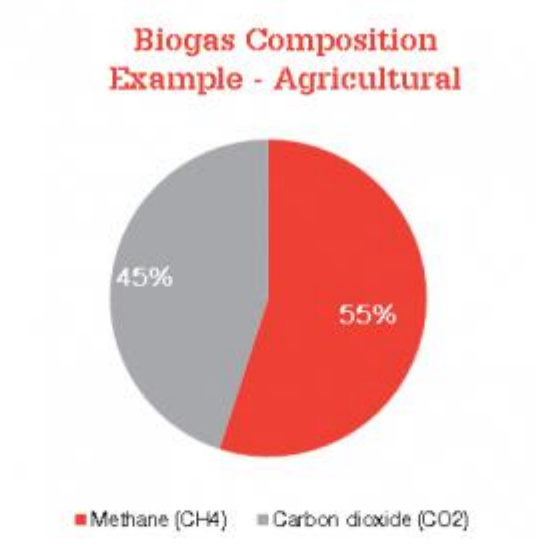
1.7 Tipos de Gráficos

1.7.1 Series de Tiempo

Una gráfica de series de tiempo del NASDAQ 100 INDEX que es constituido por 100 de las más grandes compañías listado en el NASDAQ stock Exchange. En el siguiente grafico podemos observar el índice durante los últimos 6 meses la cual es una buena referencia para inversionistas.

1.7.2 Grafico de Pastel o Pie

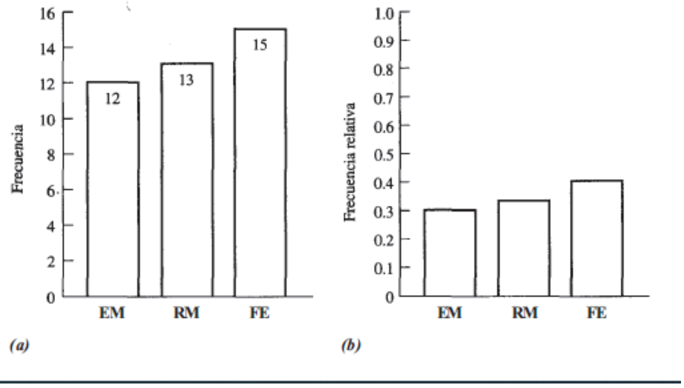
Los garficos de Pastel o circulares tambien se utilizan para visualizar datos cualitativos. Para construir una grafica circular, se divide el circulo en las proporciones adecuadas las cuales muestras diferentes categorias que en su mayor parte suman hasta un 100%.



Fuente: www.clarke-energy.com

1.7.3 Grafico de Barras

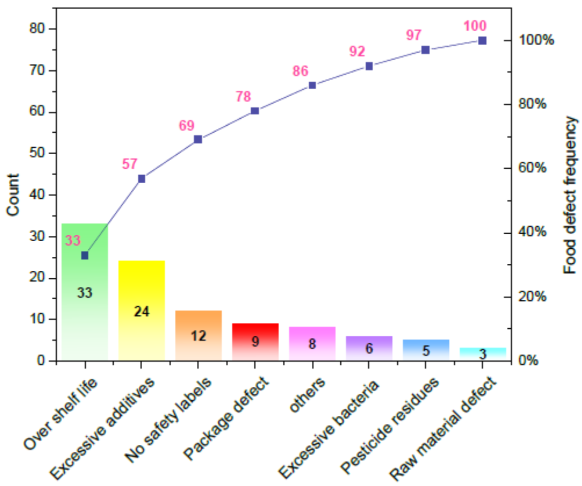
Se puede utilizar un gráfico de barra verticales. Cada categoría está tomada por una barra vertical, todas de la misma anchura. Las alturas de las barras dependen del número de observaciones por categoría. El eje vertical del grafico puede representar frecuencias, frecuencias relativas o porcentajes, como se puede observar a continuación:



Fuente: Tomado de (Milton, 2001)

1.7.4 Grafico de Pareto

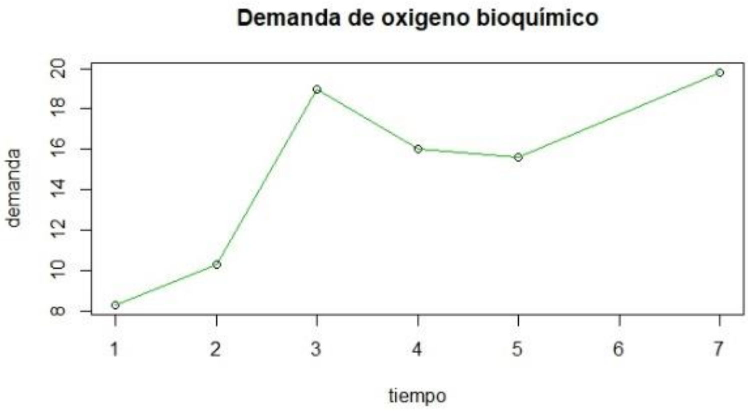
El grafico de Pareto es un gráfico de barras para datos cualitativos, donde se ordenan de acuerdo con las frecuencias de mayor a menor. Las escalas verticales de la gráfica de Pareto representan tanto frecuencias como frecuencias relativas. Para este tipo de grafico existe El principio de Pareto el cual es 80/20 que significa que aproximadamente el 80% de los efectos son por el 20% de causas.



Fuente: www.originlab.com

1.7.5 Gráficas de puntos

Uno de los más sencillos resúmenes gráficos de datos son las gráficas de puntos. En el ejemplo horizontal se presenta el intervalo de los datos. Cada dato se representa por un punto colocado sobre este eje. La gráfica de puntos muestra los detalles de los datos y son útiles para comparar la distribución de los datos de dos o más variables.



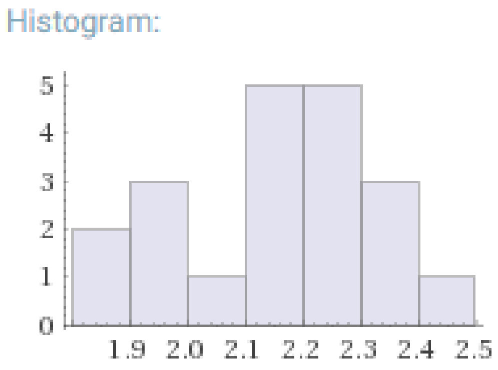
1.8 Histograma

Una presentación gráfica usual para datos cuantitativos es el histograma. Esta gráfica se hace con datos previamente resumidos mediante una distribución de frecuencia, de frecuencia relativa o de frecuencia porcentual. Un histograma se construye colocando la variable de interés en el eje horizontal y la frecuencia, la frecuencia relativa hola frecuencia porcentual en el eje vertical. La frecuencia la frecuencia relativa, o frecuencia porcentual de cada clase se indica dibujando un rectángulo cuya base está determinada por los límites de clase sobre el eje horizontal y cuya altura es la frecuencia, la frecuencia relativa o la frecuencia porcentual correspondiente.

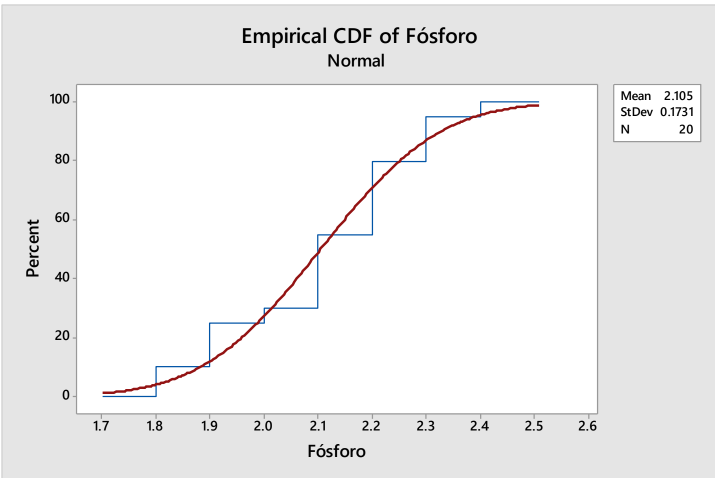
Tabla de Frecuencias

Fosforo	frecuencia	Porcentaje	CumPct
1.8	2	10.00	10.00
1.9	3	15.00	25.00
2.0	1	5.00	30.00
2.1	5	25.00	55.00
2.2	5	25.00	80.00
2.3	3	15.00	95.00
2.4	1	5.00	100.00
	N=20		

La figura de abajo es un histograma de las concentraciones medidas de fosforo (P) para n=20 muestras idénticas de aguas residuales con una concentración conocida de 2 mg/l son:



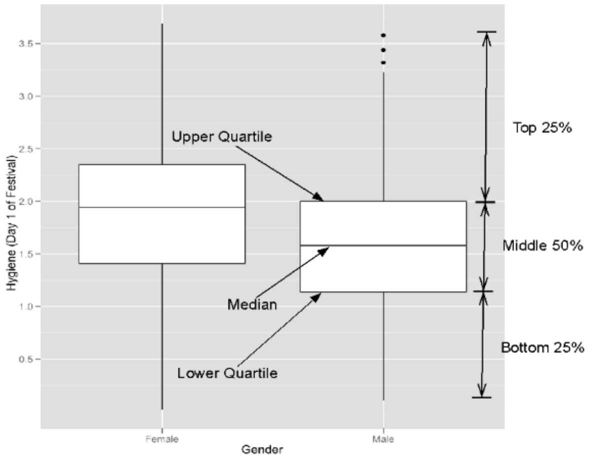
Uno de los usos más importantes de un histograma proveer información acerca de la forma de la distribución. Un gráfico Q-Q plot construido a partir de distribuciones de frecuencia relativa. Se muestra un conjunto de datos que siguen una Distribución normal

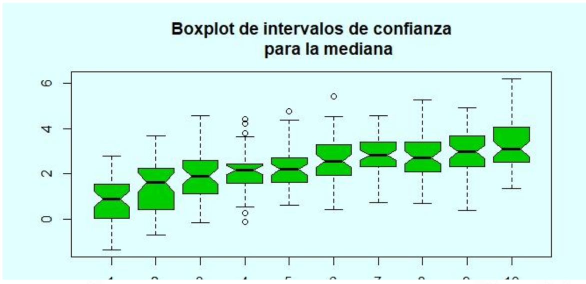


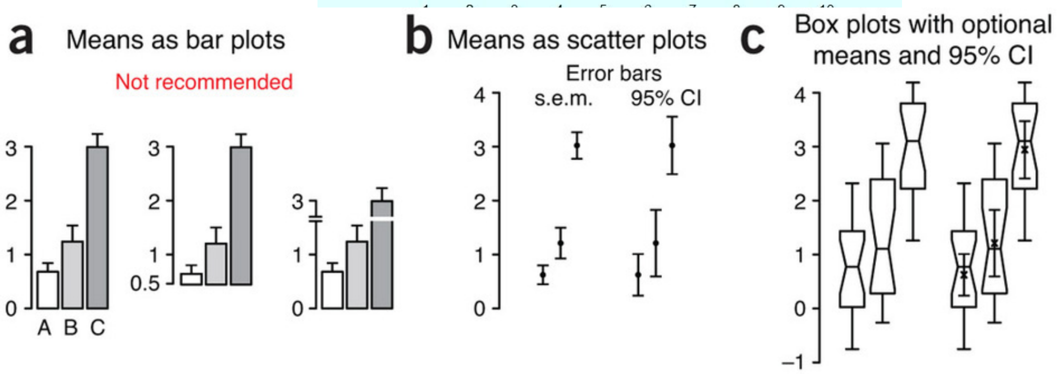
1.9 Diagrama de Cajas o Box Plot.

Este grafico es de mucha ayuda porque ayuda a visualizar la posición de los datos, como también la mediana, y los datos aberrantes.

Utilizado bastante en artículos científicos ayuda a describir cómo se comporta la población visualizando si tienen una distribución normal, el sesgo y curtosis de la misma.







1.10 Medidas de Tendencia Central

Existen varios tipos de medias donde se encuentran la media aritmética, media geométrica la media cortada, y la media harmónica teniendo sus ventajas y desventajas como por ejemplo la media geométrica tiene menor media que la aritmética, pero mayor media que la harmónica.

$$H \leq G \leq A$$

1.10.1 Media Aritmética

Sea una muestra $x_1, x_2, x_3, x_4, x_5..., x_n$ se denota la media aritmética de una muestra a:

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

El mango, una reconocida fruta tropical exótica, se consume mayormente como fruta fresca, pero también puede ser utilizado para preparar mermeladas y confituras, además de sus grandes cualidades alimenticias, el mango ecuatoriano se destaca por su excelente calidad y exquisito sabor.

Las variedades que se cultivan principalmente en el Ecuador son las siguientes: Tommy Atkins, Haden, Kent, Keitt.

Ejemplo:

Se tiene una muestra de tamaño n=21 de la variedad Tommy Atkins:

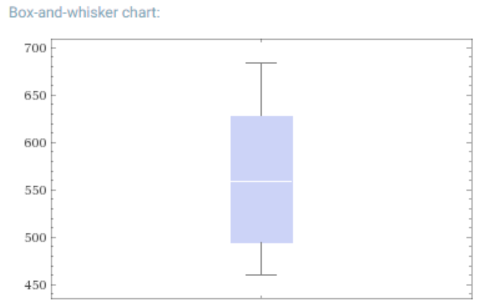
555, 460, 560, 650, 680, 490, 580, 477, 662, 669, 559, 550, 488, 549, 684, 620,469, 562, 496, 521, 569.

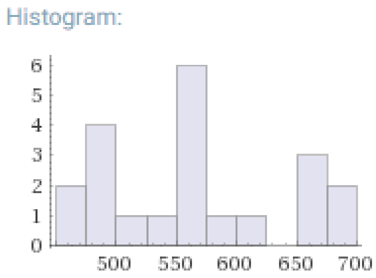
Ordenamos los datos:

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

$$= \frac{460 + 469 + 477 + 488 + 490 + 496 + 521 + 549 + 550 + 555 + 559 + 560 + 562 + 569 + 580 + 620 + 650 + 662 + 669 + 680 + 684}{21}$$

$$\bar{x} = \frac{11850}{21} = 564.28\ gramos$$





1.10.2 Media Cortada

Cortando parte de la cola inferior de la distribución y crea un conjunto de datos asimétricos, uno con valores más conocidos por encima que por debajo de la mediana. La simetría puede ser regreso recortando la cola superior de la distribución.

La media Recortada se puede usar para estimar la media si el subyacente de la distribución es simétrica (no necesariamente normal). En este caso, es un estimador insesgado, pero no tiene mínima varianza.

1.10.3 Media Geométrica

La media geométrica G de n números positivos $x_1, x_2, x_3, x_4, x_5, \ldots, x_n$

$$\bar{x}_g = \sqrt[n]{x_1 x_2 x_3 \ldots x_n}$$

En ocasiones se trabaja con cantidades que cambian en ciertos periodos, como tasas de interés, tasas de crecimiento de insectos en un período t.

1. Básicamente se la utiliza para obtener promedio de índices, porcentajes
2. Incrementos porcentuales, producción u otras actividades etc. (López & González, 2015)

Ejemplo:

Sea una muestra n=11 de porcentajes de daño de un tipo de hongo en trigo

X={0.40, 0.35, 0.2, 0.55, 0.6, 0.48, 0.55, 0.60, 0.40, 0.25, 0.6}

$$\overline{x}_g = \sqrt[11]{(0.4)(0.35)(0.2)(0.55)(0.6)(0.48)(0.55)(0.6)(0.4)(0.25)(0.6)}$$

$$\overline{x}_g = 0.4277$$

Ejercicio

Plomo en agua del grifo. Los datos a continuación son mediciones de plomo en el agua del grifo en un complejo de departamentos. Del Total de n=140 apartamentos muestreados, 93 tenían una concentración por debajo de 5 μg/L. Tomado de (Brown & Mac Berthouex, 2010)

a) Estime la concentración mediana de plomo en los 140 apartamentos.
b) Estime la concentración media aritmética de plomo en los 140 apartamentos.

Pb (μg/L)	Punto medio (x_i)	frecuencia (f_i)	F acumulada	$x_i f_i$
0 - 4.9	2.45	93	93	227.85
5.0-9.9	7.45	26	119	193.7
10-14.9	12.45	6	125	74.7
15-19.9	17.45	4	129	69.8
20-29.9	24.95	7	136	174.65
30-39.9	34.95	1	137	34.95
40-49.9	44.95	1	138	44.95
50-59.9	54.95	1	139	54.95
60-69.9	64.95	0	139	0
70-79.9	74.95	1	140	74.95

a) $\tilde{x} = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} * a_i$

$\tilde{x} = 0 + \frac{\frac{140}{2} - 0}{93} * 5$= 3.763 (μg/L)

b)

$$\bar{x} = \sum \frac{x_i f_i}{N} = \frac{227.85 + 193.7 + 74.7 + 69.8 + 174.65 + 34.95 + 44.95 + 54.95 + 0 + 74.95}{140}$$

$\bar{x} = \sum \frac{950.5}{140} = 6.789$ (μg/L)

1.10.4 Mediana

Sea $\tilde{x}$ la mediana de un conjunto de datos que implica que es el valor intermedio de los datos ordenados de forma creciente (o decreciente).

Para calcular la mediana $\tilde{x}$ sigue las siguientes reglas:

1. ***Si n es par***, la mediana $\tilde{x}$ se obtiene calculando la media de los dos números que se encuentran en la mitad.
2. ***Si n es impar***, la mediana $\tilde{x}$ es el número que se localiza en el centro o exactamente en la mitad de los datos ordenados.

Ejemplo:

Sea la muestra $x_1\ x_2\ x_3 \ldots x_n$ de tamaño n=13 del peso (gr) de cabezas de ajo blanco 37, 42.5, 38.6, 55, 62.3, 39.9, 42.5, 62.8, 88.3, 82.4, 76.4, 91.2, 72.1

Se agrupan los datos X= { 37, 38.6, 39.9, 42.5, 42.5, 55, 62.3, 62.8, 72.1, 76.4, 82.4, 88.3, 91.2}

La información fue tomada de (Balzarini, Monica; Di Rienzo, Julio; Tablada & Bruno, 2011)

1.10.5 Moda

Sea la muestra $x_1\ x_2\ x_3 \ldots x_n$ se define a la **moda** como el valor más frecuente entre los datos. Cuando existen dos valores con la misma frecuencia se dice **bimodal.** Si los datos tienen más de dos valores que presentan la misma frecuencia entonces todos los valores son moda y se le conoce como **multimodal.**

Ejemplo

Una muestra n=14 del número de flores por planta **X= {3**, 5, 8, 6, 5, 3, 9, 4, 8, 3, 4, 7, 9, 7**}**

La moda sería el valor que ocurre con más frecuencia en la muestra en este caso **seria 3 flores**.

1.11 Medidas de Dispersión

1.11.1 Rango

El Rango de un conjunto de datos ordenados es la diferencia entre valor máximo y el valor mínimo.

Rango= X_n-X_1

Ejemplo

Tomando el ejemplo de la variable de numero de flores por planta X= {3, 3, 3, 4, 4, 5, 5, 6, 7, 7, 8, 8, 9, 9**}**

El Rango= X_n-X_1 es igual a Rango= 9-3 =**6 flores**.

1.11.2 Rango Intercuartil

Rango Intercuartil de un conjunto de datos ordenados es la diferencia entre el Cuartil 3 y el Cuartil 1.

Rango Intercuartil= Q_3- Q_1

1.11.3 Varianza.

Existe otro mecanismo para solucionar el efecto de cancelación para entre diferencias positivas y negativas. Si elevamos al cuadrado cada diferencia antes de sumar, desaparece la cancelación: Esta fórmula tiene una desventaja, y es que sus unidades no son las mismas que las de las observaciones, ya que son unidades cuadradas. Esta dificultad se soluciona, tomando la raíz cuadrada de la ecuación anterior:

$$S^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

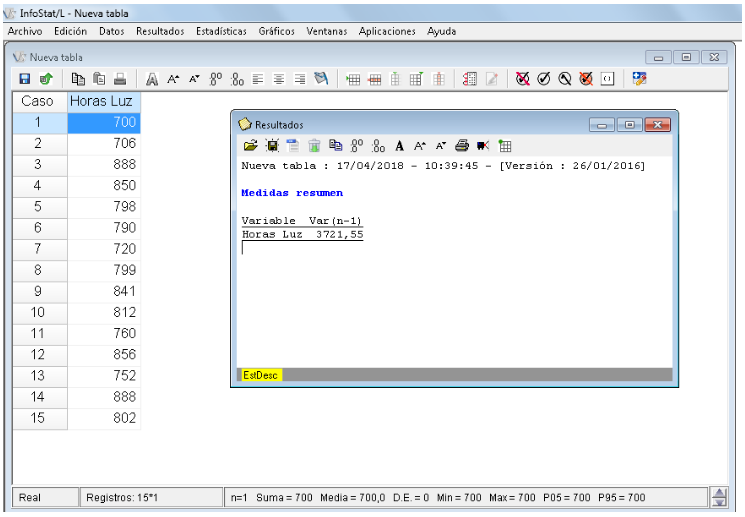
El Programa de Palma Africana ha puesto a disposición de los productores materiales genéticos, como el Híbrido Tenera– INIAP, adaptado a una zona tropical húmeda en donde los suelos son de origen volcánico, con pH de 5,5 a 6,5 de textura franco a franco arenoso, con topografía ondulada, buenas condiciones de drenaje, convenientes promedios de temperatura, 24°C, humedad relativa de 84 a 88%, número de horas luz de 700 a 900 y precipitaciones entre 2500 y 3200 mm anuales; por sus buenas características este material fue escogido por los palmicultores de las zonas del Oriente y del Noroccidente Ecuatoriano.

Supongamos que tenemos una muestra de horas luz en un cultivo de palma africana donde el tamaño de la muestra es n=15

700, 706, 888, 850, 798, 790, 720, 799, 841, 812, 760, 856, 752, 888, 802

a) Obtenga la Varianza muestral.

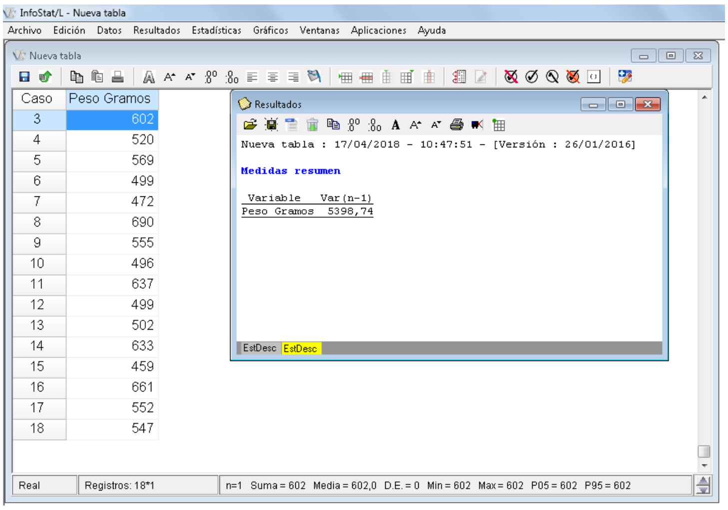
Utilizando el Software Estadístico Infostat obtenemos el siguiente resultado



Con respecto a la variedad TOMMY ATKINS la cual es originaria de la Florida, supuestamente del Haden. Es una fruta de 13 cm de largo y 450 a 700 gramos de peso, con forma ovoide a casi redonda, color con base morado a rojizo, bastante resistente a los daños mecánicos debido a la cáscara gruesa, carece de fibra, tiene buen sabor y de pulpa jugosa.

Calculemos la Varianza para esta variedad de mango a partir de muestra n=18

483, 665, 602, 520, 569, 499, 472, 690, 555, 496, 637, 499, 502, 633, 459, 661, 552, 547.



1.11.4 Desviación estándar.

La varianza se asemeja a la desviación media absoluta en que se basa en la diferencia entre cada valor del conjunto de datos y la media del grupo. Pero se distingue de ella en un muy importante aspecto: cada diferencia se eleva al cuadrado antes de sumarse. En el caso de una población, la varianza se representa con V(X) o, más habitualmente, con la letra griega minúscula σ^2 ("sigma cuadrada"). La fórmula es

$$S = \sqrt{S^2}$$

1.11.5 Usos de la desviación estándar.

La desviación estándar nos permite determinar, con un buen grado de precisión, dónde están localizados los valores de una distribución de frecuencias con relación a la media. El teorema de Chebyshev dice que no importa qué forma tenga la distribución, al menos 75% de los valores caen dentro de + 2 desviaciones estándar a partir de la media de la distribución, y al menos 89% de los valores caen dentro de + 3 desviaciones estándar a partir de la media. Con más precisión: • Aproximadamente 68% de los valores de la población cae dentro de + 1 desviación estándar a partir de la media. • Aproximadamente 95% de los valores estará dentro de + 2 desviaciones estándar a partir de la media. • Aproximadamente 99% de los valores estará en el intervalo que va desde tres desviaciones estándar por debajo de la media hasta tres desviaciones estándar por arriba de la media.

1.11.6 Coeficiente de Variación

El coeficiente de variación es una medida de dispersión la cual nos dice que porcentaje (%) están dispersos las observaciones.

A continuación, veamos un ejemplo donde para las observaciones de 27 nitratos, el promedio de la muestra es de

$$\bar{y} = \frac{6.9 + 7.8 + \cdots + 8.1 + 7.9}{27} = 7.51mg/L$$

La varianza de la muestra es

$$S^2 = \frac{(6.9 - 7.51)^2 + \cdots + (7.9 - 7.51)^2}{27 - 1} = 1.9138\ (mg/L)^2$$

La muestra estándar de derivación es

$$s = \sqrt{1.9138} = 1.38mg/L$$

La varianza demuestra y la desviación estándar de la muestra tienen V=27-1 =26 grados de libertad.

Los datos reportados con dos cifras significativas. El promedio de varios valores se debe calcular con una cifra más que la de los datos, la desviación estándar se debe calcular al menos en cifras significativas.

Calculemos el %CV el cual se define como: $\%CV = \frac{s}{\bar{x}} * 100$

$$\%CV = \frac{1.38}{7.51} * 100 = 0.183 = 18.3\%$$

Es decir, es una dispersión aceptable tal como explica en el siguiente parte del capítulo 1.

1.11.6.1 Escala de Valoración del Coeficiente de Variación (CV)

Una medida de dispersión como CV es de suma importancia, es por eso que hay que tener en cuenta sus niveles o escala de valoración como se muestra a continuación.

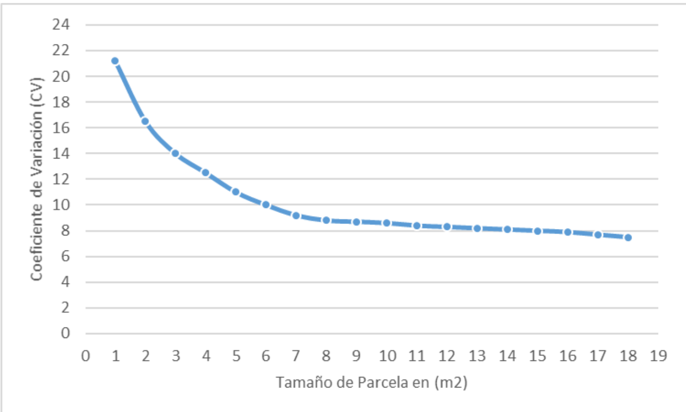
CV ≤ 10%	Poca Dispersión	Distribución Homogénea
10%<CV≤30%	Dispersión Aceptable	
30<CV≤ 50%	Dispersión Alta	Distribución Heterogénea
CV> 50%	Dispersión muy Alta	

Adicionalmente, se recomienda que un trabajo experimental en campo su %CV sea < 30%, y en laboratorio sea %CV < 10%, en laboratorio es mucho más bajo por las condiciones controladas que este incluye.

1.11.7 Relación Coeficiente de variación vs tamaño de parcela en mt^2

La obtención del tamaño adecuado de las unidades de muestreo va a depender del tipo del cultivo con el cual se desarrollarán experimentos en campo. Existen varios métodos como el de máxima curvatura y de regresión lineal en donde el objetivo es determinar el tamaño de la parcela, esto para disminuir la variabilidad en los datos, para eso se mide el coeficiente de variación antes mencionado.

A continuación, presentamos un gráfico que muestra la relación:



1.12 Ejercicios Complementarios Capitulo 1

1) La Temperatura Para el desarrollo normal del cultivo de papa, se requiere una temperatura entre los 6 a 18°C y una precipitación por ciclo de entre 600 a 1,200 mm de agua. (Fuente: INIAP, 2002). Para el año 2014, el promedio de temperatura de las principales provincias productoras de papa, Carchi y Cotopaxi, se encontraron dentro del rango óptimo que requiere el cultivo para su desarrollo; como se observa en la figura 18. Carchi registró una temperatura promedio anual 12.62°C. Mientras que en la provincia de Cotopaxi se registró una temperatura promedio anual de 14.81°C. Fuente: Inamhi, 2014

Con estos Antecedentes, tenemos una muestra de temperaturas durante 20 días en °C

10	12	14	10	11
8	7	12	15	17
10	12	13	14	9
10	9	16	12	15

a) Obtenga la media, mediana y moda
b) Calcule, la varianza, desviación estándar y %CV
c) Concluya sobre los resultados obtenidos en a) y b)

2) El rango medio de precipitación mensual que requiere el cultivo de papa es 100 mm promedio mensual. En la provincia de Carchi, la precipitación acumulada del año 2014 fue de 856.63mm. Durante los meses de enero a junio, la precipitación acumulada fue de 461.33 mm, valor que no fue suficiente para cubrir las necesidades hídricas del cultivo.

Con estos Antecedentes, tenemos una muestra de precipitación media durante 15 días en mm

100	120	140	100	110
89	79	112	150	117
105	112	130	114	90

a) Obtenga la media, mediana y moda
b) Calcule, la varianza, desviación estándar y %CV
c) Concluya sobre los resultados obtenidos en a) y b)

3) En la provincia de Cotopaxi la precipitación acumulada del año 2014 fue de 621.50 mm. Durante los meses de enero a junio, la precipitación fue de 402.40 mm, valor que no abasteció las necesidades hídricas del cultivo como se observa en la Figura 20.

Se tomó una muestra de precipitación media de los últimos 10 años en los meses de enero a junio en mm en este sector.

2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
500	400	600	525.6	723	525	700	402.4	500	600

a) Obtenga la media, mediana y moda
b) Calcule, la varianza, desviación estándar y %CV
c) Concluya sobre los resultados obtenidos en a) y b)

CAPITULO 2

2.1 DISTRIBUCION NORMAL Y TRANSFORMACIONES

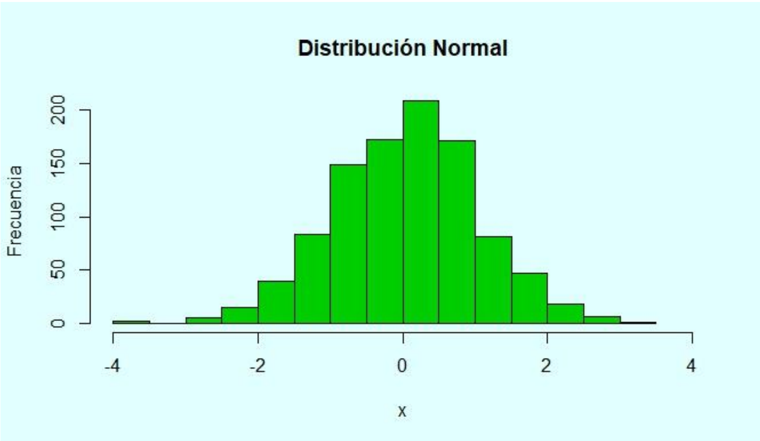
Las observaciones repetidas que difieren debido a un error experimental a menudo varían sobre algún valor central con una distribución de probabilidad con forma de campana que es simétrica y en la cual pequeñas desviaciones ocurren mucho más frecuentemente que las grandes. Una distribución de frecuencia poblacional continua que repite esta condición es la distribución normal (también a veces llamada distribución gaussiana). La distribución normal se caracteriza completamente por su media y varianza y a menudo se describe mediante la notación N (μ,σ^2), que se lee "una distribución normal con media μ y varianza σ^2 ".

La geometría de la curva normal es la siguiente:

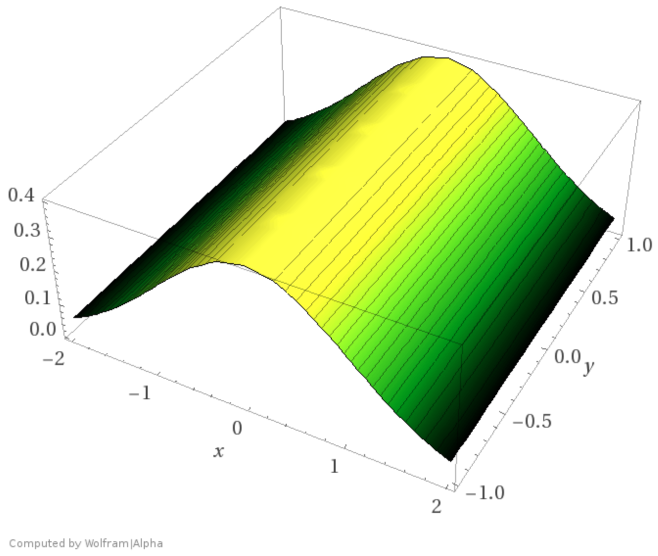
1. El eje vertical (densidad de probabilidad) se escala de manera tal que el área debajo de la curva es la unidad (1.0).

2. La desviación estándar: mide la distancia desde la media hasta el punto de inflexión.

3. Debido a la simetría, las probabilidades son las mismas para las desviaciones negativas y $\alpha_1 = \alpha_4$ y $\alpha_1 + \alpha_2 = \alpha_3 + \alpha_4$.



Es conveniente trabajar con desviaciones normales estandarizadas, $z = (y - \mu)/\sigma$, donde z tiene la distribución N(0,1), debido a que las áreas bajo la curva normal estandarizada están tabuladas.



Ejemplo, Distribución Normal. Gráficamente Determinar si los siguientes datos pudieron haber venido de una distribución normal

A continuación de obtuvieron los siguientes resultados de un conjunto de datos utilizando Statgraphic v.18

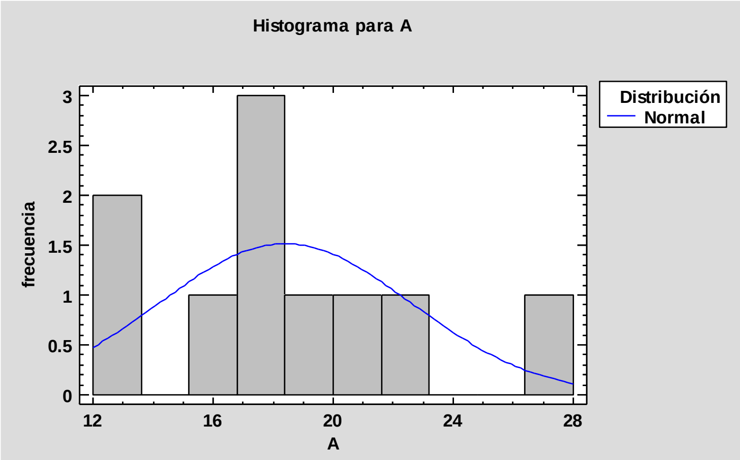
Muestra A

Pruebas de Normalidad para A

Prueba	*Estadístico*	*Valor-P*
Estadístico W de Shapiro-Wilk	0.944748	0.606951

El StatAdvisor
Esta ventana muestra los resultados de diversas pruebas realizadas para determinar si A puede modelarse adecuadamente con una distribución normal. La prueba de Shapiro-Wilk está basada en la comparación de los cuartiles de la distribución normal ajustada a los datos.

Debido a que el valor-P más pequeño de las pruebas realizadas es mayor ó igual a 0.05, no se puede rechazar la idea de que A proviene de una distribución normal con 95% de confianza.



Muestra B

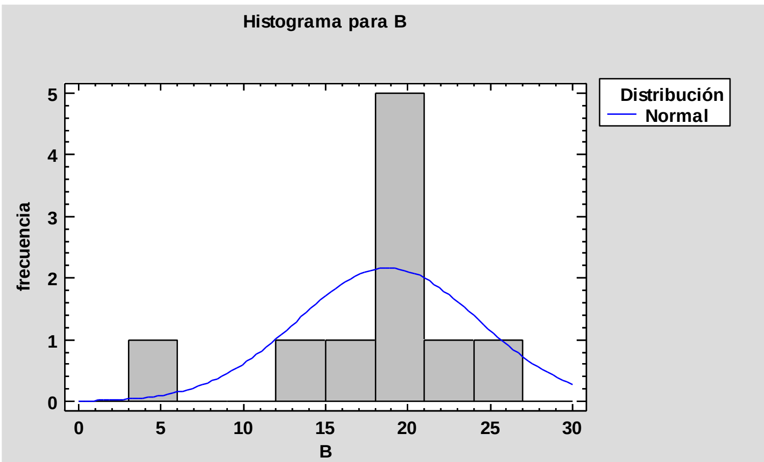
Pruebas de Normalidad para B

Prueba	*Estadístico*	*Valor-P*
Estadístico W de Shapiro-Wilk	0.802515	0.0155552

El StatAdvisor

Esta ventana muestra los resultados de diversas pruebas realizadas para determinar si B puede modelarse adecuadamente con una distribución normal. La prueba de Shapiro-Wilk está basada en la comparación de los cuartiles de la distribución normal ajustada a los datos.

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B proviene de una distribución normal con 95% de confianza.



2.2 La Distribución *t Student*

La estandarización de una variable aleatoria normal requiere que tanto η y σ sean conocidos. En la práctica, sin embargo, no podemos calcular $z = (\bar{x} - \mu)/\sigma$ porque σ se desconoce. En su lugar, sustituimos *s* y calculamos σ la estadística *t* :

$$t = \frac{\bar{x} - \mu}{s}$$

El valor de μ puede ser conocido (por ejemplo, porque es un estándar primario) o puede suponerse cuando se construye una hipótesis que se probará (por ejemplo, se supone que la diferencia entre dos tratamientos es cero). Bajo ciertas condiciones, que se dan a continuación, tiene *t* una distribución conocida, llamada Distribución del estudiante, o simplemente la distribución *t*. (Brown & Mac Berthouex, 2010)

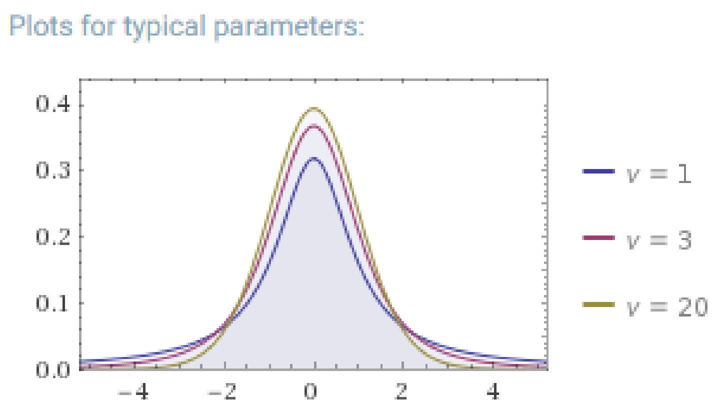
La distribución *t* es en forma de campana y simétrica (como la distribución normal), pero las colas de la distribución *t* son más anchas que las colas de la distribución normal. el ancho de la distribución *t* depende del grado de incertidumbre es s^2, que se mide por los grados por la libertad *v* en la que se basa esta estimación de s^2. Cuando el tamaño de la muestra es infinito ($v = \infty$), no hay incertidumbre en s^2 (porque $s^2 = \sigma^2$) y la distribución *t* se convierte en la distribución normal estándar. Cuando el tamaño de la muestra es pequeño ($v \leq 30$), la variación en s^2 aumenta. Esto se refleja en la expansión de la distribución *t* a medida que disminuye el número de grados de libertad de s^2. El área de la cola bajo la curva en forma de campana de la distribución *t* es la probabilidad de que *t* exceda un valor dado. Una vista de la tabla *t* se reproduce en la Tabla 3 de Anexos.

Las condiciones bajo las cuales la cantidad $t = (\bar{x} - \mu)/s$ tiene una distribución *t* con *v* grados de libertad son:

1. $\bar{x}$ se distribuye normalmente con media μ y varianza σ^2.

2. *s* se distribuye independientemente de la media; es decir, la varianza de la muestra no aumenta ni disminuye a medida que la media aumenta o disminuye.

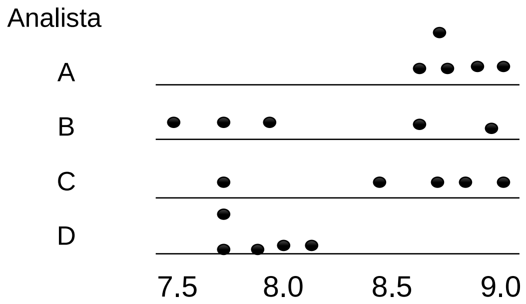
3. La cantidad s^2, que tiene v grados de libertad, se calcula normalmente y las observaciones distribuidas independientemente tienen varianza σ^2.



2.3 Exactitud, Sesgos y Precisión

La exactitud es una función de ambos sesgo y precisión. Como se muestra en el ejemplo y en la figura 2, sesgos los errores sistemáticos y la precisión en los grados de dispersión en los datos. Los métodos exactos pueden tener buena y sesgo cercano a cero. Inexactos pueden tener pobre precisión, inaceptables sesgos o ambos.

El sesgo (error sistemático) puede eliminarse, una vez que se identifique, mediante controles cuidadosos de la técnica experimental. No se puede promediar haciendo más mediciones. A veces, no se puede identificar el sesgo porque el valor real subyacente se desconoce.



Sesgos	precisión	Exactitud
Grande	Bueno	Pobre
Pequeño	Pobre	Pobre
Grande	Pobre	Pobre
Ausente	Bueno	Bueno

Figura 2. la exactitud es una función de sesgos y buena precisión. (Brown & Mac Berthouex, 2010)

2.4 Normalidad, Aleatoriedad, e Independencia

Las tres propiedades importantes en los cuales recaen muchos procedimientos estadísticos son normalidad, aleatoriedad e independencia. De estos, la normalidad es la que más parece preocupar a la gente. No es siempre lo más importante.

Normalidad significa que se supone que el termino de error en una medida y que se obtienen de una distribución de probabilidad normal de una distribución de probabilidad normal. Esta es la distribución familiar, simétrica en forma de campana, una tendencia a la distribución de errores. Este es el efecto limítrofe central. Se basa en el supuesto de que hay varias fuentes de error, que no es la única fuente dominante, y que el error general es una combinación lineal de errores distribuidos independientemente. Estas condiciones parecen ser muy restrictivas, pero frecuentemente no siempre existen. Aun cuan no existen, carecen de normalidad no es necesariamente un serio problema. Las transformaciones están disponibles para hacer errores no normales “de tipo normal” (Brown & Mac Berthouex, 2010).

Muchos de ellos utilizaron procedimientos estadísticos, incluidos aquellos que se basaron directamente en la comparación de los valores promedio (como las pruebas *t* para comparar los valores promedio y el análisis de las diversas pruebas para comparar varios valores) son robustos a las desviaciones de la normalidad. Robusto significa que tiende a generar conclusiones correctas incluso cuando se aplica a los datos que normalmente no se distribuyen.

Aleatorias significa que las observaciones se extraen de una población

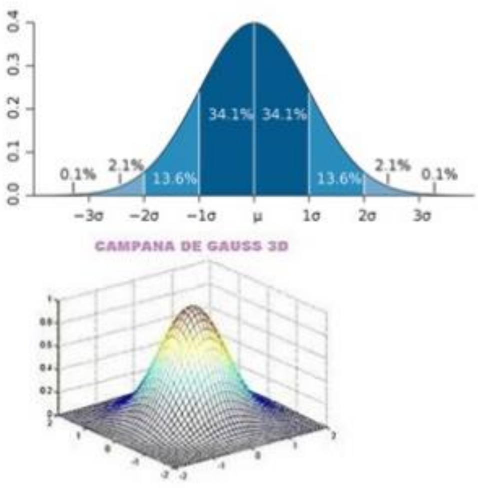
De una manera que da a cada elemento de una población de igual oportunidad de ser aleatorización de muestreo es mejor de seguro que las observaciones serán independientes.

Ejemplos

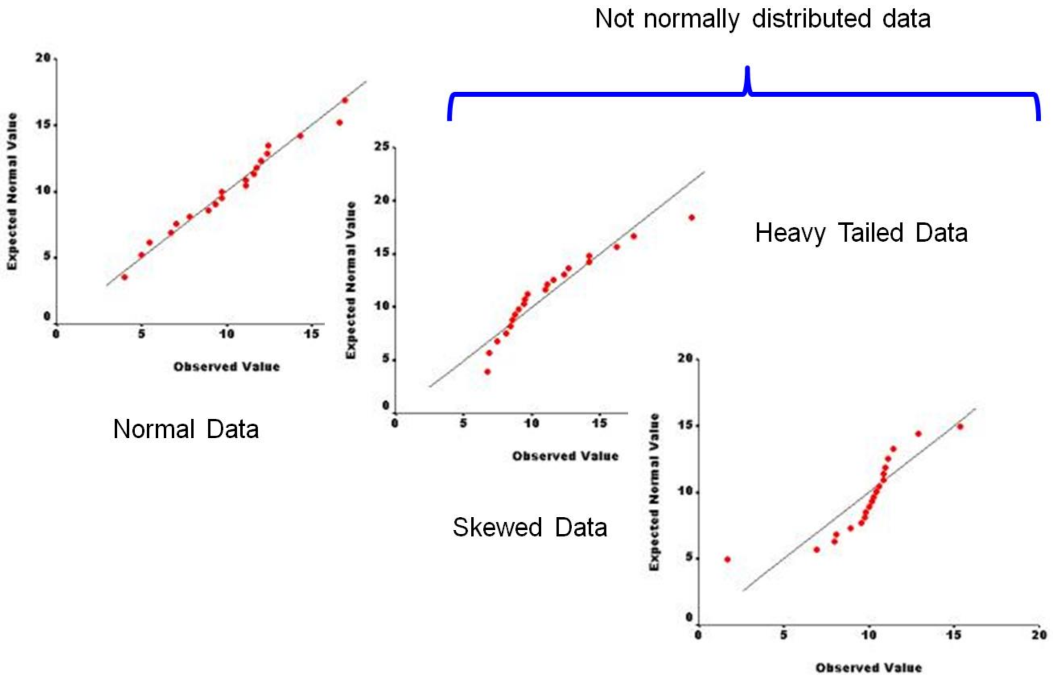
Los errores en los datos de laboratorio de nitrato se verifican para determinar la aleatoriedad al trazar los errores $e_i = y_i - n$ si los errores so aleatorios, la gráfica no han tenido ningún patrón. En una gráfica de este tipo que muestra e_i en el orden de las observaciones, no hay ninguna relación aleatoria.

Imagínate formas en que los errores de las mediciones de nitrato pueden ser no aleatorios. Supongamos, por ejemplo, el proceso de medición se ha desviado de manera que las medicines totales sean altas y las posteriores una gráfica de los errores por cada análisis tomadas una tendencia (errores positivos seguidos de resultados negativos.

2.5 Normalidad de Datos



La distribución Normal o Distribución de Gauss es la forma en que se distribuyen en la naturaleza los diversos valores numéricos de las variables continuas.



Se debe comprobar por tanto la normalidad de la variable dependiente si la muestra **no es muy grande.**

Se puede comprobar:

- Que el máximo y el mínimo quedo dentro del intervalo definido Media ± 3 desviaciones estándar.
- Que la asimetría en valor absoluto sea menor a 2 veces su error estándar: |Asimetría |< 2 error estándar de asimetría.
- Que la curtosis en valor absoluto sea menor a 2 veces su error estándar: |Curtosis |< 2 error estándar de curtosis.

  Estos requisitos más bien son válidos con la muestra es pequeña. Si no se cumple con la condición de normalidad, se puede optar por la transformación de los datos utilizando logaritmo para una aproximación a la normalidad de los mismos.

He aquí algunos ejemplos utilizando variables cuantitativas en el programa Infostat.

Shapiro-Wilks (modificado)

```
Ho: Los datos siguen una distribución normal
V
H1: Los datos NO siguen una distribución normal
```

Ejemplo

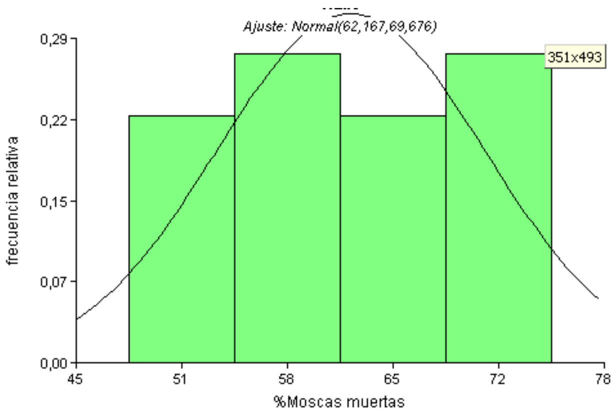
Medidas resumen

Resumen	%Moscas muertas
n	18,00
Media	62,17
D.E.	8,35
Var(n-1)	69,68
CV	**13,43**
Mín	48,00
Máx	75,00
Mediana	61,50

Prueba normalidad Variable % Moscas muertas

Shapiro-Wilks (modificado)

Variable	n	Media	D.E.	W*	p(Unilateral D)
%Moscas muertas	18	62,17	8,35	0,93	0,3876



Ho: Los datos siguen una distribución normal
vs
H1: Los datos NO siguen una distribución normal

0.3876 >0.05 por lo tanto acepto la hipótesis nula

Los datos son normales

He aquí otro ejemplo utilizando variables cuantitativas el cual es Rendimiento tn/ha en el programa Infostat.

16 14 9 12 13 12 12 7 9

12 10 9 8 9 8 10 7 8 7 8

TEST de NORMALIDAD

Hipotesis:

Ho: Los datos siguen una distribución Normal

V

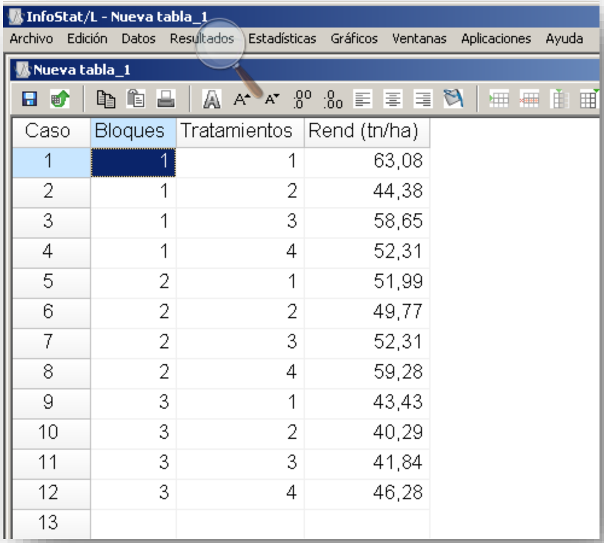
H1: Los datos NO siguen una distribución Normal

Shapiro-Wilks (modificado)

Variable	n	Media	D.E.	W*	p(Unilateral D)
Rend (tn/ha)	20	10,00	2,55	0,89	0,0569

El p-valor 0.0569 > 0.05 por lo tanto Acepto Ho, siendo los datos normales

Rendimiento (tn/ha) en una variedad de caña de azúcar sometida a cuatro tratamientos y 3 repeticiones (Bloques) en un experimento DBCA



Caso	Bloques	Tratamientos	Rend (tn/ha)
1	1	1	63,08
2	1	2	44,38
3	1	3	58,65
4	1	4	52,31
5	2	1	51,99
6	2	2	49,77
7	2	3	52,31
8	2	4	59,28
9	3	1	43,43
10	3	2	40,29
11	3	3	41,84
12	3	4	46,28
13			

TEST de NORMALIDAD

Hipotesis

Ho: Los datos siguen una distribución Normal

vs

H1: Los datos NO siguen una distribución Normal

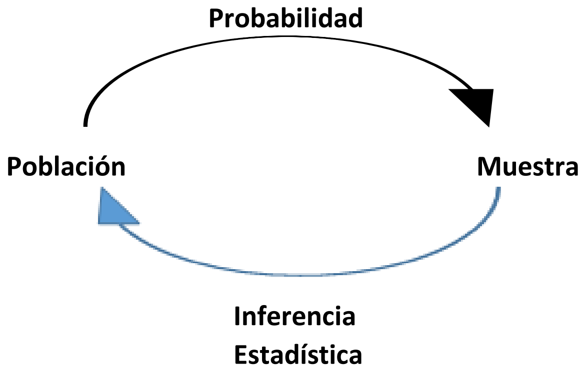
Shapiro-Wilks (modificado)

Variable	n	Media	D.E.	W*	p(Unilateral D)
Rend (tn/ha)	12	50,30	7,36	0,92	0,4419

El p-valor 0.4419 > 0.05 por lo tanto Acepto Ho, siendo los datos normales

2.6 Estadística Inferencial

Estudia los métodos necesarios para extraer o inferir conclusiones validas e información sobre una población a partir del estado experimental de una dicha población.



No se puede hacer estimaciones hasta saber que la muestra se comparta con la población.

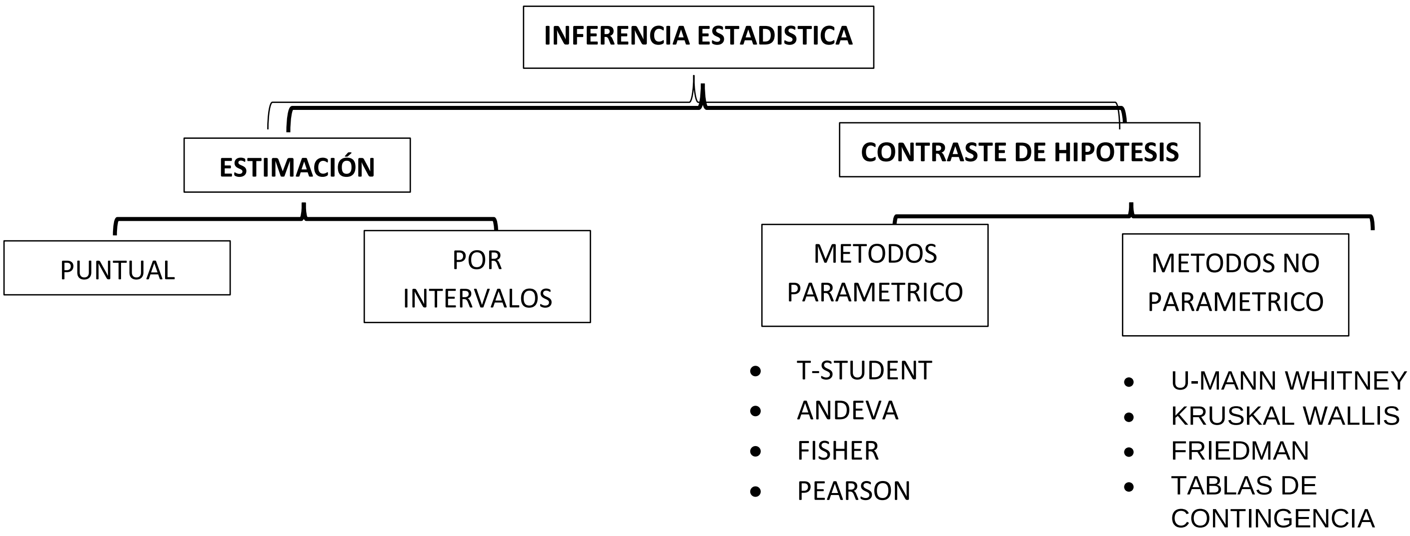
Métodos dependen de la información que se tiene y como este se comparta.

1. Se conoce la distribución de la población
   Se determina los diferentes parámetros de dicha distribución ej: (media, varianza)

Para esto se utilizan los <u>Métodos Paramétricos</u>.

Cuando la distribución de la población es desconocida se utilizan los <u>Métodos no Paramétricos</u>

2. Procedimientos – estimación de parámetros
   - Estimación Puntual – valor especifico
   - Estimación por intervalos de confianza



2.7 Estimación por intervalos de confianza

Interesa dar una estimación y precisar la incertidumbre de dicha estimación, en vez de calcular un único estimador, se determinan dos estimadores. Al valor conocido que toma el intervalo aleatorio en una muestra en particular se le llama estimación por intervalos.

$\rho(L_1 < \beta < L_2) = 1{-}\propto$

$L_1\, L_2 = Intervalos\ de\ confianza$

$1{-}\propto = nivel\ de\ confianza$

$$P\left(\bar{X} - Z_{\propto/2}.\, {}^{\sigma}/_{\sqrt{n}} < \mu < \bar{X} + Z_{\propto/2}.\, {}^{\sigma}/_{\sqrt{n}}\right) = 1{-}\propto$$

Ejemplos

Se encuentra que la concentración media de zinc que se obtiene en una muestra de mediciones en 36 sitios diferentes de un rio es de 206 gramos por ml. Calcule los intervalos de confianza del 95% y 99% para la concentración media de zinc en el rio. Suponga que la desviación estándar de la población es de 0.3 gramos por ml.

n=36
$\bar{X}$=2.6
$1-\alpha$ =95%
$1-\alpha$ =99%
σ=0.3 gr/ml

$1-\alpha$ =95% $\rightarrow Z_{\propto/2} = 1.96$

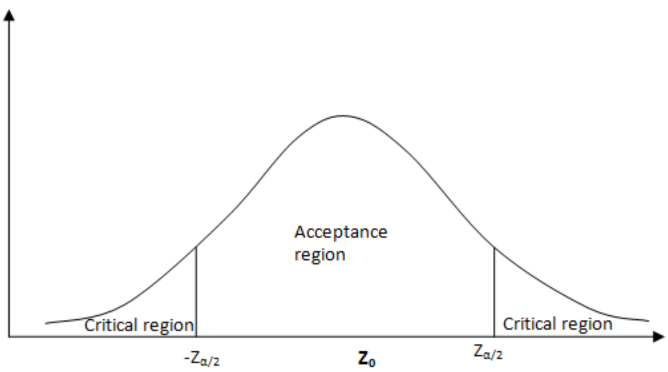
$1-\alpha$ =99% $\rightarrow Z_{\propto/2} = 2.58$

$$\mathbf{95}\% \rightarrow P(2.6 - (1.96).\tfrac{0.3}{\sqrt{36}} < \mu < 2.6 + (1.96).\tfrac{0.3}{\sqrt{36}}) = 0.95$$

$$P(2.502 < \mu < 2.698) = 0.95 \quad \rightarrow \quad intervalos\ de\ confianza$$

$$\mathbf{99}\% \rightarrow P(2.6 - (2.58).\tfrac{0.3}{\sqrt{36}} < \mu < 2.6 + (2.58).\tfrac{0.3}{\sqrt{36}}) = 0.99$$

$$P(2.474 < \mu < 2.729) = 0.99 \quad \rightarrow \quad intervalos\ de\ confianza$$



Se debe saber que tan grande debe ser la muestra para poder estar seguro de que el error al estimar μ sera mejr que una cantidad especifica *e.*

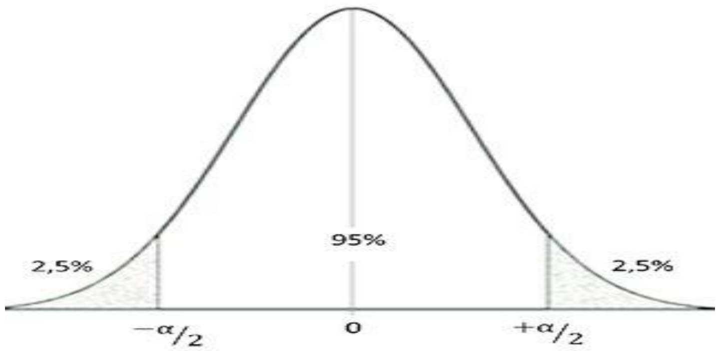
Si usamos $\bar{X}$ como una estimación de μ, podemos tener 100% de confianza en el error no excedera a una cantidad especifica *e* cuando el tamaño de la muestra sea:

$$n = \left(\frac{Z_{\propto/2}.\sigma}{e}\right)^2 \quad \text{ó} \quad n = \left[\!\left[\frac{(Z_{\propto/2})^2.\sigma^2}{e^2}\right]\!\right]$$

→la formula solo es alicable si se conoce la varianza de la población.

Si se desconoce la varianza: se usa *t-student.*

$$\bar{x} \pm Z_{\propto/2}.\left(\frac{s}{\sqrt{n}}\right)$$



CARACTERISTICAS DE LOS	INTERVALOS
Si se conoce la varianza poblacional (σ^2)	$P\left(\bar{X} - Z_{\propto/2}.\sigma/\sqrt{n} < \mu < \bar{X} + Z_{\propto/2}.\sigma/\sqrt{n}\right) = 1-\propto$
Si se conoce la varianza poblacional (σ^2), n<30 y la poblacion se distribuye normalmente	$P\left(\bar{X} - t_{\propto/2} * s\sqrt{1+\frac{1}{n}} < \mu < \bar{X} + t_{\propto/2} * s\sqrt{1+\frac{1}{n}}\right) = 1-\propto$
Si se conoce la varianza poblacional (σ^2), n≥30 y la poblacion se distribuye normalmente	$P\left(\bar{X} - Z_{\propto/2}.s/\sqrt{n} < \mu < \bar{X} + Z_{\propto/2}.s/\sqrt{n}\right) = 1-\propto$

Una muestra aleatoria de 10 barras energéticas de chocolate de cierta marca tiene una media de 230 calorías por barra y una desviación estándar de 15 calorías. Construya un intervalo de confianza del 99% para el contenido medio verdadero de calorías de esta marca de barras energéticas de chocolate. Suponga que la distribución del contenido calórico es aproximadamente normal.

n=10
$\bar{X}$=230
s=15 calorías
$1-\propto = 0.99 \rightarrow t_{\propto/2}$
v=n-1=10-1=9 grados de libertad

$$P\left(\bar{X} - t_{\propto/2}.s\sqrt{1+\frac{1}{n}} < \mu < \bar{X} + t_{\propto/2}.s\sqrt{1+\frac{1}{n}}\right) = 1-\propto$$

$$P\left(230 - (3.250)(15)\sqrt{1+\frac{1}{10}} < \mu < 230 + (3.250)(15)\sqrt{1+\frac{1}{10}}\right) = 0.99$$

$$P(178.87 < \mu < 281.13) = 0.99$$

Las estaturas de una muestra aleatoria de 50 estudiantes de la carrera de ingeniería en agronomía tienen una media de 174.5 cm y una desviación estándar de 6.9 cm. Construya un intervalo de confianza del 98% para la estatura media de todos los estudiantes universitarios.

n=50
$\bar{X}$=174.5 cm
s=6.9 cm
$1-\propto = 0.98$

$$P\left(\bar{X} - z_{\propto/2}.({}^{S}/_{\sqrt{n}}) < \mu < \bar{X} + z_{\propto/2}.({}^{S}/_{\sqrt{n}})\right) = 1-\propto$$

$$P\left(174.5 - 2.33\left({}^{6.9}/_{\sqrt{50}}\right) < \mu < 174.5 + 2.33({}^{6.9}/_{\sqrt{50}})\right) = 0.98$$

$$P(172.23 < \mu < 176.77) = 0.98$$

Una muestra aleatoria de 100 propietarios de automóviles de la provincia del Guayas revela que estas conducen su automóvil, en promedio 23.500 km por año con una desviación estándar de 3900km. Suponga que la distribución de las mediciones es aproximadamente normal. Construya un intervalo de confianza del 99% para el numero promedio de km que un propietario de un automóvil conduce anualmente en guayas.

n=100
$\bar{X}$=23.500 km
s=3.900
$1-\propto = 0.99$

$$P\left(\bar{X} - z_{\propto/2}.{}^{S}/_{\sqrt{n}} < \mu < \bar{X} + z_{\propto/2}.{}^{S}/_{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(23500 - 2.575\left({}^{3900}/_{\sqrt{100}}\right) < \mu < 23500 + 2.575({}^{3900}/_{\sqrt{100}})\right) = 0.99$$

$$P(22495.75 < \mu < 24504.25) = 0.99$$

2.8 Transformación de los datos

Para que los residuales tengan una varianza homogénea y se ajusten a una distribución normal. Con frecuencia la transformación logra el doble objetivo de normalidad y aditividad de los datos.

En estos casos, los datos, los análisis realizados con los datos transformados son perfectamente válidos. Para la presentación de resultados, las medias deben transformarse volviendo a la escala original, pero no es apropiado hacer lo mismo con los errores estándar o varianzas.

Si existen dudas sobre la transformación más adecuada es necesario examinar la relación entre varianzas y medias y elegir la transformación para la cual la relación sea mínima. No obstante, cuando no es posible hallar una transformación que normalice los datos, deben hacerse otros métodos de análisis o pruebas no paramétricas. (Fernández Escobar, Trapero, & Domínguez, 2010)

2.8.1 Logaritmo [log (x)]

Está indicada cuando existe una distribución muy sesgada a la derecha o a la izquierda, si existen valores pequeños, menores que 10, y especialmente ceros, es más adecuado la transformación log (x+1).

2.8.2 Raíz Cuadrada [$\sqrt{\boldsymbol{x}}$]

Está indicada cuando estamos tratando con conteos o recuentos de acontecimientos siguen una distribución Poisson moderadamente sesgada a la derecha o moderadamente a la izquierda, si existen valores pequeños, menores que 10, y especialmente ceros, es más adecuado la transformación $\sqrt{x+1}$.

2.8.3 Inversa [1/x]

Está indicada para casos pocos comunes en los datos presentan una alta variabilidad y las varianzas son proporcionales a las medias elevadas a la 4. Los datos presentan una distribución fuertemente sesgada a la derecha (J invertida) o moderadamente a la izquierda (en forma de J), si existen valores pequeños especialmente ceros, es más adecuado sumarle 1 para que no existe una indefinición.

2.8.4 Angular o Arcoseno [arcsen $\sqrt{\boldsymbol{x/100}}$]

Cuando los datos son proporciones o porcentajes de la muestra total, tiene una distribución binomial en vez de una distribución normal. En los datos binomiales, las varianzas tienden a ser pequeñas en los dos extremos de los intervalos de valores (cercanos a 0 y 100%), pero mayores en el medio (alrededor del 50%). Cuando el intervalo de porcentajes esta entre o y 20, o bien 80 y 100, pero no ambos, se recomienda la transformación raíz cuadrada.

Conteo de Plankton en 20 Réplicas de muestras de agua de cinco estaciones en un reservorio																				
Estación 1	0	2	1	0	0	1	1	0	1	1	0	2	1	0	0	2	3	0	1	1
Estación 2	3	1	1	1	4	0	1	4	3	3	5	3	2	2	1	1	2	2	2	0
Estación 3	6	1	5	7	4	1	6	5	3	3	5	3	4	3	8	4	2	2	4	2
Estación 4	7	2	6	9	5	2	7	6	4	3	5	3	6	4	8	5	2	3	4	1
Estación 5	12	7	10	15	9	6	13	11	8	7	10	8	11	8	14	9	6	7	9	5

A continuación, veamos un ejemplo de transformación de datos en el conteo de Plankton. El siguiente ejercicio es tomado de los siguientes autores (Brown & Mac Berthouex, 2010)

Fuente: Methods for statistical analysis of samples of Benthic Invertebrates

Estación			**1**	**2**	**3**	**4**	**5**
Datos sin Transformar		$\bar{x}$ =	0.85	2.05	3.90	4.60	9.25
		$s^2{}_x =$	0.77	1.84	3.67	4.78	7.57
Transformado	y=$\sqrt{x+c}$	$\bar{y}$ =	1.10	1.54	2.05	2.20	3.09
		$s^2{}_y =$	0.14	0.20	0.22	0.22	0.19

Tenemos los resultados de los datos podemos observar que los datos sin transformar la varianza no es constante, mientras que los datos transformados su varianza ***S^2_y es constante,*** la idea de la transformación es obtener una menor volatilidad con una varianza mayormente constante.

EL efecto de raíz cuadrada y transformación logaritmo es hacer los valores grandes en valores relativamente más pequeño. La transformación logarítmica es más potente que la de raíz cuadrada. Cuando la muestra de datos contiene ceros la transformación logarítmica se la agrega un constante **c.** Usualmente el valor de c es arbitrario escogido entre 0.5 ó 1.

2.9 Ejercicios complementarios del capítulo 2

1. Se hicieron mediciones para mostrar la densidad de esta bacteria en tres estaciones para medir los niveles de contaminación.

Ocho mediciones duplicadas en tres estaciones de muestreo

y = Bacteria/100 mL			x = log10 (Bacteria/100 mL)		
1	2	3	1	2	3
27	225	1020	1.431	2.352	3.009
11	99	136	1.041	1.996	2.134
48	41	317	1.681	1.613	2.501
36	60	161	1.556	1.778	2.207
120	190	130	2.079	2.279	2.114
85	240	601	1.929	2.38	2.779
18	90	760	1.255	1.954	2.889
130	112	240	2.144	2.049	2.38

a) Grafique los datos para ***x*** y ***y***, y comente sobre distribución.

b) Obtenga media y la varianza para ***y***.

c) Obtenga media y la varianza para ***x*** donde se sugiere usar una transformación logarítmica.

d) Comente sobre las diferencias entre **y** como variable original y ***x*** como variable transformada.

2. Plomo en el suelo, examine la distribución de las 36 mediciones de plomo (mg/kg) en el suelo y recomiende una transformación que haga que los datos sean casi simétricos y normales.

7.6	32	5	14	18	2.3	52	10	3.3	38	3.4	4.3	0.1	5.7	0.1	0.1	4.4
0.42	0.1	16	1.2	0.1	3.2	0.43	1.4	5.9	0.23	0.1	0.1	0.23	0.29	5.3	2	1

a) Grafique los datos para ***x*** y ***y***, y comente sobre distribución.

b) Obtenga media y la varianza para ***y***.

c) Obtenga media y la varianza para ***x*,** escoja 2 tipos de transformaciones para ***x***.

d) Comente sobre las diferencias entre **y** como variable original y ***x*** como variable transformada.

Ejercicios tomados de (Brown & Mac Berthouex, 2010)

CAPITULO 3

3.1 CORRELACION Y REGRESION LINEAL

3.1.2 Medición de relaciones

Existen tres tipos para medir la relación entre variables estas son:

- Diagramas de dispersión
- Covarianza
- Coeficiente de correlación de Pearson

También existe la medición de estas relaciones entre variables no paramétricas

- Spearman
- Kendall (no paramétrica medida entre dos variables)

Dos variables han sido medias y graficadas en un diagrama de dispersión sugiriendo que hay una relación lineal entre ellos.

Tabla 3.1

	Temperatura(°C)	presión(mm)
1	0	0.0002
2	20	0.0012
3	40	0.0060
4	60	0.0300
5	80	0.0900
6	100	0.2700
7	120	0.7500
8	140	1.8500
9	160	4.2000
10	180	8.8000
11	200	17.3000
12	220	32.1000
13	240	57.0000
14	260	96.0000
15	280	157.0000
16	300	247.0000
17	320	376.0000
18	340	558.0000
19	360	806.0000

Da la tabla 3.1 obtenemos utilizando RStudio la covarianza (x,y)= 19157.32

```
> cov(pressure)
            temperatura  presión
temperatura    12666.67 19157.32
presión        19157.32 50455.29
```

Uno de las desventajas de la covarianza es las unidades en que son medidas las variables, si son medidas en millas puede ser 3.25, pero si se la transforma a km, la covarianza resultaría en 10.

$$cov(X,Y) = \frac{1}{n}\sum(x_i - \overline{x})(y_i - \overline{y})$$

Una solución para esto es estandarizar la covarianza diviendola para las desviaciones estándar de cardar variable. La versión estandarizada de la covarianza es conocida como el coeficiente de correlación es un estadístico que cuantifica la fuerza de la relación lineal entre dos variables es el cual se encuentra entre $-1 \leq \rho \leq 1$.

La correlación puede, pero no necesariamente, indicar casualidad. Observando que ***y*** aumenta cuando ***x*** aumenta, no significa que un cambio en ***x*** provoca un cambio en ***y***. Ambos ***x*** y ***y*** pueden ser resultado de una tercera variable ***z***.

$$\rho(X,Y) = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}}$$

El coeficiente de correlación varía entre $-1 \leq \rho \leq 1$

Escala de medición

0 no hay relación
± 0.1 baja relación
± 0.3 media relación
± 0.5 alta relación

Ver la Tabla A6 en anexos para los valores críticos del coeficiente de correlación lineal.

3.2 Coeficiente de determinación R^2

Elevando el coeficiente de correlación al cuadrado se obtiene la proporción de una variable compartida con las otras.

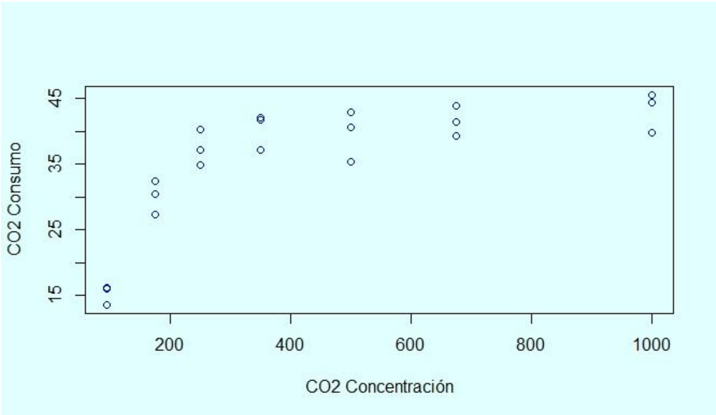
$$R^2 = \frac{(S_{xy})^2}{S_{xx}S_{yy}} = \frac{\left[\sum(x_i - \overline{x})(y_i - \overline{y})\right]^2}{\left[\sum(x_i - \overline{x})^2\right]\left[\sum(y_i - \overline{y})^2\right]}$$

$$0 \leq R^2 \leq 1$$

El coeficiente de determinación es la participación de las variables independientes en el modelo realizado, el cual puede ser leído en porcentaje.

Concentración ambiental de dióxido de carbono (μl/l)

	Plant	concentración	Consumo
1	Qn1	95	16
2	Qn1	175	30.4
3	Qn1	250	34.8
4	Qn1	350	37.2
5	Qn1	500	35.3
6	Qn1	675	39.2
7	Qn1	1000	39.7
8	Qn2	95	13.6
9	Qn2	175	27.3
10	Qn2	250	37.1
11	Qn2	350	41.8
12	Qn2	500	40.6
13	Qn2	675	41.4
14	Qn2	1000	44.3
15	Qn3	95	16.2
16	Qn3	175	32.4
17	Qn3	250	40.3
18	Qn3	350	42.1
19	Qn3	500	42.9
20	Qn3	675	43.9
21	Qn3	1000	45.5



Con este diagrama de dispersión procederemos en RStudio a obtener la correlación lineal entre concentracion y consumo de CO2

> Data_CO2

A tibble: 21 x 4

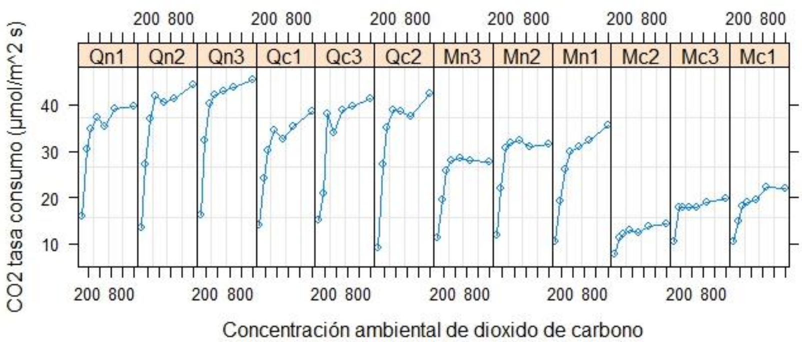
	Plant	concentracion	consumo
1	Qn1	95	16
2	Qn1	175	30.4
3	Qn1	250	34.8
4	Qn1	350	37.2
5	Qn1	500	35.3
6	Qn1	675	39.2
7	Qn1	1000	39.7
8	Qn2	95	13.6
9	Qn2	175	27.3
10	Qn2	250	37.1

... with 11 more rows

```
> cor(Data_CO2$concentracion, Data_CO2$consumo)
[1] 0.7038936
```

Debido el conjunto de datos tiene tres variables dos numéricas y otra de tipo string es por eso que debemos decirle a R que escoja solo las dos columnas con datos numéricos y por defecto utiliza la correlación de Pearson.

La correlación es positiva 0.703



Utilizando **R**Studio obtenemos que la correlación lineal entre temperatura °C y presión de mercurio en mm es 0.7577 la cual es positiva y directamente proporcional.

```
> cor(presion, method = 'pearson')
```

```
            temperatura   presion
temperatura   1.0000000 0.7577923
presion       0.7577923  1.0000000
```

3.3 Introducción a la regresión Lineal

Una forma razonable de relación entre la respuesta ***Y*** y el regresor ***X*** es la relación lineal, existe la regresión lineal simple, y múltiple. El modelamiento a partir de la regresión lineal dependerá de la correlación entre las variables independientes y la variable dependiente. La correlación explicada en este capítulo nos dirá si la relación entre las variables es significativa y su fuerza lineal para ser escogidas en el modelo.

Ejemplos

Modelo de Regresión Lineal Simple

$$Y_{Daño} = b_0 + b_1 X_{Dosis}$$

Modelo de Regresión Lineal Múltiple

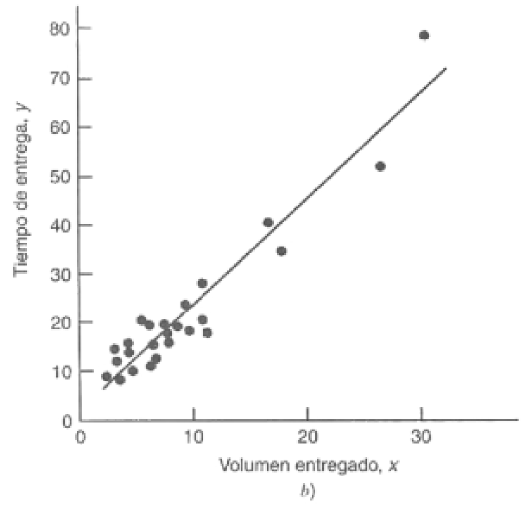
$$Y_{Daño} = b_0 + b_1 X_{Dosis} + b_2 X_{poda}$$

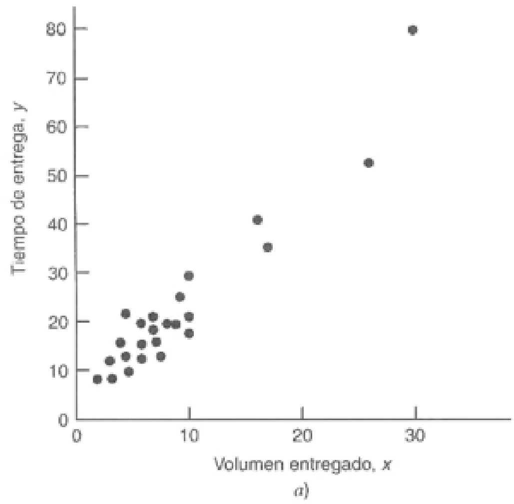
3.4 Regresión Lineal Simple

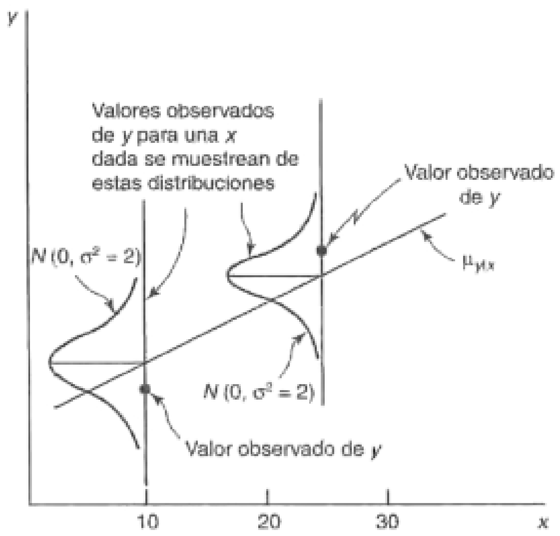
Se centra sobre la dependencia de una variable respuesta respecto a un conjunto de variables regresoras o predictoras. Mediante un modelo de regresión se mide el efecto de cada una de las variables regresoras sobre la respuesta. Uno de los objetivos es la estimación para la predicción del valor medio de la variable dependiente, con base en el conocimiento de las variables independientes o predictoras (Díaz Monroy & Morales Rivera, 2012)

En la práctica a menudo se requiere resolver problemas que implican conjuntos de variables de las cuales se sabe que tienen alguna relación inherente entre sí. Por ejemplo, en una situación industrial quizá se sepa que el contenido de alquitrán en el flujo de salida de un proceso químico está relacionado con la temperatura en la entrada. Podría ser de interés desarrollar un método de pronóstico, es decir, un procedimiento que permita estimar el contenido de alquitrán para varios niveles de temperatura de entrada a partir de información experimental.

Desde luego, es muy probable que para muchos ejemplos concretos en los que la temperatura de entrada sea la misma, por ejemplo 130ºC, el contenido de alquitrán de salida no sea el mismo. Esto es muy similar a lo que ocurre cuando se estudian varios automóviles con un motor del mismo volumen; no todos tienen el mismo rendimiento de combustible. (Walpole Ronald, Myers Raymond, Myers Sharon, 2012)







3.5 Usos de la Regresión

Los modelos de regresión se usan con varios fines, que incluyen los siguientes:

1. Descripción de datos.

2. Estimación de parámetros.

3. Predicción y estimación.

4. Control.

Es común que los ingenieros y los científicos usen ecuaciones para resumir o describir un conjunto de datos. El análisis de regresión es útil para plantear esas ecuaciones. Por ejemplo, se puede reunir una cantidad considerable de tiempo y volumen de entrega, por lo cual un modelo de regresión sería probablemente un resumen mucho más conveniente y útil de esos datos, más que una tabla o una gráfica.

Muchas aplicaciones de regresión requieren de la predicción de la variable de respuesta.
Por ejemplo, se podría tratar de predecir el tiempo en el que se puede entregar una cantidad específica de cajas de refrescos. Estas predicciones pueden ser útiles para planear actividades de entrega, como, por ejemplo, las rutas y los programas, o para evaluar la productividad en las operaciones de entrega.
Ya se han discutido los peligros de extrapolar cuando se usa un modelo de regresión para pronosticar, debidos a errores en el modelo o a la ecuación (véase la Fig. 1.5). Sin embargo, aun cuando sea correcta la forma del modelo, las malas estimaciones de los parámetros de éste pueden seguir causando mal desempeño de la predicción. (Montgomery, Douglas; Peck, Elizabeth; Vining, 2006)

Los modelos de regresión se pueden usar para fines de control. Por ejemplo, un ingeniero químico podría aplicar el análisis de regresión para establecer un modelo que relacionara la resistencia del papel a la tensión con la concentración de fibra corta (es decir, de madera dura) en la pulpa. Esta ecuación se podría usar después para controlar la resistencia dentro de valores adecuados, variando la concentración de fibra corta. Cuando se usa una ecuación de regresión para fines de control, es importante que las variables estén relacionadas en forma causal. Nótese que podría no necesitarse una relación de causa a efecto si sólo se usara la ecuación para predicción. En este caso sólo es necesario que las relaciones que existían en los datos originales con los que se formuló la ecuación de regresión sigan siendo válidas.

Ejemplo utilizando **RStudio**

```
> Agua<- c(8,16,24,32,40,48)

> Rendimiento<- c(4.1,4.5,5.1,6,6.8,7.6)

> Regresion<-lm(Rendimiento~Agua)

> summary(Regresion)
```

Call:

```
lm(formula = Rendimiento ~ Agua)
```

Residuals:

```
      1        2        3        4        5        6
0.22381 -0.09905 -0.22190 -0.04476  0.03238  0.10952
```

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.153333	0.164036	19.22	4.32e-05 ***
Agua	0.090357	0.005265	17.16	6.76e-05 ***

---Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

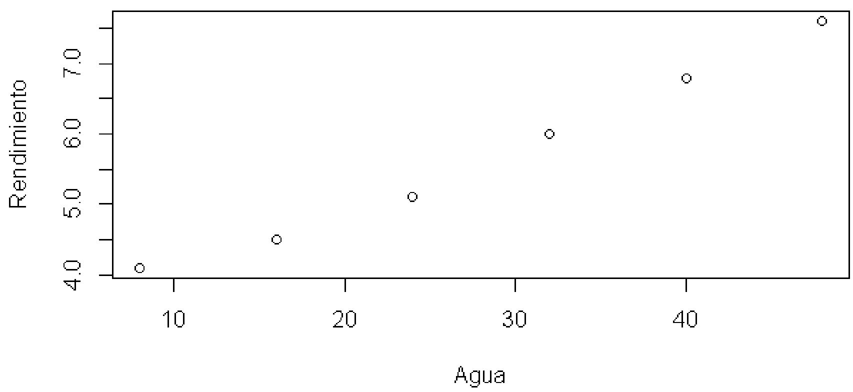
Residual standard error: 0.1762 on 4 degrees of freedom

Multiple R-squared: 0.9866,   Adjusted R-squared: 0.9833

F-statistic: 294.5 on 1 and 4 df,   p-value: 6.763e-05

Utilizando RStudio

```
> data<-data.frame(Agua, Rendimiento)
> cor(data)
                 Agua       Rendimiento
Agua             1.0000000  0.9932777
Rendimiento      0.9932777  1.0000000
> plot(data)
> line(data)
```



Ejemplo de una Regresión Lineal Simple

Se desea probar la efectividad de un nuevo fungicida para el control de roya en trigo. Se probaron distintas dosis en gramos de principio activo por ha (gr.p.a./ha) en 10 parcelas de 100 plantas cada una. A los 15 días de la aplicación se realizó una evaluación del daño, como el tamaño promedio de las manchas en hoja bandera. Los datos son los siguientes:

DOSIS (X)	100	125	200	250	275	300	325	350	375	400
DAÑO (Y)	50	48	39	35	30	25	20	12	10	5

Modelo

$$Y_i = b_0 + b_1 X_i + e_i$$

Formulas:

a) Encuentre los Coeficientes

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

b) Coeficiente de Determinación

$$R^2 = \frac{(S_{xy})^2}{S_{xx} S_{yy}} = \frac{\left[\sum (x_i - \bar{x})(y_i - \bar{y})\right]^2}{\left[\sum (x_i - \bar{x})^2\right]\left[\sum (y_i - \bar{y})^2\right]}$$

$$0 \le R^2 \le 1$$

X	Y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}).(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
100	50	-170	22.6	-3842	28900	510.76
125	48	-145	20.6	-2987	21025	424.36
200	39	-70	11.6	-812	4900	134.56
250	35	-20	7.6	-152	400	57.76
275	30	5	2.6	13	25	6.76
300	25	30	-2.4	-72	900	5.76
325	20	55	-7.4	-407	3025	54.76
350	12	80	-15.4	-1232	6400	237.16
375	10	105	-17.4	-1827	11025	302.76
400	5	130	-22.4	-2912	16900	501.76
$\bar{x}$ =**270**	$\bar{y}$ =**27.4**			S_{xy} -14230	S_{xx} 93500	S_{yy} 2236.4

c) Definir el Modelo final

R Square	0.968386
Bo	68.49198
B1	-0.15219

d) Predecir el daño (tamaño promedio de las manchas) que se hallará si se aplican 260 gr.p.a./ha.

$$Y_{Daño} = 68.491 + (-0.1521)(260)_{Dosis}$$

$$Y_{Daño} = 68.491 - 39.5694$$

$$\boldsymbol{Y_{Daño} = 29.371}$$

3.6 Utilizando RStudio y Statgraphic

Aplicando regresión lineal simple (producción-fertilizantes) Se dispone de los siguientes datos experimentales obtenidos en un campo de cultivo que relacionan la producción con la cantidad de fertilizante aplicado.

Fertilizantes (kg/hect.)	**Producción (kg/hect.)**
x	**y**
100	40
200	45
300	50
400	65
500	70
600	70
700	80

Hallar:

a) La nube de puntos y dibujar una recta que pase “lo más cerca posible” de todos sus puntos y en especial por el centroide.

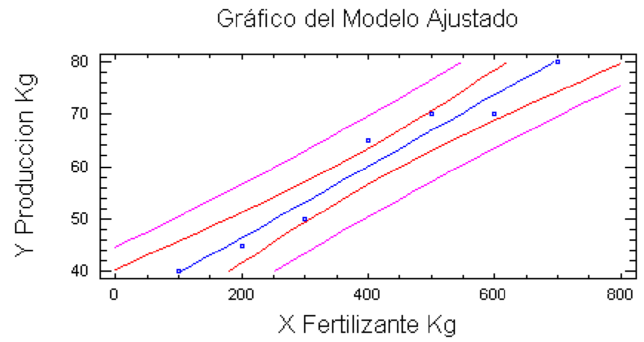
b) La recta de regresión de y sobre x. Interpretar los coeficientes.

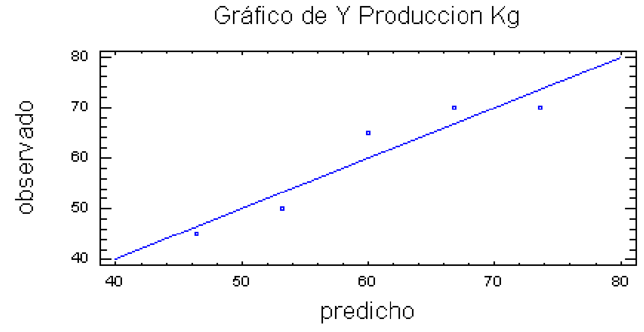
c) La varianza residual.

d) I.C. al 95% para β1.

e) I.C. al 95% para σ2 y para σ.

f) Coeficiente de correlación y coeficiente de determinación. Interpretar los resultados. g) Si se aplican 350kg/hect. de fertilizante, ¿qué producción se obtendrá? ¿Y con 1000 kg/hect.? h) Realiza un contraste de hipótesis con un nivel de significación del 5% para comprobar si la producción depende del fertilizante.





Análisis de Regresión - Modelo Lineal Y = a + b*X

Variable dependiente: Y *Producción Kg*

Variable independiente: X *Fertilizante Kg*

Parámetro	Estimación	Error estándar	Estadístico T	P-Valor
Ordenada	32,8571	2,94508	11,1566	0,0001
Pendiente	0,0678571	0,00658539	10,3042	0,0001

Análisis de la Varianza

Fuente	Suma de cuadrados	GL	Cuadrado medio	Cociente-F	P-Valor
Modelo	1289,29	1	1289,29	106,18	0,0001
Residuo	60,7143	5	12,1429		
Total (Corr.)	1350,0	6			

Coeficiente de Correlación = 0,977255

R-cuadrado = 95,5026 porcentaje

R-cuadrado (ajustado para g.l.) = 94,6032 porcentaje

Error estándar de est. = 3,48466

Error absoluto medio = 0,0979391

Estadístico de Durbin-Watson = 1,32289 (P=0,0287)

Autocorrelación residual en Lag 1 = 0,174299

El StatAdvisor

La salida muestra los resultados del ajuste al modelo lineal para describir la relación entre Y Produccion Kg y X Fertilizante Kg. La ecuación del modelo ajustado es

Y Producción Kg = 32,8571 + 0,0678571*X Fertilizante Kg

Dado que el p-valor en la tabla ANOVA es inferior a 0.01, existe relación estadísticamente significativa entre Y Producción Kg y X Fertilizante Kg para un nivel de confianza del 99%.

El estadístico R-cuadrado indica que el modelo explica un 95,5026% de la variabilidad en Y producción Kg. El coeficiente de correlación es igual a 0,977255, indicando una relación relativamente fuerte entre las variables. El error estándar de la estimación muestra la desviación típica de los residuos que es 3,48466. Este valor puede usarse para construir límites de la predicción para las nuevas observaciones seleccionando la opción Predicciones del menú del texto.

El error absoluto medio (MAE) de 0,0979391 es el valor medio de los residuos. El estadístico Durbin-Watson (DW) examina los residuos para determinar si hay alguna correlación significativa basada en el orden en el que se han introducido los datos en el fichero. Dado que el p-valor es inferior a 0.05, hay indicio de una posible correlación serial. Represente los residuos frente al orden de fila para ver si hay algún modelo que pueda verse.

3.7 Ejercicios complementarios capítulo 3

1. Para estudiar el efecto de la temperatura sobre el vigor durante la germinación, se dispusieron semillas de alfalfa en germinadores a distintas temperaturas. A los 6 días se midió la longitud de las plántulas (mm), obteniéndose los siguientes datos:

Temperatura ºC (X)	12	12	12	18	18	18	24	24	24	28	28	28
Longitud de Plántulas (mm) (Y)	13	18	15	20	24	15	22	27	31	24	25	28

Formulas:

a) Coeficientes

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

b) Coeficiente de Determinación

$$R^2 = \frac{(S_{xy})^2}{S_{xx}S_{yy}} = \frac{\left[\sum(x_i - \bar{x})(y_i - \bar{y})\right]^2}{\left[\sum(x_i - \bar{x})^2\right]\left[\sum(y_i - \bar{y})^2\right]}$$

$$0 \leq R^2 \leq 1$$

c) Covarianza

$$\text{cov}(X,Y) = \frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}S_{xy}$$

$$\rho(X,Y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

$$-1 \leq \rho \leq 1$$

d) Predecir la longitud (mm) que se hallará si la temperatura alcanza 30 ºC

$$Y_{long} = (\quad) + (\quad)X_{\text{ºC}}$$

2. Los siguientes datos proporcionan la recuperación de bromuro adicionado a muestras con contenido vegetal, medido mediante un método cromatográfico gas-liquido. La cantidad de bromuro potásico añadido a cada tipo de vegetal fue la misma.

Tomate	777	790	759	790	770	758	764	768	762	µg g^{-1}
Pepino	782	773	778	765	789	797	782	792	793	µg g^{-1}

Fuente: (Roughan, J. A., Roughan, P. A. and Wilkins, J. P. G. 1983 *Analyst* 108:742)

a) Coeficientes
b) Correlación y Covarianza
c) R^2
d) Predecir la cantidad de bromuro en tomate que se hallará si la cantidad de bromuro en pepino es 795 µg g^{-1}

$$Y_{tomate} = (\quad) + (\quad)X_{\text{pepino}}$$

CAPITULO 4

4.1 MUESTREO

4.1.1 Población

Está compuesta de individuos como, personas, animales, insectos, frutas, semillas, arboles, etc. Elementos que guardan características similares y medibles.

4.1.2 Censo

Un **censo** es una tentativa de investigación que trata de incluir los datos relacionados con todo miembro de la población meta definida.

4.1.3 Muestra

Una **muestra** es una tentativa de investigación que busca elaborar juicios sobre un grupo más grande, mediante la comunicación con un grupo más pequeño de elementos extraído de la población meta total.

* **Muestreo. - S**e basa más en el método científico que en la intuición.

El muestreo es eficaz y permite que el equipo de investigación proyecte los resultados procedentes de un grupo pequeño hacia una población meta más grande, con lo que se ahorra tiempo y dinero.

* **Unidad Muestral. -** Es todo elemento o entidad de la muestra.

* **Tamaño Muestral. -** Es el número de elementos de la población que conforman la muestra y se denota con **n**.

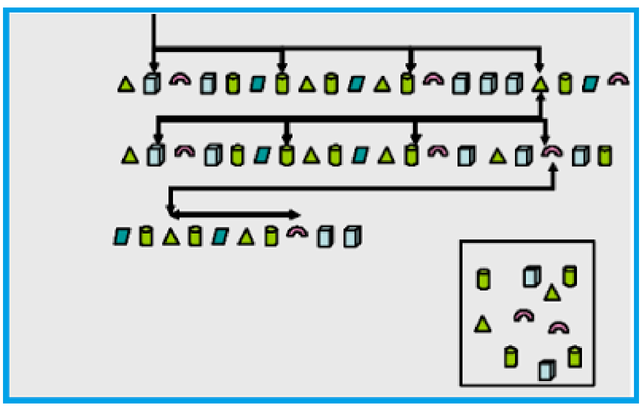
4.2 Muestreo Probabilístico

4.2.1 Muestreo Aleatorio Simple (MAS)

Los investigadores utilizan una tabla de números aleatorios, u otro procedimiento de selección aleatoria que garantice que cada unidad muestral que integra la población de tamaño N definida tenga una oportunidad, igual, y diferente de cero, de ser elegida en la muestra.

4.2.2 Muestreo Aleatorio Sistemático (MASI)

Siriviéndose de alguna forma de lista ordenada de los miembros de la población meta definida, los investigadores seleccionan un punto de partida aleatorio para el primer miembro muestreado. Tras determinar cuál tiene que ser el valor de “intervalo de salto” constante para garantizar la representatividad, este intervalo se aplica para elegir a cada enésimo miembro a partir del punto de partida aleatorio hasta completar la muestra necesaria.



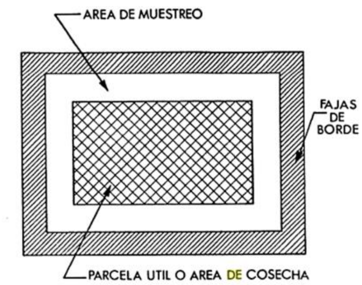
Ejemplo:

Un experimento consistió en contar el número de flores por planta de una POBLACION

N = 50 plantas. Los valores resultantes del conteo fueron los siguientes

μ=5.86

10	8	6	3	9	7	5	4	6	9
8	10	7	9	10	6	8	6	3	2
4	3	2	7	5	5	4	3	7	6
6	7	8	8	6	7	7	9	8	6
5	3	2	1	4	3	6	8	7	0



a) Tome una muestra de tamaño n= 10 a través del MAS
b) Tome una muestra de tamaño n= 10 a través del M. Sistemático

$$error = e = |\bar{x} - \mu|$$

μ: media poblacional
e: error experimental

Ejercicio:

Se tiene dos variedades de Mango donde su población es de tamaño **N=170** cada uno, la variable que se medio fue **Peso en Gramos.**

TOMMY ATKINS (grs)

539	574	561	458	528	536	570	544	475	522
452	508	517	599	581	475	469	539	546	529
484	465	576	529	463	483	596	470	537	505
575	477	460	582	545	596	521	500	588	500
451	585	493	559	451	495	558	591	535	460
549	557	549	590	574	598	511	503	572	580
589	513	589	450	600	463	472	505	503	488
458	548	486	540	561	573	482	555	470	492
467	588	581	502	577	550	470	577	458	567
560	547	483	580	558	593	479	524	575	482
555	454	588	525	525	454	473	597	573	593
593	493	505	513	589	549	507	535	587	535
511	491	520	556	466	595	519	573	464	503
491	557	468	482	539	457	523	526	584	533
532	554	462	497	454	562	538	468	584	523
554	597	459	487	477	526	552	504	450	568
495	590	592	486	512	594	521	500	576	462

KENT (grs)

531	532	503	541	544	497	513	542	474	481
471	547	503	480	507	501	495	484	474	510
537	534	537	488	514	532	516	526	494	509
545	497	549	536	486	494	518	525	532	525
486	518	528	475	524	520	478	536	533	509
480	513	546	496	549	539	479	517	510	541
526	519	515	489	512	474	471	510	537	487
497	532	486	476	517	544	516	507	491	525
548	480	509	546	538	528	503	499	516	504
528	530	483	509	502	496	549	550	502	520
529	549	539	517	536	486	550	497	478	535
506	532	537	501	494	524	493	519	489	514
538	514	550	532	541	475	505	526	528	550
507	538	486	532	513	544	537	536	498	491
475	540	508	488	479	474	522	486	531	535
494	487	549	542	539	481	510	483	537	527
525	537	505	498	531	504	487	525	521	476

a) Calcule la media poblacional tanto para las variedades TOMMY y KENT

b) Un Ingeniero Agrónomo desea tomar una muestra para medir la media del peso de los frutos si sabe que tiene una desviación típica de 25 gr, con un **95%** de confianza, ¿Cuál sería el tamaño de muestra para un error de 5 gr en la **Variedad TOMMY?**

c) Tome una muestra de acuerdo al tamaño que obtuvo en el literal **b)** utilice M.A.S y obtenga la media muestral de la variedad TOMMY.

d) Un Ingeniero Agrónomo desea tomar una muestra para medir la media del peso de los frutos si sabe que tiene una desviación típica de 20 gr, con un **99%** de confianza, ¿Cuál sería el tamaño de muestra para un error de 5 gr en la **Variedad KENT**?

e) Tome una muestra de acuerdo al tamaño que obtuvo en el literal **d)** y utilice Muestreo Sistemático y obtenga la media muestral de la variedad KENT.

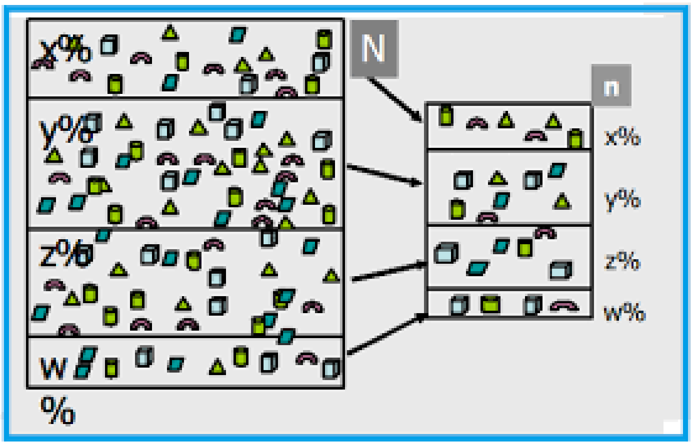
f) Compare las medias muéstrales de las **Variedades,** saque la diferencia entre la media poblacional y las dos medias muéstrales de las dos Variedades.

e=|$\bar{x}$ - μ| para MAS y

e=|$\bar{x}$ - μ| para Muestreo Sistemático.

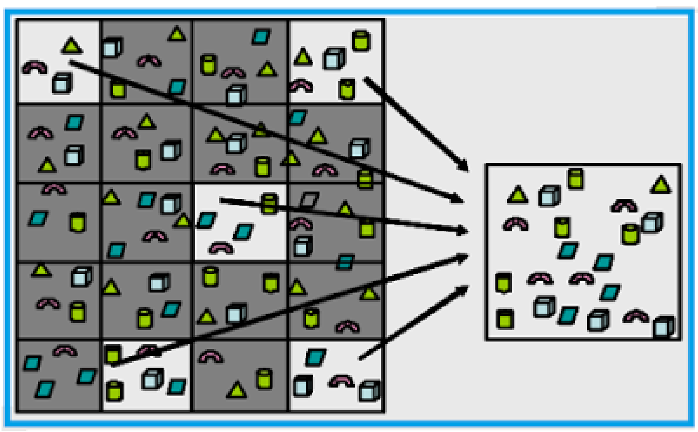
4.2.3 Muestreo Aleatorio Estratificado (MAE)

Cuando se cree que la población meta definida tiene una distribución anormal (o bifurcada) para una o más de sus características distintivas (por ejemplo, edad, ingresos, propiedad de productos), los investigadores tienen que identificar subpoblaciones, a las que llama *estratos o segmentos*. Después de que se han frecuentado los estratos, se elabora una muestra aleatoria sencilla por cada estrato. Se aplican factores de ponderación proporcional y desproporcional para estimar los valores de población totales.



4.2.4 Muestreo de Conglomerados o por Áreas

Este método requiere que la población meta definida se segmente en áreas geográficas, cada una de las cuales se considera muy similar a las otras. Los investigadores seleccionan al azar algunas áreas y luego levantan un censo de los elementos de cada área; como alternativa pueden elegir más áreas y tomar muestras de cada una. Este método muestral es atractivo cuando los investigadores pueden identificar fácilmente las áreas que son muy semejantes.



4.2.5 Muestreo de Captura y Recaptura

En biología, conocer el número de individuos de una población determinada es de vital importancia; sin embargo, en muchos de los seres vivos, es imposible tener un conteo de su población, ya sea por su elevado número o su movilidad. Es por esto que se tiene que recurrir a técnicas de estimación de la población. Uno de los modelos con mayor importancia es el de Captura-Recaptura. Peterson fue el primero en utilizar esta técnica en 1896 para estudiar la migración de peces y luego para estimar el tamaño de la población y su tasa de mortalidad. (Rienzo, Alejandro, Alicia, Margot, & Pilar, 2008)

Utilizando el método más simple (Petersen) las dos muestras son independientes, los animales marcados en la primera captura, se vuelven a mezclar en el hábitat (misma población), de tal forma que el hecho de ser seleccionados (marcados) no está relacionado con la probabilidad de ser seleccionado en la segunda muestra, la probabilidad de ser atrapado en la segunda muestra no depende de su historia de captura.

La marca o señal debe ser lo suficientemente resistente para soportar el tiempo entre la captura y recaptura.

Procedimiento (modelo Petersen)

1) Obtener una muestra aleatoria de n_1 individuos los cuales son marcados.

2) Regresar los individuos marcados al medio para que se mezclen con los no marcados.

3) Capturar una nueva muestra aleatoria de tamaño n_2 y contar las recapturas.

4) La proporción recapturada nos indicará el tamaño de la población total.

$$Por\, proporcionalidad:$$

$$\frac{n_1}{N} = \frac{m}{n_2}$$

$$por\, lo\, que:$$

$$\tilde{N} = \frac{n_2 n_1}{m}$$

N = Tamaño de la población
n_1 = Tamaño de la primera muestra y total de elementos marcados en la población
n_2 = Tamaño de la segunda muestra
m = Individuos marcados en la segunda muestra
$m \leq n_2$

Ejemplo.

En un lago con tilapias (***Oreochromis niloticus***) se desea estimar el total de tilapias que se encuentran, para esto se toma una red y se captura una muestra aleatoria de n_1 48 se los marca y se los vuelve a devolver al lago después se toma una muestra aleatoria de n_2 50 peces donde se cuenta los peces marcados. ¿Calcule el total de peces estimado en el lago?

$$\tilde{N} = \frac{n_1 n_2}{m}$$

$$\tilde{N} = \frac{(48)(50)}{23}$$

$$\tilde{N} = 104.34 \; \approx 104\, peces$$

4.2.6 Muestreo de Suelos

En la toma de muestra de suelos hay tipos y cantidad de muestras a tomar

- **Muestra simple:** Es la que se obtiene con una sola extracción de suelo. Son usadas en trabajos de investigación y en suelos muy homogéneos. Sé recomienda cuatro muestras por hectárea, de 1 kilogramo de suelo cada una.
- **Muestra compuesta:** Se refiere a la muestra de suelo obtenida por la extracción de varias muestras simples o submuestras, reunidas en un recipiente y bien mezcladas, de donde se retiran de 0,5 a 1 kg de suelo. Son las más usadas para la planificación de la fertilización. Se recomienda 15-20 submuestras por parcela de muestreo.

4.2.6.1 Localización y profundidad de muestreo

- Características de los muestreos en diferentes cultivos.
- Para cultivos anuales, retirar las muestras de los surcos a una profundidad de 20 cm. Si el sistema es de siembra directa, se recomienda muestrear a 2 profundidades, de 0 a 10 y de 10 a 20 cm. Para cultivos perennes, realizar el muestreo en la zona de fertilización.

4.2.6.2 Sitios de Muestreo

- El muestreo de suelos se deberá realizar al azar y en las siguientes formas.

4.3 Determinación del Tamaño Muestral

Depende de los parámetros que se desean estimar Ej: media, diferencia de medias

4.3.1 Fórmulas para calcular el Tamaño de muestra para la media µ

Tamaño de muestra

Cuando no se conoce la población

(población infinita)

$$n = \frac{Z^2 \ \sigma^2}{e^2}$$

Tamaño de muestra

Cuando se conoce la población

(población finita)

$$n = \frac{Z^2 \ \sigma^2 \ N}{e^2(N-1) + Z^2 \ \sigma^2}$$

Para Estimar µ

z = nivel de confianza elegido
σ= desviación estándar
e= error máximo
N= tamaño de la población

Ejemplos

- N= 120 árboles (Universo)
- Zα/2=1.96 con (1-α) *100% de Confianza
- Donde α=0,05; **95% de confianza**
- e=± 5 cm; aproximadamente no más de 5 cm de error

Pob. Infinita

$$n_0 = \frac{Z_{\alpha/2}{}^2 \sigma^2}{e^2} = \frac{(1.96)^2 (259.85)}{5^2} = 39.92 \cong 40$$

Pob. Finita

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{39.92}{1 + \frac{39.92}{120}} = 29.9 \cong 30$$

90% confianza= 1.64
95% confianza= 1.96
99% confianza= 2.58

Un médico quiere estimar el peso promedio de los recién nacidos en cierto hospital. Un estudio anterior de diez niños mostro que la desviación estándar de sus pesos es de 150 gr. ¿Qué tan grande debe ser una muestra para que el medico tenga el 95% de confianza de que el error de estimación es a lo mas de 40gr?

$$Z(1.96)\ 1-\alpha = 95\%$$

$$\alpha = 5\%$$

$e = 40$
σ= 150
Z= 1.96

$n = \frac{(1.96)^2\ (150)^2}{(40)^2} = 54.02 \cong 54$ niños

Siempre será 1.96 cuando es 0.95

Para el ejemplo anterior. ¿Cuánto sería el tamaño de muestra si se conoce que el total de recién nacidos en cierto mes es de 200 niños?

$$n = \frac{Z^2\ \ \sigma^2\ N}{e^2(N-1) + Z^2\ \ \sigma^2}$$

$$n = \frac{(1.96)^2\ \ (150)^2\ (200)}{(40)^2(200-1) + (1.96)^2\ \ (150)^2} = 42.7$$

Serian aproximadamente 43 niños en la muestra.

4.3.2 Fórmula para calcular el tamaño de muestra para estimar *P*

Tamaño de muestra cuando no se conoce la población

(población infinita)

$$n = \frac{Z^2\ \ p\ (1-p)}{e^2}$$

Tamaño de la muestra cuando se conoce la población

(población finita)

$$n = \frac{Z^2\ \ p(1-p)\ N}{e^2(N-1) + Z^2\ \ p\ (1-p\ )}$$

Para estimar Proporción

Z=nivel de confianza elegido
P=proporción de una categoría de la variable
e=error máximo

¿A cuántas familias tendríamos que estudiar para conocer la preferencia del mercado en cuanto a las marcar de shampoo para bebé, si se desconoce la población total? Asumamos los siguientes datos, nivel de confianza 95%, una precisión (error muestral) del 3% y la proporción esperada es del 5%.

Z=1.96
P=0.05
e=0.03

$$n = \frac{(1.96)^2\ \ (0.05)\ (1-0.05)}{(0.03)^2} = 202.75 \cong 203\ familias$$

¿Cómo hubiera cambiado el ejemplo anterior, si se desconoce la proporción esperada?

Se supone un 50% de proporción

Z=1.96
P=0.50
e=0.03

$$n = \frac{(1.96)^2 \ \ (0.50)(1-0.50)}{(0.03)^2} \ = 1067.11 \cong 1067 \ familias$$

Un investigador está interesado en estimar la proporción de muertes debido a cáncer de estómago en relación con el número de defunciones por cualquier tipo de neoplasia. Su experiencia le indica que sería sorprendente que tal proporción supere el valor de 1/3. ¿Qué tamaño de muestra debe tomar para estimar la anterior proporción en un nivel de confianza del 99%, para que el valor estimado no difiera del valor real en más de 0.03?

Z=2.575
P=1/3
e=0.03

$$n = \frac{Z^2 \ \ p\,(1-p)}{e^2}$$

$$n = \frac{(2.575)^2 \ \ (1/3)\,(1-1/3)}{(0.03)^2} = 1637{,}19 \cong 1637 \ \text{ personas}$$

4.3.3 Tamaños de Muestra para Encuestas

Para el cálculo del tamaño de la muestra utilizaremos un p=0.5 como incertidumbre por lo general es utilizado. Aquí un ejemplo

En una muestra aleatoria de 500 familias en el distrito de los Ceibos se encuentra que 340 familias están suscritas a seguros Confianza. ¿Qué tan grande se requiere que sea una muestra si se quiere tener 95% de confianza, que la estimación de *p* esté dentro de 0.06?

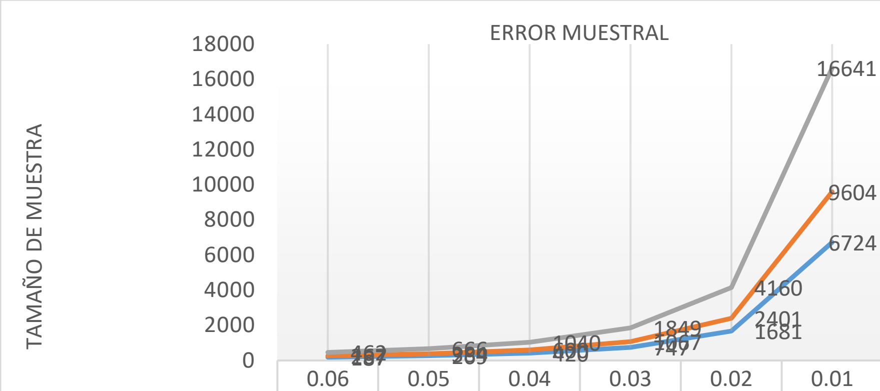
Z=1.96
P=340/500=0.68
e=0.06

$$n = \frac{Z^2 \ \ p\,(1-p)}{e^2}$$

$$n = \frac{(1.96)^2 \ \ (0.68)\,(1-0.68)}{(0.06)^2} = 232.20 \cong 232$$

Serian 232 familias en el distrito ceibos a encuestar.

A continuación, se muestra un gráfico con diferentes niveles de errores y confianza para encuestas.



	0.06	0.05	0.04	0.03	0.02	0.01
90% de Confianza	187	269	420	747	1681	6724
95% de Confianza	267	384	600	1067	2401	9604
99% de Confianza	462	666	1040	1849	4160	16641

CAPITULO 5

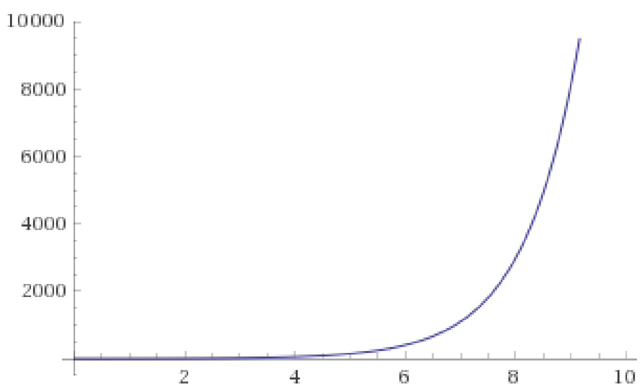
FUNCION EXPONENCIAL: CRECIMIENTO Y DECRECIMIENTO

5.1 Función Exponencial

La función f(x)= e^x (donde e es el número irracional e≈ 2.718281) se llama función exponencial y está definida por todos los números reales R.

Para todos los números x y y se cumple:

1. **$e^0=1$**
2. **$e^{x+y}= e^x e^y$**
3. **$(e^x)^y= e^{xy}$**
4. **$e^{-x} = \frac{1}{e^x}$**
5. **La función f(x) es derivable y f'(x) =e^x**
6. $\lim_{x\to+\infty} e^x = \infty$ $\qquad$ $\lim_{x\to-\infty} e^x = 0$



5.2 Función Logaritmo

Como f(x)= e^x es una función continua y creciente (la derivada f'(x) = e^x toma valores positivos) la inversa existe y la llamaremos función logaritmo, g(x) =ln(x). El dominio de la función logaritmo es el conjunto de todos los números reales positivos.

Si: e^x =y ←→ x = lny

Y tenemos: $e^{\ln y}$= y $\qquad$ ln e^x= x

La función logaritmo tiene las siguientes propiedades, si x e y son mayores a cero se cumple que:

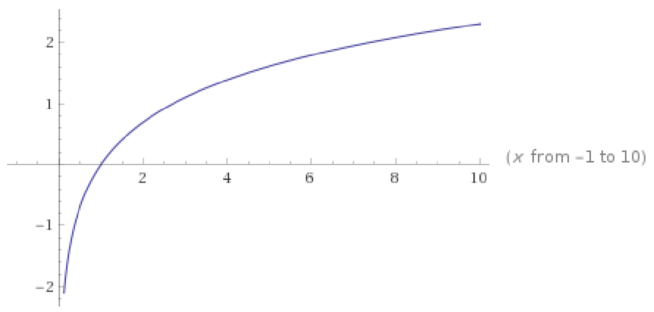
1. ln(xy) = lnx +lny
2. Si m y n son enteros positivos, entonces:
    a. $\ln(x^{-1})$=-lnx
    b. $\ln(x^m)$=m lnx

c. $\ln(\sqrt[n]{x}) = \frac{1}{n} lnx$

3. si g(x) = lnx entonces g'(x)= $\frac{1}{x}$

4. $\lim_{x\to+\infty} lnx = +\infty$ $\lim_{x\to 0^+} lnx = -\infty$

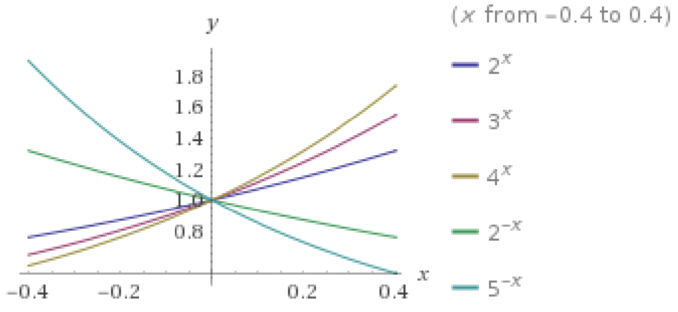
Representación gráfica de ln(X)



5.2.1 Propiedades del Logaritmo

Usamos la definición del logaritmo común: b^a=x si y solo si *log*b(x) = a

Usando la regla de cambio de base: *log*b(x)= $\frac{\log_a x}{\log_a b}$

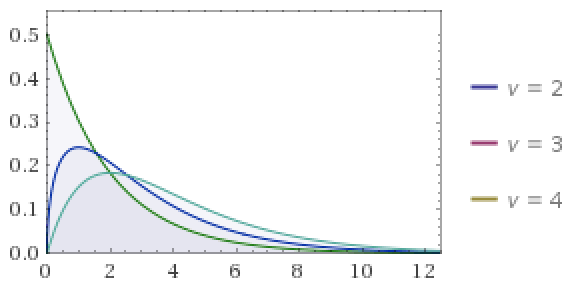


5.3 Chi Cuadrado

En 1900 Karl Pearson propuso el siguiente estadístico de prueba, que es una función de los cuadrados de las desviaciones de los números observados con respecto a sus valores esperados, ponderados por el reciproco de sus valores esperados:

$$X^2 = \sum_{i=1}^{k} \frac{[n_i - E(n_i)]^2}{E(n_i)}$$

Usualmente esta prueba es usada para datos enumerativos o de conteo como por ejemplo observaciones físicas o químicas que no se pueden medir en una escala continua y por tanto producen datos enumerativos o de clasificación. (Mendenhall, William III; Wackerly, Dennis; Scheaffer, 2009)



plot chi square | Computed by Wolfram|Alpha

Donde v son los grados de libertad

Se puede utilizar también el estadístico X^2 para verificar si los datos de la muestra indican que un modelo particular para una distribución de la población no ajuste a los datos. Un ejemplo de tal prueba, llamada la prueba de bondad de ajuste, se presenta en el siguiente ejemplo.

Ejemplo

Los datos siguientes muestran las frecuencias de conteo para 400 observaciones acerca del número de colonias bacterianas por campo en un microscopio, utilizando muestras de una capa delgada de leche. Pruebe la hipotesis de que los datos provienen de una distribución de Poisson. (Utilice α=5%)

Numero de colonias por campo	Frecuencia de observación
0	56
1	104
2	80
3	62
4	42
5	27
6	9
7	9
8	5
9	3
10	2
11	0
19	1
	400

Fuente: C. I. Bliss y R. A. Fisher, "Fitting the Negative Binomial Distribution to Biological Data," *Biometrics*, vol 9 (1953), pp. 176-200.

Utilizando Minitab v17

Goodness-of-Fit Test for Poisson Distribution

```
Data column: Número de Colonias
Frequency column: Frecuencia

Poisson mean for Número de Colonias = 2.44
```

Número de Colonias	Observed	Poisson Probability	Expected	Contribution to Chi-Sq
0	56	0.087161	34.864	12.8130
1	104	0.212672	85.069	4.2129
2	80	0.259460	103.784	5.4506
3	62	0.211028	84.411	5.9501
4	42	0.128727	51.491	1.7493
5	27	0.062819	25.128	0.1395
6	9	0.025546	10.219	0.1453
>=7	20	0.012586	5.035	44.4853

N	N*	DF	Chi-Square	P-Value
400	0	6	74.9460	0.000

$$X^2 = \sum_{i=1}^{7} \frac{[n_i - E(n_i)]^2}{E(n_i)}$$

Hipotesis

Ho: Los datos poseen una distribución de Poisson

vs

H_1: Los datos no poseen una distribución de Poisson

Los grados de libertad fueron k-1=6 donde se sigue la regla que si $X^2 > X^2tab$ se rechaza la hipotesis nula, 74.946 > 12.592

Ya que $X^2_{0.05}$ = 12.592, con 6 grados de libertad podemos rechazar la Hipotesis nula Ho, es decir, el conteo de observaciones por colonias de bacterias por campo sigue una distribución de Poisson.

5.4 Crecimiento y Decrecimiento de Bacterias, Virus e Insectos.

Para este tipo de poblaciones se utiliza la función exponencial para poder estimar la dinámica de estas poblaciones como en este ejemplo donde el Análisis de experimentos con fertilizantes, se suelen interpretar esos ensayos por la ley de MItscherlich:

$$\boldsymbol{y = A(1 - 10^{-c(x+b)})}$$

Donde y es la producción, x es la dosis del nutriente, A es la producción máxima teórica posible cuando aumenta indefinidamente la dosis de un nutriente, c es el llamado coeficiente de eficacia (es un parámetro típico del nutriente en cuestión) y b es el tenor de ese nutriente contenido en el suelo en forma asimilable para las plantas.

A continuación, tenemos una población de una especie sigue la siguiente función:

$$\boldsymbol{N(t) = a + \frac{t}{e^{t/2}}}\,, \qquad \boldsymbol{t \geq 0}$$

Donde N(t) es el número de individuos en la población (medida en miles) y el tiempo (medido en meses) y a una constante positiva.

a) Calcular a sabiendo que inicialmente había 3000 individuos
b) Grafique N(t) de t=0 hasta t=13 (meses)
c) Si se sabe que una población está en peligro de extinción cuando el número de individuos es menor que 1000. ¿Tiene esta población peligro de extinción?

Dentro de dos meses, la población de una colonia de insectos en un área remota alcanzara $\boldsymbol{3.2 * 10^4}$. Si la población de la colonia se duplica cada dos meses, ¿Cuál era la población hace ocho meses?

a) $3.6 * 10^2$
b) $1.0 * 10^3$
c) $2.0 * 10^3$
d) $1.6 * 10^4$
e) $2.6 * 10^4$

Se nos dice que una población se duplica cada 2 meses en el futuro, tendríamos:

Hace 8 meses: X insectos

Hace 6 meses: 2X insectos

Hace 4 meses: 4X insectos

Hace 2 meses: 8X insectos

Ahora: insectos X: 16X

2 meses en el futuro: 32 X insectos

Entonces, en ese periodo de tiempo, la cantidad de insectos se convierte en 32 veces más de lo que comenzó. Ahora podemos dividir $3.6 * 10^2$ para 32 para determinar cuál era la población hace 8 meses.

$$\frac{3.2 * 10^4}{32} = 0.1 * 10^4 = 1 * 10^3$$

El decrecimiento de un cultivo de bacterias después de aplicar una dosis de antibióticos es tal que a cada hora disminuye (4/5) del número de las mismas. En estas condiciones había 600 bacterias al iniciar el experimento.

a) ¿Cuántas bacterias habrá en el cultivo cuando transcurra 600 minutos?

600 minutos son 10 horas
P_0 es igual a 600 bacterias

$$P(t) = Po * a^{t/tr}$$

$$P(t) = 600 * \left(\frac{4}{5}\right)^{10/1}$$

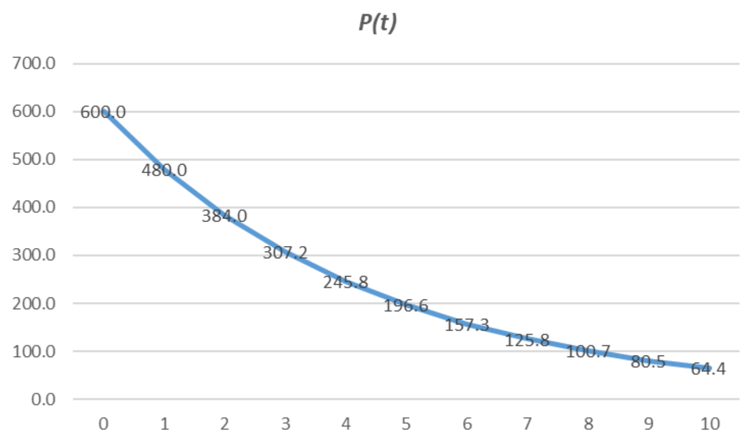
$$P(t) = 600 * 0.1074$$

$$P(t) = 64.42 \approx 64$$

Al transcurrir 600 minutos es decir 10 horas el decrecimiento de este tipo de bacteria después de la aplicación de una dosis de antibiótico se reduce su población a 64 bacterias.

b) Esbocé un gráfico para P(t) desde t=0 hasta t=10 (horas)

t	*P(t)*
0	600.0
1	480.0
2	384.0
3	307.2
4	245.8
5	196.6
6	157.3
7	125.8
8	100.7
9	80.5
10	64.4



La población de una colonia de bacterias alcanzara $\mathbf{1.28 * 10^6}$. Si la población de la colonia se duplica cada cuatro horas, ¿Cuál era la población hace doce horas?

$$P(t) = Po * a^{t/tr}$$

$$\mathbf{1.28 * 10^6} = Po * 2^{12/4}$$

$$\mathbf{1.28 * 10^6} = Po * 2^3$$

$$Po = \frac{\mathbf{1.28 * 10^6}}{8}$$

$$Po = 16000 = \mathbf{1.6 * 10^5}$$

Se nos dice que una población se duplica cada 4 horas en el futuro, tendríamos:

Hace 12 horas: X bacterias

Hace 8 horas: 2X bacterias

Hace 4 horas: 4X bacterias

Ahora: X: 8X bacterias

Entonces, en ese periodo de tiempo, la cantidad de bacterias se convierte en 8 veces más de lo que comenzó. Ahora podemos dividir $1.28 * 10^6$ para 8 para determinar cuál era la población hace 12 horas.

$$\frac{1.28 * 10^6}{8} = 1.6 * 10^5$$

Cada año durante cuatro años, una especie de insecto aumento su población dentro de una localidad en un número igual a la mitad de la población del año anterior. Si había 16.200 insectos de una especie en la localidad final del periodo de cuatro años, ¿Cuantos insectos de la especie se encontraban en la localidad al comienzo del periodo de cuatro años?

$$P(t) = Po * a^t$$

$$16200 = Po * \left(\frac{3}{2}\right)^4$$

$$16200 = Po * \left(\frac{81}{16}\right)$$

$$Po = \left(\frac{16}{81}\right) \quad * 16200$$

$$Po = 3200$$

Una población de una colonia de bacterias aumenta en un 20% cada 3 minutos. Si a las 9:00 am la colonia tenía una población de 144,000, ¿Cuál era la población de la colonia a las 8:54 am?

a) 100,000
b) 112,000
c) 120,000
d) 121,000
e) 136,000

$$P(t) = Po * (1 + i)^{t/tr}$$

$$144{,}000 = Po * 1.2^{6/3}$$

$$144{,}000 = Po * 1.2^{2}$$

$$144{,}000 = Po * 1.44$$

$$\frac{144{,}000}{1.44} = Po \;\rightarrow Po = 100{,}000$$

La población de un determinado pueblo aumenta en 50% cada 50 años. Si la población en 1950 era de 810, ¿En qué año era la población 160? Con 2 decimales.

a) 1651
b) 1709
c) 1738
d) 1800
e) 1912

$$P(t) = Po * (1 + i)^{t/tr}$$

$$810 = 160 * 1.5^{t/50}$$

$$\frac{810}{160} = \sqrt[50]{1.5^{t}}$$

$$5.06 = 1.5^{t/50}$$

Usamos la definición del logaritmo común: b^a=x si y solo si log_b(x) = a

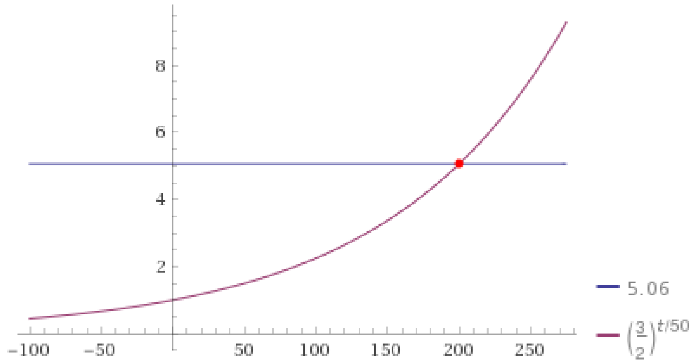
$$\log_{1.5} 5.06 = \frac{t}{50}$$

Usando la regla de cambio de base: log_b(x)= $\frac{\log_a x}{\log_a b}$

$$\frac{\log 5.06}{\log 1.5} = \frac{t}{50}$$

$$\frac{\log 5.06}{\log 1.5} * 50 = t$$

$4.24 * 50 = t$ ➔ $t = 212$ años atrás es decir 1738.



5.5 Ejercicios complementarios del Capítulo 5

1. El crecimiento de un cultivo de bacterias es tal que a cada hora se duplica el número de las mismas. En estas condiciones había 600 bacterias al iniciar el experimento. ¿Cuántas bacterias habrá en el cultivo cuando transcurra 720 minutos?

$$N(t) = N_0 \,.\, a^{t/tr}$$

2. Ninoska está estudiando el crecimiento de una población de insectos. Durante la primera semana hay 400 insectos, la segunda semana hay 1200 y las semanas siguientes se sigue triplicando la población. ¿Cuántos insectos habrá para la quinta semana?

$$N(t) = N_0 \,.\, a^{t/tr}$$

3. El decrecimiento de un cultivo de bacterias después de aplicar una dosis de antibióticos es tal que a cada hora disminuye (3/6) del número de las mismas. En estas condiciones había 680 bacterias al iniciar el experimento.

a) ¿Cuántas bacterias habrá en el cultivo cuando transcurra 240 minutos?

b) Esbocé un gráfico para P(t) desde t=0 hasta t=8 (horas)

4. Un estudiante realiza un experimento en laboratorio de biología y descubre que la proporción del número de insectos en una población dada que tiene la característica *x* con respecto al número de insectos en la población que no tiene la característica *x* es 5:3, y que 3/8 de los insectos que tienen la característica x son los insectos machos. ¿Qué proporción de la población total de insectos son insectos machos que tienen la característica *x*?

CAPITULO 6

6.1 PRUEBA DE HIPOTESIS Y ESTADISTICA NO PARAMETRICA

6.1.2 Prueba de Hipotesis

La hipotesis es una aseveración o conjetura sobre un conjunto de datos. La prueba de Hipotesis sirve para la formación de un procedimiento de decisión que se base en los datos y que pueda producir una conclusión acerca de algún sistema científico.

Aceptación o rechazo de hipotesis estadísticas:

Son complementarias

Hipotesis nula (Ho): cualquier hipotesis que desea probar

Hipotesis alternativa (H_1): el rechazo de Ho conduce la hipotesis alternativa

Se decide si es cierto o no

Hipotesis nula= generalmente representa la pregunta que se responderá o la teoría que se probará.

Hipotesis nula Ho anula o se opone a Hi y a menudo es el complemento lógico de Hi.

→ rechazar Ho a favor de Hi debido a evidencia suficiente en los datos.

Ejemplos

H_0: el tamaño promedio de los cangrejos extraídos cumple la normativa

H_1: el tamaño promedio de los cangrejos extraídos no cumple la normativa

H_0: existe una reducción del contaminante luego de la aplicación del tratamiento

H_1: no existe una reducción del contaminante luego de la aplicación del tratamiento

Nivel de significancia: la probabilidad (α) más alta de rechazar Ho cuando es cierto se llama nivel de significancia.

Región Critica o de Rechazo: parte de la curva *z* o de la curva *t-student* donde se rechaza Ho.

Estadístico de Prueba: un solo número calculado a partir de los datos muestrales.

Valor *p*: probabilidad calculada usando la prueba estadística.

Tipos de Pruebas de Hipotesis:

1. de dos colas o bilaterales
2. de una cola o unilateral

1.Para el de una cola a la derecha

El investigador desea comprobar la hipotesis de un valor mayor en el parámetro que el de la hipotesis nula, en este caso el nivel de significancia se carga todo hacia el lado derecho, para definir las regiones de aceptación y de rechazo.

Prueba De Hipotesis

H_0: $\mu = \mu_0$

H_1: $\mu > \mu_0$

2.Para el de una cola a la izquierda

El investigador desea comprobar la hipotesis de que el parámetro sea menor que el de la hipótesis nula, en este caso el nivel de significancia se carga todo hacia el lado izquierdo, para definir las regiones de aceptación y de rechazo.

H_0: $\mu = \mu_0$

H_1: $\mu < \mu_0$

3.Para el de dos colas

El investigador desea comprobar la hipotesis de un cambio en el parámetro, es decir no importa si es mayor o menor y 0 que se busca es si hay diferencia con el valor planteado. El nivel de significancia se divide en dos y existen dos regiones de rechazo.

H_0: $\mu = \mu_0$

H_1: $\mu \neq \mu_0$

Si σ es conocido y los datos son normales, aplicamos el teorema del limite central y dependiendo de σ se desea probar.

Ho: $\mu = \mu o$	H_1: $\mu < \mu o$	una cola a la izquierda
Ho: $\mu = \mu o$	H_1: $\mu > \mu o$	una cola a la derecha
Ho: $\mu = \mu o$	H_1: $\mu \neq \mu o$	dos colas

Se compara $z = \frac{\bar{X} - \mu o}{\frac{\sigma}{\sqrt{n}}}\ con\ Z\propto/_2\ ó\ Z_\propto$

6.1.3 Reglas para Rechazar Ho

1. Se selecciona el nivel de significancia ∝ (o nivel de confianza 1 -∝).
2. Encuentra el valor estadístico critico correspondiente (por ejemplo, el $Z \propto$ en la tabla de la distribución normal estándar).
3. Calcula el valor estadístico de la muestra.
4. Si Z cae en el rango critico $Z \propto$, entonces se echaza Ho.

Ejemplos

Se quiere probar si el compostaje obtenido de los residuos orgánicos de una zona comercial es de buena calidad, para lo cual se medirá 5 muestras y se comparará las medias muestrales con los parámetros estadísticos de temperatura, PH, relación carbono-nitrógeno y materia orgánica.

Suponiendo que:

- **a.** la media muestral de temperatura fue de 43.5° y la desviación muestral fue de 2.9°C.
- **b.** el valor promedio de PH de la muestra fue de 6.8 con una desviación muestral de 0.31.
- **c.** la relación promedio de carbono-nitrógeno fue de 24 y la desviación muestral fue de 5.1%.
- **d.** el promedio de materia orgánica fue del 29% con una desviación muestral de 5.1%.

Parámetros:

Temperatura >40°

PH >6.5

Relación C-N >20

Materia Orgánica >25

¿Se puede concluir que el compostaje es de buena calidad si se quiere un nivel de confianza del 95% de los resultados? Se sabe que los datos se distribuyen normalmente.

1) Primer parámetro

n=5

$$t = \frac{\bar{X} - \mu o}{\frac{\sigma}{\sqrt{n}}} = \frac{43.5 - 40}{\frac{2.9}{\sqrt{5}}} = 2.70$$

$\bar{X}$**=**43.5°C

S=2.9°C

$1 - \propto = 95\%$

µo > 40°C

si t >2.132, se rechaza la Ho

si t<2.132, no se rechaza la Ho

Ho: µ= 40°C

H_1: µ >40°C

Conclusión

Se concluye como el estadístico de prueba 2.70 >2.132, entonces se rechaza la Ho, lo que significa que el compostaje si cumple el nivel de temperatura deseado.

2) **Segundo Parámetro**

Zonas Criticas

$\propto=0.05$
$t_{1-\propto}=2.132$
V=n-1=5-1=4

1. Plantear Ho e H_1

Ho: μ= 6.5

H1: μ >6.5

2. Región de rechazo/critico

$\propto=0.05$
$T\propto=2.132$

n= 5

$\bar{X}$**=**6.8

s=0.31
μ=6.5
$1-\propto=95\%$

sí t > 2.132 se rechaza la Ho

sí t < 2.132, no se rechaza la Ho

Calcular el estadístico de prueba

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{6.8 - 6.5}{\frac{0.31}{\sqrt{5}}} = 2.16$$

Conclusión

2.16 >2.132 → se rechaza Ho

Se rechaza la hipotesis nula. El nivel del PH del compostaje cumple con los parámetros.

3) **Tercer Parámetro**

$\bar{X}$**=**24
s=1.12
μ >20
n=5

1.Plantear Ho e Hi

Ho: μ= 20

Hi: μ >20

2. Región de rechazo/

$\propto=0.05$
$T\propto=2.132$

sí t >2.132 se rechaza la Ho

sí t<2.132, no se rechaza la Ho

Calcular el estadístico de prueba

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{24 - 20}{\frac{1.12}{\sqrt{5}}} = 7.99 \quad \rightarrow \quad 7.99 > 2.132 \quad \rightarrow \quad se\ rechaza\ Ho$$

Conclusión

Se rechaza la hipotesis nula y por consiguiente se acepta la hipotesis alternativa esto quiere decir que la relación carbono-nitrógeno es de buena calidad

4) Cuarto parámetro

Ho: μ=25%
H_1: μ >25%

n=5
$\bar{X}$= 29% =0.29
s= 5.1% = 0.051

si t >2.132 se rechaza la Ho

si t <2.132, no se rechaza la Ho

μ:>25%
∝=0.05
t∝=2.132

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{0.29 - 0.25}{\frac{0.051}{\sqrt{5}}} = 1.75$$

Se acepta la hipotesis nula, no se puede rechazar ya que no hay evidencia suficiente para rechazarla.

Conclusión

Se concluye que el compostaje cumple con Temperatura, Nitrógeno y PH, pero no cumple con el parámetro de materia orgánica para sea un compostaje de buena calidad.

Ejercicio

Se requiere comprobar estadísticamente si los pescadores de los sectores de la ciudad de Guayaquil (sector A y B) cumplen con las normas ambientales vigentes con respecto al tamaño permitido del cangrejal para la extracción, para lo cual se pudo obtener la siguiente información durante el muestreo realizado

a) En el sector A se registró el tamaño de 2500 cangrejos dando un promedio muestral de 8.1 con una desviación de 3.07 cm.
b) En el sector B se registró el tamaño de 2900 cangrejos dando un promedio de 6.9 con una desviación de 3.22 cm

¿Ambos sectores cumplen con la normativa ambiental el cual especifica que el tamaño mínimo permitido de extracción es de 7cm? Asuma un nivel de confianza del 95% ¿Existen diferencias en el tamaño de los cangrejos extraídos entre los sectores Ay B?

a.-
$\bar{X}$=8.1
s=3.07
n=2500
μ=>7
1-∝: 0.95

Ho: μ=7

H_1: μ >7

∝=0.05
t=1.645

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{8.1 - 7}{\frac{3.07}{\sqrt{2500}}} = 17.92 \quad \rightarrow \quad 17.92 > 1.645$$

El sector A cumple con la normativa (se rechaza Ho)

b.-
$\bar{X}$=6.9
s=3.32
n=2500
1-∝: 0.95

Ho: μ=7

H1: μ > 7

∝=0.05
t∝=1.65

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{6.9 - 7}{\frac{3.32}{\sqrt{2900}}} = -1.623 \quad \rightarrow \quad -1.623 > 1.645$$

el sector B no cumple con la normativa (no se rechaza Ho)

6.2 Valor P de la Prueba

Es la probabilidad de observar un valor externo de la estadística a prueba y se supone que la hipotesis nula es cierta

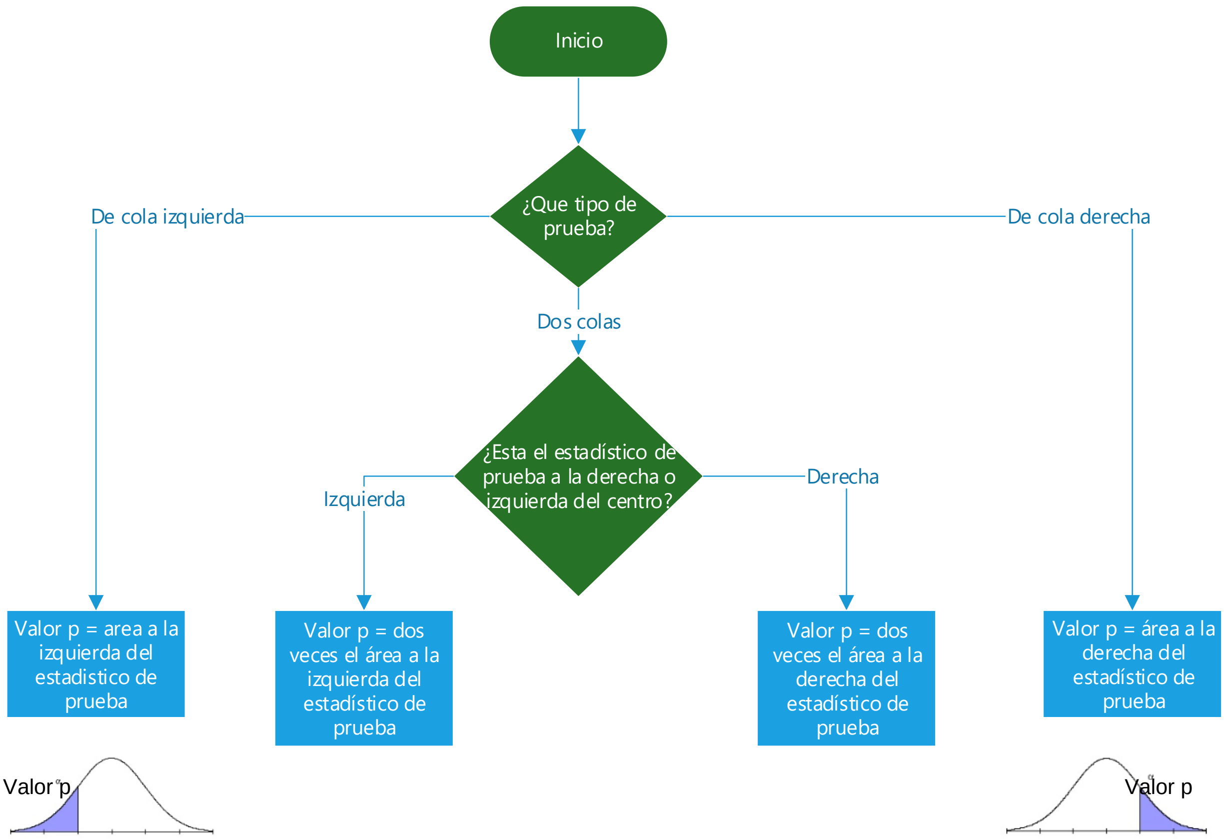
sí Ho es cierto y la alternativa es H_1: μ < μo ¿Cuál es la probabilidad de observar z< -2.41?

El área desde z=-2.41 hacia el externo izquierdo nos da un valor de 0.00798 por 6 que ese es el valor P

Si p<∝ → se rechaza Ho

Si p >∝→ no se rechaza

6.2.1 Calculo de Valores P



6.3 Comparación de dos medias poblacionales a través de t-student

Los resultados de un método analítico nuevo se pueden contrastar mediante comparación con los obtenidos utilizando un segundo método (quizá uno de referencia). En este caso tenemos dos medias muestrales $\bar{x}_1$ y $\bar{x}_2$. Tomando como hipotesis nula que los dos métodos proporcionan el mismo resultado, es decir Ho: μ1= μ2, se necesita probar si ($\bar{x}_1$ - $\bar{x}_2$) difiere significativamente de cero. (Miller, James; Miller, 2002)

Ho: μ1= μ2 | Ho: T1=T2

Vs

H1: μ1≠ μ2 | H1: T1≠ T2

Test de normalidad

Para la prueba de normalidad dependerá del tamaño de muestra, KS cuando n >= 30 y Shapiro Wilks n<30

Ho: Los datos siguen una distribución Normal

Vs

H1: Los datos NO siguen una distribución Normal

Prueba homogeneidad de Varianzas (LEVENE)

Ho: $\sigma 1^2 = \sigma 2^2$ → Homocedasticidad

Vs

H1: $\sigma i^2 \neq \sigma 2^2$ → Heterocedasticidad

Ejercicio

A continuación, haremos la prueba de hipotesis con media de dos poblaciones utilizando los datos del ejercicio anterior del tamaño de los cangrejos:

Formulas

$$z = \frac{\bar{x}A - \bar{x}B - do}{Sp\sqrt{\frac{1}{nA} + \frac{1}{nB}}}$$

$$Sp = \sqrt{\frac{(nA-1){S_A}^2 + (nB-1){S_B}^2}{nA + nB - 2}}$$

nA:2500
nB 2900
$\bar{x}A$ 8.1
$\bar{x}B$ 6.
S_A 3.07
S_B 3.32
1-∝: 0.95

Ho: $\mu_A = \mu_B$

H_1: $\mu_A \neq \mu_B$

sí -1.96 < 3 < 1.96 no rechazo Ho, caso contrario se rechaza Ho

$$Sp = \sqrt{\frac{(2500-1)(3.07)^2 + (2900-1)(3.32)^2}{2500+2900-2}} = 3.20$$

$$z = \frac{8.1 - 6.9}{(3.20)\sqrt{\frac{1}{2500} + \frac{1}{2900}}} = 13.74$$

conclusión

Si hay una diferencia de tamaño de los cangrejos en los grupos A y B, se rechaza Ho

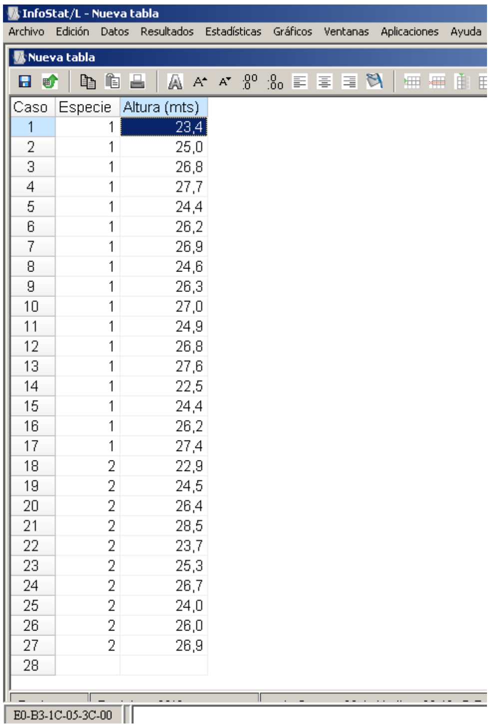
Otro ejemplo aplicado

Los datos de la siguiente tabla se refieren a las alturas (en metros) de árboles en muestras aleatorios e independientes de dos especies forestales diferentes (1 y 2). Verifique si las alturas medias de los árboles de las dos especies no difieren entre si, considerándose un nivel de significancia del 5%.

Ho: $\mu 1 = \mu 2$ (La altura promedio de los árboles de las dos especies son iguales)

Vs

H1: $\mu 1 \neq \mu 2$ (La altura promedio de los árboles de las dos especies son diferentes)



Caso	Especie	Altura (mts)
1	1	23,4
2	1	25,0
3	1	26,8
4	1	27,7
5	1	24,4
6	1	26,2
7	1	26,9
8	1	24,6
9	1	26,3
10	1	27,0
11	1	24,9
12	1	26,8
13	1	27,6
14	1	22,5
15	1	24,4
16	1	26,2
17	1	27,4
18	2	22,9
19	2	24,5
20	2	26,4
21	2	28,5
22	2	23,7
23	2	25,3
24	2	26,7
25	2	24,0
26	2	26,0
27	2	26,9
28		

Shapiro-Wilks (modificado)

Variable	n	Media	D.E.	W*	p(Unilateral D)
Altura (mts)	27	25,67	1,59	0,94	0,3272

Test de normalidad

Para la prueba de normalidad dependerá del tamaño de muestra, KS cuando n >= 30 y Shapiro Wilks n<30, en este caso la muestra es n< 30 por lo que usamos la prueba Shapiro Wilk.

Ho: Los datos siguen una distribución Normal

Vs

H1: Los datos NO siguen una distribución Normal

Si p-valor< α rechazo H_0 → 0.3272 > 0.05 aceptar H_0

Prueba F para igualdad de varianzas

Variable	Grupo(1)	Grupo(2)	n(1)	n(2)	Var(1)	Var(2)	F	p	prueba
Altura (mts)	{1}	{2}	17	10	2,38	2,97	0,80	0,6682	Bilateral

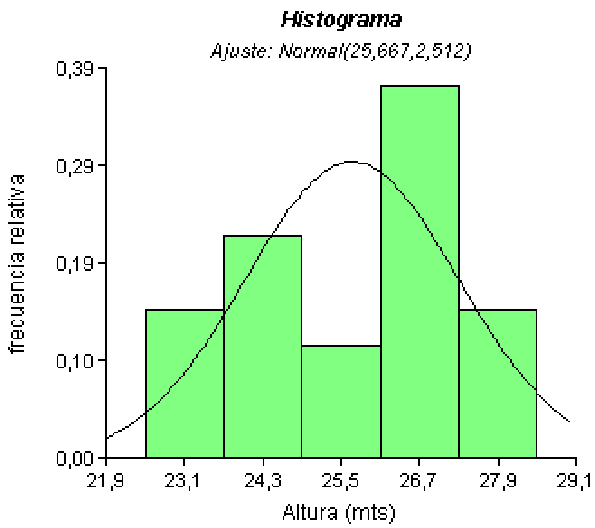
Prueba homogeneidad de Varianzas (LEVENE)

Ho: $\sigma 1^2 = \sigma 2^2$ → Homocedasticidad

Vs

H1: $\sigma i^2 \neq \sigma 2^2$ → Heterocedasticidad

Si p-valor< α rechazo H_0 → 0.6682 > 0.05 acepto H_0, es decir tienen igualdad de varianzas



Prueba T para muestras Independientes

Variable: Altura (mts) - Clasific: Especies - prueba:Bilateral

	Grupo 1	Grupo 2
	1	2
n	17	10
Media	25,77	25,49
Media(1)-Media(2)	0,28	
LI(95)	-1,04	
LS(95)	1,60	
pHomVar	0,6682	
T	0,44	
gl	25	
p-valor	**0,6657**	

Si p-valor< α rechazo H_0 → 0.6657 > 0.05 acepto H_0, es decir la media de altura en metros son iguales estadísticamente lo que concluye que no hay significancia entre las especies forestales.

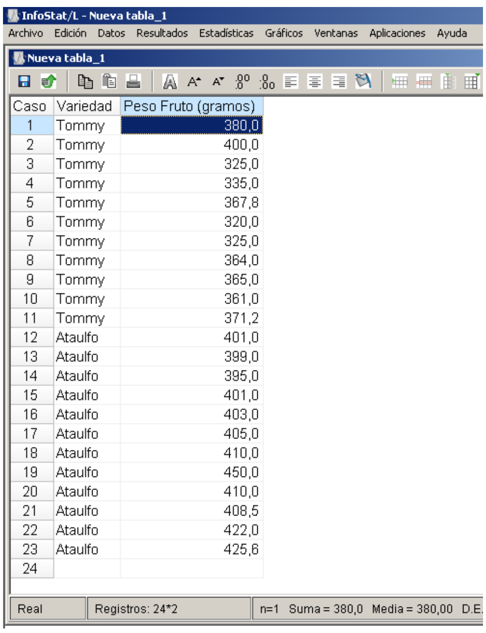
Los datos de la siguiente tabla se refieren al peso del fruto (gramos) en muestras aleatorios e independientes de dos variedades de mango, Tommy Atkins y Ataulfo. Verifique si las medias de los pesos del fruto de los dos árboles de mango no difieren entre sí, considerándose un nivel de significancia del 5%.

Hipotesis

Ho: $\mu 1 = \mu 2$ (La media del peso del fruto de las dos variedades son iguales)

Vs

H1: $\mu 1 \neq \mu 2$ (La media del peso del fruto de las dos variedades son diferentes)



Caso	Variedad	Peso Fruto (gramos)
1	Tommy	380,0
2	Tommy	400,0
3	Tommy	325,0
4	Tommy	335,0
5	Tommy	367,8
6	Tommy	320,0
7	Tommy	325,0
8	Tommy	364,0
9	Tommy	365,0
10	Tommy	361,0
11	Tommy	371,2
12	Ataulfo	401,0
13	Ataulfo	399,0
14	Ataulfo	395,0
15	Ataulfo	401,0
16	Ataulfo	403,0
17	Ataulfo	405,0
18	Ataulfo	410,0
19	Ataulfo	450,0
20	Ataulfo	410,0
21	Ataulfo	408,5
22	Ataulfo	422,0
23	Ataulfo	425,6
24		

Shapiro-Wilks (modificado)

Variable	n	Media	D.E.	W*	p(Unilateral D)
Peso Fruto (gramos)	23	384,53	34,80	0,92	0,1705

Test de normalidad

Para la prueba de normalidad dependerá del tamaño de muestra, KS cuando n >= 30 y Shapiro Wilks n<30

Ho: Los datos siguen una distribución Normal

Vs

H1: Los datos NO siguen una distribución Normal

Si p-valor< α rechazo HO → 0.1705 > 0.05 aceptar HO

Prueba homogeneidad de Varianzas (LEVENE)

Ho: $\sigma 1^2 = \sigma 2^2$ → Homocedasticidad

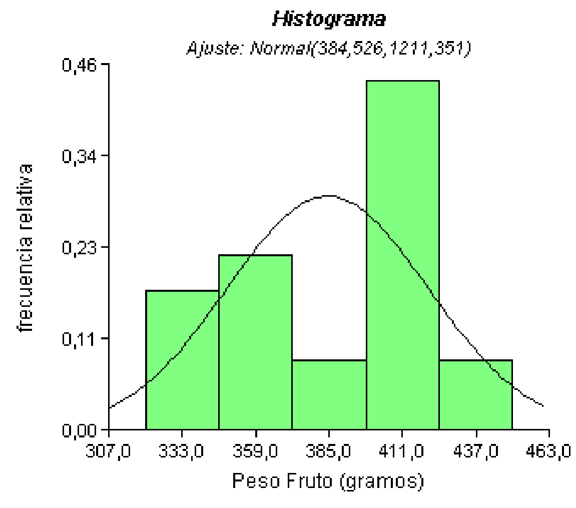
Vs

H1: $\sigma i^2 \neq \sigma 2^2$ → Heterocedasticidad

Si p-valor< α rechazo HO → 0.0.0972 > 0.05 aceptar HO, es decir tienen igualdad de varianzas

Prueba F para igualdad de varianzas

Variable	Grupo(1)	Grupo(2)	n(1)	n(2)	Var(1)	Var(2)	F	p	prueba
Peso Fruto(gramos)	{Ataulfo}	{Tommy}	12	11	233,01	671,09	0,35	0,0972	Bilateral



Variable:Peso Fruto (gramos) - Clasific:Variedad - prueba:Bilateral

	Grupo 1	Grupo 2
	Ataulfo	Tommy
n	12	11
Media	410,84	355,82

Media(1)-Media(2)	55,02
LI(95)	36,78
LS(95)	73,27
pHomVar	0,0972
T	6,27
gl	21
p-valor	<0,0001

Si p-valor< α rechazo Ho → 0.0001 < 0.05 rechazo Ho, es decir la media de peso de fruto en gramo son diferentes estadísticamente lo que concluye que hay significancia entre las variedades de mango.

6.4 Métodos No Paramétricos

No suponen conocimiento de ninguna clase acerca de las distribuciones de las poblaciones subyacentes y excepto, quizás que estas son continuas.

	Métodos paramétricos	Métodos no paramétricos
1 Muestra	Z, t-student	Prueba de signos
2 Muestras	Z, t-student	Suma de rangos de Wilcoxon
Más de 2 muestras	ANOVA	Kruskal-Wallis

Los ejercicios son tomados de (Walpole Ronald, Myers Raymond, Myers Sharon, 2012)

6.4.1 Prueba de Signos

Se usa para hacer pruebas de hipotesis acerca de la mediana de una población de una variable continua.

La media es reemplazada por la mediana como el parámetro de ubicación pertinente a probar.

Ho: la mediana poblacional es igual a un valor dado.

H_1: la mediana es menor (mayor o distinto) del valor dado.

Basada en la distribución binomial con probabilidad de éxito p=1/2

Si $\tilde{x} > 0.05$ *no se rechaza Ho*

Si $\tilde{x} < 0.05$ *se rechaza Ho*

para calcular se determina las diferencias de los datos con respecto al valor dado de la mediana y se cuentan los signos positivos y negativos.

Cuando la hipotesis alternativa (H_1) es mayor que y el número de diferencias positivas es mayor/ menor que las diferencias negativas, entonces el valor p se calcula por:

$Hi: \tilde{\mu} > \tilde{\mu}o$ $\quad P_1 = \Sigma\binom{n}{i}\left(\frac{1}{2}\right)^n \quad +> -(\,cuando\ hay\ mas\ numeros\ positivos\ que\ negativos)$

$P_2 = \Sigma\binom{n}{i}\left(\frac{1}{2}\right)^n \quad +< -(cuando\ hay\ mas\ numeros\ negativos\ que\ positivos)$

$Hi: \tilde{\mu} < \tilde{\mu}o$ $\quad P_1 = (X \geq x)$ $\quad P_2$ = P(X$\leq x$)

$Hi: \tilde{\mu} \neq \tilde{\mu}o$ $\quad Valor\ P = 2P_2$ $\quad Valor\ P = 2P_1$

$+= \ - \rightarrow \ valor\ P = 1$

Obs. No se consideran números iguales a la mediana

Ejemplos

1. Los siguientes datos representan el número de horas que funciona una maquina antes de requerir una recarga. $1^{-}2, 2^{+}2, 0^{-}9, 1^{-}3, 2^{+}0, 1^{-}6, 1^{-}5, 2^{+}0, 1^{-}2, 1^{-}7$. a un nivel de significancia de 0.05 utilice la prueba de signos para probar la hipotesis de que esta maquina funciona con una mediana de 1.8 horas antes de requerir una recarga.

Ho:$\tilde{\mu}$ =1.8

H_1: $\tilde{\mu}$ ≠1.8

X=3 positivos

$Valor\ p = 2P_1\left[(X \leq 3)\ cuando\ p = \frac{1}{2}\right] = 2\left[\sum_{i=0}^{3}\binom{10}{i}\left(\frac{1}{2}\right)^{10}\right] = 2[10C0 + 10C1 + 10C2 + 10C3]\left(\frac{1}{2}\right)^{10} =$ $0.3438 > 0.05$ si es mayor, no se rechaza Ho.

2. Los siguientes datos representan el tiempo en minutos que un paciente tiene que esperar durante 12 visitas al consultorio de un médico antes de ser atendido:

-	-			+	+
17	15	20	20	32	28
12	26	25	25	35	24
-	+	+	+	+	+

Utilice la prueba de signos a un nivel de significancia de 0.05 para probar la afirmación del médico de que la mediana del tiempo de espera de sus pacientes no es mayor a 20 minutos.

Ho: $\tilde{\mu}$ =20

H_1: $\tilde{\mu} < 20$

X=7 signos positivos

n=10

$Valor\ p = \sum_{i=7}^{10} \binom{10}{i} \left(\frac{1}{2}\right)^{10} = (10C7 + 10C8 + 10C9 + 10C10)\left(\frac{1}{2}\right)^{10} = 0.1718 \quad > 0.05$ no se rechaza Ho.

3. Un inspector de alimentos examina 16 latas de cierta marca de jamón para determinar el porcentaje de impureza externas se registraron los siguientes datos:

2.4	2.3	3.1	2.2	2.3	1.2	1.0	2.4
1.7	1.1	4.2	1.9	1.7	3.6	1.6	2.3

Realice una prueba de signos a un nivel de significancia de 0.05 para probar la hipotesis nula de que la mediana del porcentaje de impureza en esta marca es de 2.5%, en comparación con la hipotesis alternativa de que la mediana del porcentaje de impurezas no es de 2.5%.

Ho: $\tilde{\mu} = 2.5$

H_1: $\tilde{\mu} \neq 2.5$ valor p= 2 [p(X<=3)]

X = 3 positivos

$Valor\ p = 2p_1 \qquad Valor\ p = 2\left[\sum_{i=0}^{3} \binom{16}{i} \left(\frac{1}{2}\right)^{16}\right]$

$= 2(16C0 + 16C1 + C16C2 + 16C3)\left(\frac{1}{2}\right)^{10} = 0.021 \ < 0.05$ Se rechaza Ho

4. Se tomaron 10 muestras de un baño de platinado utilizado en un proceso de manufacturación y se determinó el ph del baño. Los valores de ph de la muestra son los siguientes:

+	**+**	**-**	**+**	**+**	**-**	**+**	**+**	**+**	**+**
7.91	7.85	6.82	8.01	7.46	6.95	7.05	7.35	7.25	7.42

Los ingenieros del departamento de manufactura creen que el ph tiene un valor medio de 7.0¿los datos de la muestra indican que este anunciado es correcto? Emplee la prueba de hipotesis para comprobar esta hipotesis.

Ho: $\tilde{\mu} = 7.0$ $\qquad Valor\ P = 2P_2$

H_1: $\tilde{\mu} \neq 7.0$ $\qquad Valor\ P = 2\left[\sum_{i=8}^{10} \binom{10}{i} \left(\frac{1}{2}\right)^{10}\right]$

X=8 número de positivos $\qquad = 2(10C8 + 10C9 + 10C10)\left(\frac{1}{2}\right)^{10} = 0.1094 > 0.05$

n=10 $\qquad no\ se\ rechaza\ Ho$

6.4.2 Prueba no Paramétricas para dos Muestras

6.4.2.1 Prueba de la suma de rangos de Wilcoxon (prueba U-Mann Whitney)

- Aplicada a dos muestras independientes, versión no paramétrica de la habitual t-student
- La hipotesis nula (Ho) es que la mediana de las dos poblaciones son iguales y la hipotesis (Hi) puede ser que la mediana de la población 1 sea mayor (menor o diferente) de la mediana de la población 2.

Pasos

1. Determina el tamaño de la muestra (n_1 y n_2). Si n_1 y n_2 son menores que 20, se considera muestras pequeñas
2. Ordenar los datos en rango, del menor al mayor valor en caso de que existen empates se saca promedio.
3. Calcular los valores de U_1 y U_2 de modo que se elija el más pequeño para comparar con los valores críticos de U-Mann Whitney de la tabla
4. En caso de muestras grandes (n_1 y n_2 mayor a 20) calcular el valor z, pues en estas condiciones se distribuye normalmente.
5. Decidir si se acepta o rechaza la hipotesis

$H1:\ \tilde{\mu}1 < \ \tilde{\mu}_2 \rightarrow\ U_1\ =\ n_1 n_2 + \frac{n_1(\ n_1 - 1)}{2} - R_1$ $\quad\quad \rightarrow H1:\ \tilde{\mu}1 >\ \tilde{\mu}2\ \rightarrow\ U_2\ =\ n_1 n_2 + \frac{n_2(\ n_2 - 1)}{2} -\ R_2$

$H1:\ \tilde{\mu}_1 \neq\ \tilde{\mu}_2 \quad U = \min(U_1, U_2)$ para determinar se debe sacar U_1 y U_2

n_1 y $n_2 \rightarrow$ tamaño respecto de cada muestra

$R_1\ y\ R_2 \rightarrow$ suma de los rangos de las observaciones de las muestras 1 y 2 respectivamente

El estadístico U se define como el mínimo de U_1 y U_2 . si el valor de U_1 y U_2 o U es menor o igual que el valor critico tabulado, se rechaza la hipotesis nula al nivel de significancia que se indica en la tabla

U_1, U_2, U < valor critico de tabla rechaza

En casos mayores a 20 (n> 20)

$$z = \frac{U - mu}{\sigma_U}$$

mu: media
σ_U: desviacion estándar

$$mu = \frac{n_1 n_2}{2} \qquad\qquad \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Ejemplos

1. Un experimento utiliza dos métodos para ensenar a leer a un grupo de 0 niños de 6 años, quieren ingresar por primera vez a la escuela. El experimento quiere demostrar que el procedimiento descrito por el es el más efectivo en función de la fluidez, comprensión, análisis y síntesis cuyos resultados se muestran

(Ordenar los datos de menor a mayor) para encontrar la suma de rangos

Tradicional (grupo 1)	80	95	25	70	90	$\rightarrow R_1$=19	n_1=5
Inventado por el investigador (grupo 2)	95	100	93	110	45	$\rightarrow R_2$=36	n_2=5

Ho: $\tilde{\mu}_1 = \tilde{\mu}_2$

H1: $\tilde{\mu}_1 \neq \tilde{\mu}_2$

$$U_1 = (5)(5) + \frac{5(5+1)}{2} - 19 = 21$$

$$U_2 = (5)(5) + \frac{5(5+2)}{2} - 36 = 4$$

U= min(21; 4) → U=4

Buscar en la tabla

n_1= 5 4 > 2 no se rechaza Ho

n_2 = 5

Como U=4 es mayor al criterio de la tabla (valor igual a 2) entonces no se rechaza la hipotesis y se concluye que los dos métodos dan iguales resultados a un nivel de confianza del 95%.

2. Se encontró que el contenido de nicotina de dos marcas de cigarrillos, medido en miligramos es el siguiente:

(ordenar los datos de menor a mayor)

Marca A	2.1	4.0	6.3	5.4	4.8	3.7	6.1	3.3	5.4	R_A=93
Marca B	.1	0.6	3.1	2.5	4.0	6.2	1.6	2.2	1.9	R_B=78

A un nivel de significancia de 0.5 prueba de hipotesis de que las medianas del contenido de nicotina de las dos marcas son iguales, en comparación con la hipotesis alternativa de que son diferentes.

Ho: $\tilde{\mu}_1 = \tilde{\mu}_2$

Hi: $\tilde{\mu}_1 \neq \tilde{\mu}_2$

$$U_1 = (8)(10) + \frac{8(9)}{2} - 93 = 23$$

$$U_2 = (8)(10) + \frac{8(11)}{2} - 78 = 57$$

U= min(23; 57) → U=23

$\left.\begin{array}{l} n_1 = 8 \\ n_2 = 10 \end{array}\right]$ 23 > 17 Si U es menor a U en la tabla entonces rechaza H_0

Se concluye con no hay evidencia estadística para rechazar la hipotesis nula, por lo que no hay diferencias entre las dos marcas en el contenido de nicotina.

6.5 Ejercicios complementarios capítulo 6

1. En una comparación de dos métodos para la determinación de cromo en muestras de hierba de centeno se obtuvieron los siguientes resultados (mg kg^{-1}):

   Método 1: Media=1.48, desviación estándar =0.28

   Método 2: Media=2.33, desviación estándar =0.31

   Para cada método se tomaron 5 muestras, utilice un α =5%

   a) Plantee las hipotesis
   b) ¿Estos dos métodos proporcionan resultados cuyas medias difieren significativamente?

2. Los siguientes datos proporcionan la recuperación de bromuro adicionado a muestras con contenido vegetal, medido mediante un método cromatográfico gas-liquido. La cantidad de bromuro potásico añadido a cada tipo de vegetal fue la misma.

Tomate	777	790	759	790	770	758	764	768	762	µg g^{-1}
Pepino	782	773	778	765	789	797	782	792	793	µg g^{-1}

(Roughan, J. A., Roughan, P. A. and Wilkins, J. P. G. 1983 *Analyst* 108:742)

   a) Contrastar si la recuperación en los dos vegetales tiene varianzas que difieren significativamente
   b) Contrastar si las tasas de recuperación medias difieren significativamente

Bibliografía

Albornoz, V. (2011). La población del Ecuador 1950 - 2010. *Carta Economica*, 1.

Balzarini, Monica; Di Rienzo, Julio; Tablada, M., & Bruno, C. (2011). *Estadística y Biometría Ilustraciones* (1era ed.). Buenos Aires: Brujas.

Brown, L., & Mac Berthouex, P. (2010). Statistics for Environmental Engineers, Second Edition. In *Statistics for Environmental Engineers, Second Edition* (Second Edi). https://doi.org/10.1201/9781420056631

Díaz Monroy, L. G., & Morales Rivera, M. A. (2012). *Estadística multivariada: inferencia y métodos* (Tercera Ed). Bogotá: Editorial Universidad Nacional de Colombia.

Fernández Escobar, R., Trapero, A., & Domínguez, J. (2010). *Experimentación en la agricultura* (Junta de A). Sevilla.

López, E., & González, B. (2015). *Estadística: Fundamentos y Aplicaciones en Agronomia y ciencias afines*. Ciudad de Guatemala.

Mendenhall, William III; Wackerly, Dennis; Scheaffer, R. (2009). Mathematical statistics with applications. In *Computational Statistics & Data Analysis* (Seventh Ed, Vol. 13). https://doi.org/10.1016/0167-9473(92)90162-9

Miller, James; Miller, J. (2002). *Estadistica y Quimiometria para Quimica Analitica* (4th_ed). Madrid: Pearson Education.

Milton, S. J. (2001). *Estadística para Biologia y Ciencias de la Salud* (Tercera Ed). Madrid: Mc Graw-Hill.

Montgomery, Douglas; Peck, Elizabeth; Vining, G. (2006). *Introducción al Analisis de Regresión Lineal*. Ciudad de Mexico.

Rienzo, D., Alejandro, J., Alicia, L., Margot, E., & Pilar, M. (2008). *Estadística para las Ciencias Agropecuarias* (Septima Ed). Cordoba: Brujas.

Spiegel, Murray R.; Stephens, L. J. (2009). *Estadística* (Cuarta Edi). Mexico: Mc Graw-Hill.

Triola, F. M. (2009). *Estadística* (Decima Edi). Mexico.

Walpole Ronald, Myers Raymond, Myers Sharon, Y. K. (2012). *Probabilidad y Estadística para Ingenieria y Ciencias* (Novena Edi). https://doi.org/10.1192/bjp.112.483.211-a

Tablas Estadísticas

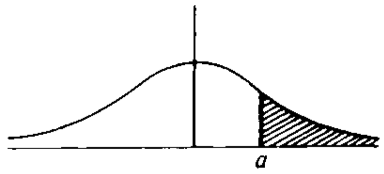
Las tablas estadísticas de este texto son una cortesía las cuales agradecemos, fueron tomadas de la Universidad Nacional de Quilmes.
http://materias.unq.edu.ar/pye/Trabajos%20Pr%C3%A1cticos/Tablas%20de%20Estadistica.pdf

Tabla A-6 Valores críticos del coeficiente de Correlación r de Pearson

n	$\alpha = .05$	$\alpha = .01$
4	.950	.999
5	.878	.959
6	.811	.917
7	.754	.875
8	.707	.834
9	.666	.798
10	.632	.765
11	.602	.735
12	.576	.708
13	.553	.684
14	.532	.661
15	.514	.641
16	.497	.623
17	.482	.606
18	.468	.590
19	.456	.575
20	.444	.561
25	.396	.505
30	.361	.463
35	.335	.430
40	.312	.402
45	.294	.378
50	.279	.361
60	.254	.330
70	.236	.305
80	.220	.286
90	.207	.269
100	.196	.256

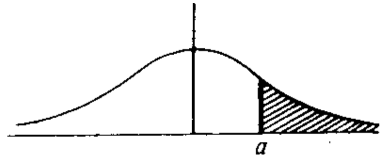
NOTA: Para probar H_0: $\rho = 0$ contra H_1: $\rho \neq 0$, se rechaza H_0 si el valor absoluto de r es mayor que el valor crítico que se indica en la tabla.

Tabla 1. Distribución normal (0; 1). *P* (X ≥ *a*)



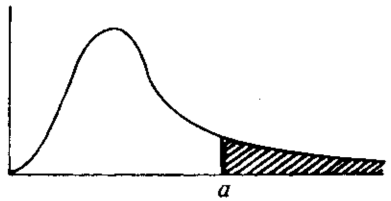
	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4841	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4091	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2644	0,2611	0,2579	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2297	0,2266	0,2236	0,2207	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1563	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1094	0,1075	0,1057	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0352	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233

Tabla 1 (Continuación). Distribución normal (0; 1). P (X ≥ a)



	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,0082	0,0080	0,0078	0,0076	0,0073	0,0071	0,0070	0,0068	0,0066	0,0064
2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
2,6	0,0047	0,0045	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036
2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
3,0	0,0014	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
3,1	0,0010	0,0009	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0007	0,0007
3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004
3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
3,6	0,0002	0,0002	0,0002	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
3,9	0,0001	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

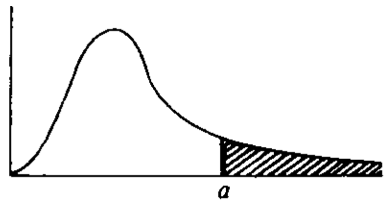
Tabla 2. Distribución X^2. $P\ (X^2 \geq a)$



Grados de libertad	Probabilidades										
	0,99	0,975	0,95	0,90	0,75	0,50	0,25	0,10	0,05	0,025	0,01
1	1,571*	9,821*	39,320*	0,016	0,102	0,455	1,323	2,706	3,841	5,024	6,635
2	0,020	0,051	0,103	0,211	0,575	1,386	2,773	4,605	5,991	7,378	9,210
3	0,115	0,216	0,352	0,584	1,213	2,366	4,108	6,251	7,815	9,348	11,345
4	0,297	0,484	0,717	1,064	1,923	3,357	5,385	7,779	9,488	11,143	13,277
5	0,554	0,831	1,145	1,610	2,675	4,351	6,626	9,236	11,070	12,833	15,086
6	0,872	1,237	1,635	2,204	3,455	5,348	7,841	10,645	12,592	14,449	16,812
7	1,239	1,690	2,167	2,833	4,255	6,346	9,037	12,017	14,067	16,013	18,475
8	1,646	2,180	2,733	3,490	5,071	7,344	10,219	13,362	15,507	17,535	20,090
9	2,088	2,700	3,325	4,168	5,899	8,343	11,389	14,684	16,919	19,023	21,666
10	2,558	3,247	3,940	4,865	6,737	9,342	12,549	15,987	18,307	20,483	23,209
11	3,053	3,816	4,575	5,578	7,584	10,341	13,701	17,275	19,675	21,920	24,725
12	3,571	4,404	5,226	6,304	8,438	11,340	14,845	18,549	21,026	23,337	26,217
13	4,107	5,009	5,892	7,041	9,299	12,340	15,984	19,812	22,362	24,736	27,688
14	4,660	5,629	6,571	7,790	10,165	13,339	17,117	21,064	23,685	26,119	29,141
15	5,229	6,262	7,261	8,547	11,036	14,339	18,245	22,307	24,996	24,788	30,578
16	5,812	6,908	7,962	9,312	11,912	15,338	19,369	23,542	26,296	28,845	32,000
17	6,408	7,564	8,672	10,085	12,792	16,338	20,489	24,769	27,587	30,191	33,409
18	7,015	8,231	9,390	10,865	13,675	17,338	21,605	25,989	28,869	31,526	34,805
19	7,633	8,907	10,117	11,651	14,562	18,338	22,718	27,204	30,143	32,852	36,191

* Dividir entre 1000.

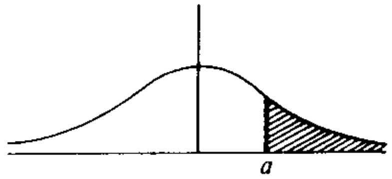
Tabla 2 (Continuación). Distribución X^2. $P\ (X^2 \geq a)$



Grados de libertad	Probabilidades										
	0,99	0,975	0,95	0,90	0,75	0,50	0,25	0,10	0,05	0,025	0,01
20	8,260	9,591	10,851	12,443	15,452	19,337	23,828	28,412	31,410	34,170	37,566
21	8,897	10,283	11,591	13,240	16,344	20,337	24,935	29,615	32,670	35,479	38,932
22	9,542	10,982	12,338	14,041	17,240	21,337	26,039	30,813	33,924	36,781	40,289
23	10,196	11,688	13,090	14,848	18,137	22,337	27,141	32,007	35,172	38,076	41,638
24	10,856	12,401	13,848	15,659	19,037	23,337	28,241	33,196	36,415	39,364	42,080
25	11,524	13,120	14,611	16,473	19,939	24,337	29,339	34,382	37,652	40,646	44,314
26	12,198	13,844	15,379	17,292	20,843	25,336	30,434	35,563	38,885	41,923	45,642
27	12,879	14,573	16,151	18,114	21,749	26,336	31,528	36,741	40,113	43,194	46,963
28	13,565	15,308	16,928	18,939	22,657	27,336	32,620	37,916	41,337	44,461	48,278
29	14,256	16,047	17,708	19,768	23,567	28,336	33,711	39,087	42,557	45,722	49,588
30	14,954	16,791	18,493	20,599	24,478	29,336	34,800	40,256	43,773	46,979	50,892
40	22,164	24,433	26,509	29,050	33,660	39,335	45,616	51,805	55,758	59,342	63,691
50	29,707	32,357	34,764	37,689	42,942	49,335	56,334	63,167	67,505	71,420	76,154
60	37,485	40,482	43,188	46,459	52,294	59,335	66,981	74,397	79,082	83,298	88,379
70	45,442	48,758	51,739	55,329	61,698	69,334	77,577	85,527	90,531	95,023	100,425
80	53,540	57,153	60,391	64,278	71,144	79,334	88,130	96,578	101,879	106,629	112,329
90	61,754	65,647	69,126	73,291	80,625	89,334	98,650	107,565	113,145	118,136	124,116
100	70,065	74,222	77,929	82,358	90,133	99,334	109,141	118,498	124,342	129,561	135,807

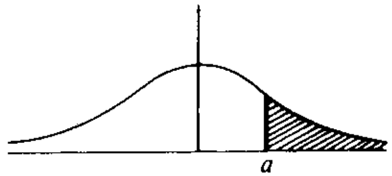
* Dividir entre 1000.

Tabla 3. Distribución *t* de Student. *P* [*t* (*n*) ≥ *a*]



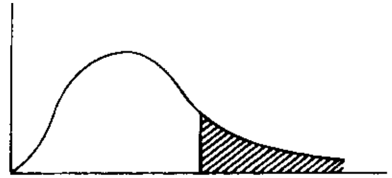
Grados de libertad	Probabilidades							
	0,40	0,25	0,15	0,10	0,05	0,025	0,01	0,005
1	0,3249	1,0000	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567
2	0,2887	0,8165	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248
3	0,2767	0,7649	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409
4	0,2707	0,7407	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041
5	0,2672	0,7267	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321
6	0,2648	0,7176	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074
7	0,2632	0,7111	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995
8	0,2619	0,7064	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554
9	0,2610	0,7027	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498
10	0,2602	0,6998	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693
11	0,2596	0,6974	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058
12	0,2590	0,6955	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545
13	0,2586	0,6938	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123
14	0,2582	0,6924	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768
15	0,2579	0,6912	1,0735	1,3406	1,7531	2,1314	2,6025	2,9467
16	0,2576	0,6901	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208
17	0,2573	0,6892	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982
18	0,2571	0,6884	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784
19	0,2569	0,6876	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609
20	0,2567	0,6870	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453
21	0,2566	0,6864	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314
22	0,2564	0,6858	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188

Tabla 3 (Continuación). Distribución *t* de Student. *P* [*t* (*n*) ≥ *a*]



Grados de libertad	Probabilidades							
	0,40	0,25	0,15	0,10	0,05	0,025	0,01	0,005
23	0,2563	0,6853	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073
24	0,2562	0,6848	1,0593	1,3178	1,7109	2,0639	2,4922	2,7969
25	0,2561	0,6844	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874
26	0,2560	0,6840	1,0575	1,3150	1,7056	2,0555	2,4786	2,7787
27	0,2559	0,6837	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707
28	0,2558	0,6834	1,0560	1,3125	1,7011	2,0484	2,4671	2,7633
29	0,2557	0,6830	1,0553	1,3114	1,6991	2,0452	2,4620	2,7564
30	0,2556	0,6828	1,0547	1,3104	1,6973	2,0423	2,4573	2,7500
35	0,2553	0,6816	1,0520	1,3062	1,6896	2,0301	2,4377	2,7238
40	0,2550	0,6807	1,0500	1,3031	1,6839	2,0211	2,4233	2,7045
45	0,2549	0,6800	1,0485	1,3006	1,6794	2,0141	2,4121	2,6896
50	0,2547	0,6794	1,0473	1,2987	1,6759	2,0086	2,4033	2,6778
60	0,2545	0,6786	1,0455	1,2958	1,6706	2,0003	2,3901	2,6603
70	0,2543	0,6780	1,0442	1,2938	1,6669	1,9944	2,3808	2,6479
80	0,2542	0,6776	1,0432	1,2922	1,6641	1,9901	2,3739	2,6387
90	0,2541	0,6772	1,0424	1,2910	1,6620	1,9867	2,3685	2,6316
100	0,2540	0,6770	1,0418	1,2901	1,6602	1,9840	2,3642	2,6259
120	0,2539	0,6765	1,0409	1,2886	1,6577	1,9799	2,3578	2,6174
150	0,2538	0,6761	1,0400	1,2872	1,6551	1,9759	2,3515	2,6090
200	0,2537	0,6757	1,0391	1,2858	1,6525	1,9719	2,3451	2,6006
300	0,2536	0,6753	1,0382	1,2844	1,6499	1,9679	2,3388	2,5923
∞	0,2533	0,6745	1,0364	1,2816	1,6449	1,9600	2,3263	2,5758

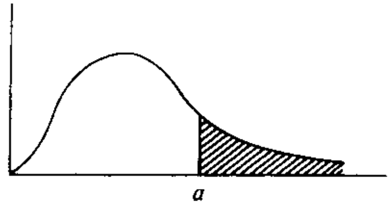
Tabla 4. Distribución *F de FISHER*. $P\,[F\,(m;\,n) \geq a] = 0{,}001$



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4053*	5000*	5404*	5625*	5764*	5859*	5929*	5981*	6023*	6056*	6107*	6158*	6209*	6235*	6261*	6287*	6313*	6340*	6366*
2	998,50	999,00	999,20	999,20	999,30	999,30	999,40	999,40	999,40	999,40	999,40	999,40	999,40	999,50	999,50	999,50	999,50	999,50	999,50
3	167,00	148,50	141,10	137,10	134,60	132,80	131,60	130,60	129,90	129,20	128,30	127,40	126,40	125,90	125,40	125,00	124,50	124,00	123,50
4	74,14	61,25	56,18	53,44	51,71	50,53	49,66	49,00	48,47	48,05	47,41	46,76	46,10	45,77	45,43	45,09	44,75	44,40	44,05
5	47,18	37,12	33,20	31,09	29,75	28,84	28,16	27,64	27,24	26,92	26,42	25,91	25,39	25,14	24,87	24,60	24,33	24,06	23,79
6	35,51	27,00	23,70	21,92	20,81	20,03	19,46	19,03	18,69	18,41	17,99	17,56	17,12	16,89	16,67	16,44	16,21	15,99	15,75
7	29,25	21,69	18,77	17,19	16,21	15,52	15,02	14,63	14,33	14,08	13,71	13,32	12,93	12,73	12,53	12,33	12,12	11,91	11,70
8	25,42	18,49	15,83	14,39	13,49	12,86	12,40	12,04	11,77	11,54	11,19	10,84	10,48	10,30	10,11	9,92	9,73	9,53	9,33
9	22,86	16,39	13,90	12,56	11,71	11,13	10,70	10,37	10,11	9,89	9,57	9,24	8,90	8,72	8,55	8,37	8,19	8,00	7,81
10	21,04	14,91	12,55	11,28	10,48	9,92	9,52	9,20	8,96	8,75	8,45	8,13	7,80	7,64	7,47	7,30	7,12	6,94	6,76
11	19,69	13,81	11,56	10,35	9,58	9,05	8,66	8,35	8,12	7,92	7,63	7,32	7,01	6,85	6,68	6,52	6,35	6,17	6,00
12	18,64	12,97	10,80	9,63	8,89	8,38	8,00	7,71	7,48	7,29	7,00	6,71	6,40	6,25	6,09	5,93	5,76	5,59	5,42
13	17,81	12,31	10,21	9,07	8,35	7,86	7,49	7,21	6,98	6,80	6,52	6,23	5,93	5,78	5,63	5,47	5,30	5,14	4,97
14	17,14	11,78	9,73	8,62	7,92	7,43	7,08	6,80	6,58	6,40	6,13	5,85	5,56	5,41	5,25	5,10	4,94	4,77	4,60

* Multiplicar por 100.

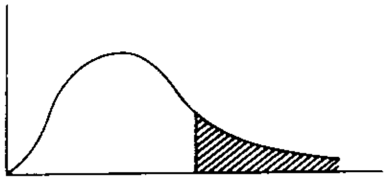
Tabla 4 (Continuación). Distribución *F de FISHER*. *P* [*F* (*m; n*) ≥ *a*] = 0,001



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
15	16,59	11,34	9,34	8,25	7,57	7,09	6,74	6,47	6,26	6,08	5,81	5,54	5,25	5,10	4,95	4,80	4,64	4,47	4,31
16	16,12	10,97	9,00	7,94	7,27	6,81	6,46	6,19	5,98	5,81	5,55	5,27	4,99	4,85	4,70	4,54	4,39	4,23	4,06
17	15,72	10,66	8,73	7,68	7,02	6,56	6,22	5,96	5,75	5,58	5,32	5,05	4,78	4,63	4,48	4,33	4,18	4,02	3,85
18	15,38	10,39	8,49	7,46	6,81	6,35	6,02	5,76	5,56	5,39	5,13	4,87	4,59	4,45	4,30	4,15	4,00	3,84	3,67
19	15,08	10,16	8,28	7,26	6,62	6,18	5,85	5,59	5,39	5,22	4,97	4,70	4,43	4,29	4,14	3,99	3,84	3,68	3,51
20	14,82	9,95	8,10	7,10	6,46	6,02	5,69	5,44	5,24	5,08	4,82	4,56	4,29	4,15	4,00	3,86	3,70	3,54	3,38
21	14,59	9,77	7,94	6,95	6,32	5,88	5,56	5,31	5,11	4,95	4,70	4,44	4,17	4,03	3,88	3,74	3,58	3,42	3,26
22	14,38	9,61	7,80	6,81	6,19	5,76	5,44	5,19	4,99	4,83	4,58	4,33	4,06	3,92	3,78	3,63	3,48	3,32	3,15
23	14,19	9,47	7,67	6,69	6,08	5,65	5,33	5,09	4,89	4,73	4,48	4,23	3,96	3,82	3,68	3,53	3,38	3,22	3,05
24	14,03	9,34	7,55	6,59	5,98	5,55	5,23	4,99	4,80	4,64	4,39	4,14	3,87	3,74	3,59	3,45	3,29	3,14	2,97
25	13,88	9,22	7,45	6,49	5,88	5,46	5,15	4,91	4,71	4,56	4,31	4,06	3,79	3,66	3,52	3,37	3,22	3,06	2,89
26	13,74	9,12	7,36	6,41	5,80	5,38	5,07	4,83	4,64	4,48	4,24	3,99	3,72	3,59	3,44	3,30	3,15	2,99	2,82
27	13,61	9,02	7,27	6,33	5,73	5,31	5,00	4,76	4,57	4,41	4,17	3,92	3,66	3,52	3,38	3,23	3,08	2,92	2,75
28	13,50	8,93	7,19	6,25	5,66	5,24	4,93	4,69	4,50	4,35	4,11	3,86	3,60	3,46	3,32	3,18	3,02	2,86	2,69
29	13,39	8,85	7,12	6,19	5,59	5,18	4,87	4,64	4,45	4,29	4,05	3,80	3,54	3,41	3,27	3,12	2,97	2,81	2,64
30	13,29	8,77	7,05	6,12	5,53	5,12	4,82	4,58	4,39	4,24	4,00	3,75	3,49	3,36	3,22	3,07	2,92	2,76	2,59
40	12,61	8,25	6,60	5,70	5,13	4,73	4,44	4,21	4,02	3,87	3,64	3,40	3,15	3,01	2,87	2,73	2,57	2,41	2,23
60	11,97	7,76	6,17	5,31	4,76	4,37	4,09	3,87	3,69	3,54	3,31	3,08	2,83	2,69	2,55	2,41	2,25	2,08	1,89
120	11,38	7,32	5,79	4,95	4,42	4,04	3,77	3,55	3,38	3,24	3,02	2,78	2,53	2,40	2,26	2,11	1,95	1,76	1,54
∞	10,83	6,91	5,42	4,62	4,10	3,74	3,47	3,27	3,10	2,96	2,74	2,51	2,27	2,13	1,99	1,84	1,66	1,45	1,00

* Multiplicar por 100.

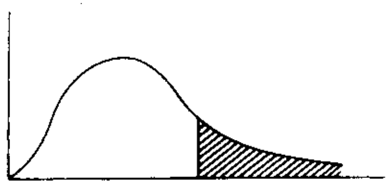
Tabla 4 (Continuación). Distribución *F de FISHER*. *P* [*F* (*m; n*) ≥ *a*] = 0,005



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	16211,00	20000,00	21615,00	22500,00	23056,00	23437,00	23715,00	23925,00	24091,00	24224,00	24426,00	24630,00	24836,00	24940,00	25044,00	25148,00	25253,00	25359,00	25465,00
2	198,50	199,00	199,20	199,20	199,30	199,30	199,40	199,40	199,40	199,40	199,40	199,40	199,40	199,50	199,50	199,50	199,50	199,50	199,50
3	55,55	49,80	47,47	46,19	45,39	44,84	44,43	44,13	43,88	43,69	43,39	43,08	42,78	42,62	42,47	42,31	42,15	41,99	41,83
4	31,33	26,28	24,26	23,15	22,46	21,97	21,62	21,35	21,14	20,97	20,70	20,44	20,17	20,03	19,89	19,75	19,61	19,47	19,32
5	22,78	18,31	16,53	15,56	14,94	14,51	14,20	13,96	13,77	13,62	13,38	13,15	12,90	12,78	12,66	12,53	12,40	12,27	12,14
6	18,63	14,54	12,92	12,03	11,46	11,07	10,79	10,57	10,39	10,25	10,03	9,81	9,59	9,47	9,36	9,24	9,12	9,00	8,88
7	16,24	12,40	10,88	10,05	9,52	9,16	8,89	8,68	8,51	8,38	8,18	7,97	7,75	7,65	7,53	7,42	7,31	7,19	7,08
8	14,69	11,04	9,60	8,81	8,30	7,95	7,69	7,50	7,34	7,21	7,01	6,81	6,61	6,50	6,40	6,29	6,18	6,06	5,95
9	13,61	10,11	8,72	7,96	7,47	7,13	6,88	6,69	6,54	6,42	6,23	6,03	5,83	5,73	5,62	5,52	5,41	5,30	5,19
10	12,83	9,43	8,08	7,34	6,87	6,54	6,30	6,12	5,97	5,85	5,66	5,47	5,27	5,17	5,07	4,97	4,86	4,75	4,64
11	12,23	8,91	7,60	6,88	6,42	6,10	5,86	5,68	5,54	5,42	5,24	5,05	4,86	4,76	4,65	4,55	4,44	4,34	4,23
12	11,75	8,51	7,23	6,52	6,07	5,76	5,52	5,35	5,20	5,09	4,91	4,72	4,53	4,43	4,33	4,23	4,12	4,01	3,90
13	11,37	8,19	6,93	6,23	5,79	5,48	5,25	5,08	4,94	4,82	4,64	4,46	4,27	4,17	4,07	3,97	3,87	3,76	3,65
14	11,06	7,92	6,68	6,00	5,56	5,26	5,03	4,86	4,72	4,60	4,43	4,25	4,06	3,96	3,86	3,76	3,66	3,55	3,44

* Multiplicar por 100.

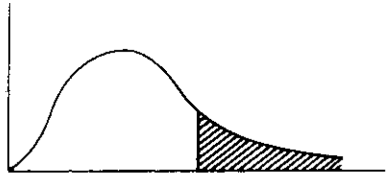
Tabla 4 (Continuación). Distribución *F de FISHER*. $P\ [F\ (m;\ n) \geq a] = 0{,}005$



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
15	10,80	7,70	6,48	5,80	5,37	5,07	4,85	4,67	4,54	4,42	4,25	4,07	3,88	3,79	3,69	3,58	3,48	3,37	3,26
16	10,58	7,51	6,30	5,64	5,21	4,91	4,69	4,52	4,38	4,27	4,10	3,92	3,73	3,64	3,54	3,44	3,33	3,22	3,11
17	10,38	7,35	6,16	5,50	5,07	4,78	4,56	4,39	4,25	4,14	3,97	3,79	3,61	3,51	3,41	3,31	3,21	3,10	2,98
18	10,22	7,21	6,03	5,37	4,96	4,66	4,44	4,28	4,14	4,03	3,86	3,68	3,50	3,40	3,30	3,20	3,10	2,99	2,87
19	10,07	7,09	5,92	5,27	4,85	4,56	4,34	4,18	4,04	3,93	3,76	3,59	3,40	3,31	3,21	3,11	3,00	2,89	2,78
20	9,94	6,99	5,82	5,17	4,76	4,47	4,26	4,09	3,96	3,85	3,68	3,50	3,32	3,22	3,12	3,02	2,92	2,81	2,69
21	9,83	6,89	5,73	5,09	4,68	4,39	4,18	4,01	3,88	3,77	3,60	3,43	3,24	3,15	3,05	2,95	2,84	2,73	2,61
22	9,73	6,81	5,65	5,02	4,61	4,32	4,11	3,94	3,81	3,70	3,54	3,36	3,18	3,08	2,98	2,88	2,77	2,66	2,55
23	9,63	6,73	5,58	4,95	4,54	4,26	4,05	3,88	3,75	3,64	3,47	3,30	3,12	3,02	2,92	2,82	2,71	2,60	2,48
24	9,55	6,66	5,52	4,89	4,49	4,20	3,99	3,83	3,69	3,59	3,42	3,25	3,06	2,97	2,87	2,77	2,66	2,55	2,43
25	9,48	6,60	5,46	4,84	4,43	4,15	3,94	3,78	3,64	3,54	3,37	3,20	3,01	2,92	2,82	2,72	2,61	2,50	2,38
26	9,41	6,54	5,41	4,79	4,38	4,10	3,89	3,73	3,60	3,49	3,33	3,15	2,97	2,87	2,77	2,67	2,56	2,45	2,33
27	9,34	6,49	5,36	4,74	4,34	4,06	3,85	3,69	3,56	3,45	3,28	3,11	2,93	2,83	2,73	2,63	2,52	2,41	2,29
28	9,28	6,44	5,32	4,70	4,30	4,02	3,81	3,65	3,52	3,41	3,25	3,07	2,89	2,79	2,69	2,59	2,48	2,37	2,25
29	9,23	6,40	5,28	4,66	4,26	3,98	3,77	3,61	3,48	3,38	3,21	3,04	2,86	2,76	2,66	2,56	2,45	2,33	2,21
30	9,18	6,35	5,24	4,62	4,23	3,95	3,74	3,58	3,45	3,34	3,18	3,01	2,82	2,73	2,63	2,52	2,42	2,30	2,18
40	8,83	6,07	4,98	4,37	3,99	3,71	3,51	3,35	3,22	3,12	2,95	2,78	2,60	2,50	2,40	2,30	2,18	2,06	1,93
60	8,49	5,79	4,73	4,14	3,76	3,49	3,29	3,13	3,01	2,90	2,74	2,57	2,39	2,29	2,19	2,08	1,96	1,83	1,69
120	8,18	5,54	4,50	3,92	3,55	3,28	3,09	2,93	2,81	2,71	2,54	2,37	2,19	2,09	1,98	1,87	1,75	1,61	1,43
∞	7,88	5,30	4,28	3,72	3,35	3,09	2,90	2,74	2,62	2,52	2,36	2,19	2,00	1,90	1,79	1,67	1,53	1,36	1,00

* Multiplicar por 100.

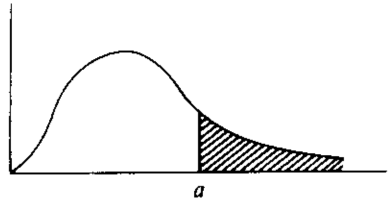
Tabla 4 (Continuación). Distribución *F de FISHER*. $P\,[F\,(m;\,n) \geq a] = 0{,}01$



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052,00	4999,50	5403,00	5625,00	5764,00	5859,00	5928,00	5982,00	6022,00	6056,00	6106,00	6157,00	6209,00	6235,00	6261,00	6287,00	6313,00	6399,00	6366,00
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	7,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,66
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00

* Multiplicar por 100.

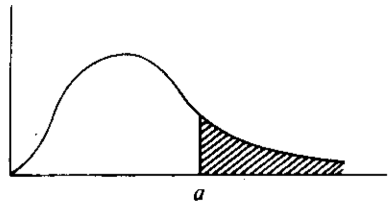
Tabla 4 (Continuación). Distribución *F de FISHER*. $P\,[F(m;\,n) \geq a] = 0{,}01$



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00

* Multiplicar por 100.

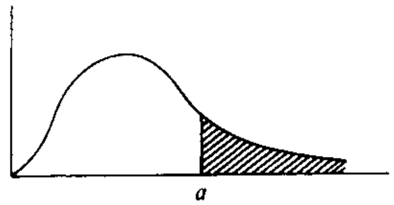
Tabla 4 (Continuación). Distribución *F de FISHER*. $P\,[F\,(m;\,n) \geq a] = 0{,}025$



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647,80	799,50	864,20	899,60	921,80	937,10	948,20	956,70	963,30	968,60	976,70	984,90	993,10	997,20	1001,00	1006,00	1010,00	1014,00	1018,00
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41	39,43	39,45	39,46	39,46	39,47	39,48	39,49	39,50
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42	14,34	14,25	14,17	14,12	14,08	14,04	13,99	13,95	13,90
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,75	8,66	8,56	8,51	8,46	8,41	8,36	8,31	8,26
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,52	6,43	6,33	6,28	6,23	6,18	6,12	6,07	6,02
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,37	5,27	5,17	5,12	5,07	5,01	4,96	4,90	4,85
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,67	4,57	4,47	4,42	4,36	4,31	4,25	4,20	4,14
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,20	4,10	4,00	3,95	3,89	3,84	3,78	3,73	3,67
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,87	3,77	3,67	3,61	3,56	3,51	3,45	3,39	3,33
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,62	3,52	3,42	3,37	3,31	3,26	3,20	3,14	3,08
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,43	3,33	3,23	3,17	3,12	3,06	3,00	2,94	2,88
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,28	3,18	3,07	3,02	2,96	2,91	2,85	2,79	2,72
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,15	3,05	2,95	2,89	2,84	2,78	2,72	2,66	2,60
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	3,05	2,95	2,84	2,79	2,73	2,67	2,61	2,55	2,49

* Multiplicar por 100.

Tabla 4 (Continuación). Distribución *F de FISHER*. $P\,[F\,(m;\,n) \geq a] = 0{,}025$



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,96	2,86	2,76	2,70	2,64	2,59	2,52	2,46	2,40
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,89	2,79	2,68	2,63	2,57	2,51	2,45	2,38	2,32
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,82	2,72	2,62	2,56	2,50	2,44	2,38	2,32	2,25
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,77	2,67	2,56	2,50	2,44	2,38	2,32	2,26	2,19
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,72	2,62	2,51	2,45	2,39	2,33	2,27	2,20	2,13
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,68	2,57	2,46	2,41	2,35	2,29	2,22	2,16	2,09
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73	2,64	2,53	2,42	2,37	2,31	2,25	2,18	2,11	2,04
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70	2,60	2,50	2,39	2,33	2,27	2,21	2,14	2,08	2,00
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67	2,57	2,47	2,36	2,30	2,24	2,18	2,11	2,04	1,97
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,54	2,44	2,33	2,27	2,21	2,15	2,08	2,01	1,94
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,51	2,41	2,30	2,24	2,18	2,12	2,05	1,98	1,91
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59	2,49	2,39	2,28	2,22	2,16	2,09	2,03	1,95	1,88
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	2,57	2,47	2,36	2,25	2,19	2,13	2,07	2,00	1,93	1,85
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55	2,45	2,34	2,23	2,17	2,11	2,05	1,98	1,91	1,83
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	2,53	2,43	2,32	2,21	2,15	2,09	2,03	1,96	1,89	1,81
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,41	2,31	2,20	2,14	2,07	2,01	1,94	1,87	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,29	2,18	2,07	2,01	1,94	1,88	1,80	1,72	1,64
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,17	2,06	1,94	1,88	1,82	1,74	1,67	1,58	1,48
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	2,05	1,94	1,82	1,76	1,69	1,61	1,53	1,43	1,31
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,94	1,83	1,71	1,64	1,57	1,48	1,39	1,27	1,00

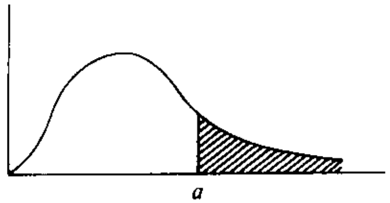
* Multiplicar por 100.

Tabla 4 (Continuación). Distribución *F de FISHER*. $P\,[F\,(m;\,n) \geq a] = 0{,}05$

Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161,40	199,50	215,70	224,60	230,20	234,00	236,80	238,90	240,50	241,90	243,90	245,90	248,00	249,10	250,10	251,10	252,20	253,30	254,30
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13

* Multiplicar por 100.

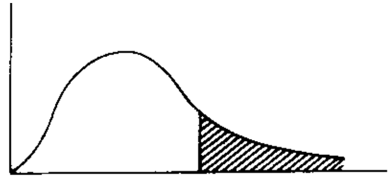
Tabla 4 (Continuación). Distribución *F de FISHER*. *P* [*F* (*m; n*) ≥ *a*] = 0,05



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

* Multiplicar por 100.

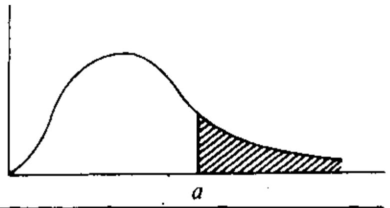
Tabla 4 (Continuación). Distribución *F de FISHER*. *P* [*F* (*m; n*) ≥ *a*] = 0,10



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19	60,71	61,22	61,74	62,00	62,26	62,53	62,79	63,06	63,33
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,41	9,42	9,44	9,45	9,46	9,47	9,47	9,48	9,49
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,20	5,18	5,17	5,17	5,16	5,15	5,14	5,13
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,90	3,87	3,84	3,83	3,82	3,80	3,79	3,78	3,76
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,27	3,24	3,21	3,19	3,17	3,16	3,14	3,12	3,10
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,90	2,87	2,84	2,82	2,80	2,78	2,76	2,74	2,72
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,67	2,63	2,59	2,58	2,56	2,54	2,51	2,49	2,47
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,50	2,46	2,42	2,40	2,38	2,36	2,34	2,32	2,29
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,38	2,34	2,30	2,28	2,25	2,23	2,21	2,18	2,16
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,28	2,24	2,20	2,18	2,16	2,13	2,11	2,08	2,06
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,21	2,17	2,12	2,10	2,08	2,05	2,03	2,00	1,97
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,15	2,10	2,06	2,04	2,01	1,99	1,96	1,93	1,90
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,10	2,05	2,01	1,98	1,96	1,93	1,90	1,88	1,85
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	2,05	2,01	1,96	1,94	1,91	1,89	1,86	1,83	1,80

* Multiplicar por 100.

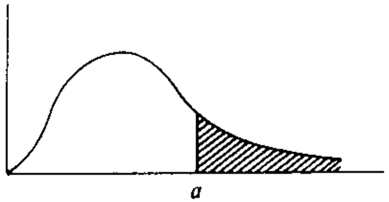
Tabla 4 (Continuación). Distribución *F de FISHER*. *P* [*F* (*m; n*) ≥ *a*] = 0,10



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	2,02	1,97	1,92	1,90	1,87	1,85	1,82	1,79	1,76
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	1,99	1,94	1,89	1,87	1,84	1,81	1,78	1,75	1,72
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	1,96	1,91	1,86	1,84	1,81	1,78	1,75	1,72	1,69
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	1,93	1,89	1,84	1,81	1,78	1,75	1,72	1,69	1,66
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96	1,91	1,86	1,81	1,79	1,76	1,73	1,70	1,67	1,63
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	1,89	1,84	1,79	1,77	1,74	1,71	1,68	1,64	1,61
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92	1,87	1,83	1,78	1,75	1,72	1,69	1,66	1,62	1,59
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	1,86	1,81	1,76	1,73	1,70	1,67	1,64	1,60	1,57
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89	1,84	1,80	1,74	1,72	1,69	1,66	1,62	1,59	1,55
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	1,83	1,78	1,73	1,70	1,67	1,64	1,61	1,57	1,53
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87	1,82	1,77	1,72	1,69	1,66	1,63	1,59	1,56	1,52
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	1,81	1,76	1,71	1,68	1,65	1,61	1,58	1,54	1,50
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85	1,80	1,75	1,70	1,67	1,64	1,60	1,57	1,53	1,49
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	1,79	1,74	1,69	1,66	1,63	1,59	1,56	1,52	1,48
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83	1,78	1,73	1,68	1,65	1,62	1,58	1,55	1,51	1,47
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,77	1,72	1,67	1,64	1,61	1,57	1,54	1,50	1,46
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,71	1,66	1,61	1,57	1,54	1,51	1,47	1,42	1,38
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,66	1,60	1,54	1,51	1,48	1,44	1,40	1,35	1,29
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65	1,60	1,55	1,48	1,45	1,41	1,37	1,32	1,26	1,19
∞	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63	1,60	1,55	1,49	1,42	1,38	1,34	1,30	1,24	1,17	1,00

* Multiplicar por 100.

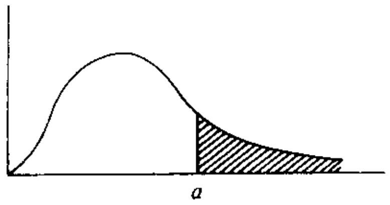
Tabla 4 (Continuación)**.** Distribución *F de FISHER*. *P* [*F* (*m; n*) ≥ *a*] = 0,25



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	5,83	7,50	8,20	8,58	8,82	8,98	9,10	9,19	9,26	9,32	9,41	9,49	9,58	9,63	9,67	9,71	9,76	9,80	9,85
2	2,57	3,00	3,15	3,23	3,28	3,31	3,34	3,35	3,37	3,38	3,39	3,41	3,43	3,43	3,44	3,45	3,46	3,47	3,48
3	2,02	2,28	2,36	2,39	2,41	2,42	2,43	2,44	2,44	2,44	2,45	2,46	2,46	2,46	2,47	2,47	2,47	2,47	2,47
4	1,81	2,00	2,05	2,06	2,07	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08
5	1,69	1,85	1,88	1,89	1,89	1,89	1,89	1,89	1,89	1,89	1,89	1,89	1,88	1,88	1,88	1,88	1,87	1,87	1,87
6	1,62	1,76	1,78	1,79	1,79	1,78	1,78	1,78	1,77	1,77	1,77	1,76	1,76	1,75	1,75	1,75	1,74	1,74	1,74
7	1,57	1,70	1,72	1,72	1,71	1,71	1,70	1,70	1,69	1,69	1,68	1,68	1,67	1,67	1,66	1,66	1,65	1,65	1,65
8	1,54	1,66	1,67	1,66	1,66	1,65	1,64	1,64	1,63	1,63	1,62	1,62	1,61	1,60	1,60	1,59	1,59	1,58	1,58
9	1,51	1,62	1,63	1,63	1,62	1,61	1,60	1,60	1,59	1,59	1,58	1,57	1,56	1,56	1,55	1,54	1,54	1,53	1,53
10	1,49	1,60	1,60	1,59	1,59	1,58	1,57	1,56	1,56	1,55	1,54	1,53	1,52	1,52	1,51	1,51	1,50	1,49	1,48
11	1,47	1,58	1,58	1,57	1,56	1,55	1,54	1,53	1,53	1,52	1,51	1,50	1,49	1,49	1,48	1,47	1,47	1,46	1,45
12	1,46	1,56	1,56	1,55	1,54	1,53	1,52	1,51	1,51	1,50	1,49	1,48	1,47	1,46	1,45	1,45	1,44	1,43	1,42
13	1,45	1,55	1,55	1,53	1,52	1,51	1,50	1,49	1,49	1,48	1,47	1,46	1,45	1,44	1,43	1,42	1,42	1,41	1,40
14	1,44	1,53	1,53	1,52	1,51	1,50	1,49	1,48	1,47	1,46	1,45	1,44	1,43	1,42	1,41	1,41	1,40	1,39	1,38

* Multiplicar por 100.

Tabla 5 (Continuación). Distribución F de FISHER. P [F(m; n) ≥ a] = 0,25



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
15	1,43	1,52	1,52	1,51	1,49	1,48	1,47	1,46	1,46	1,45	1,44	1,43	1,41	1,41	1,40	1,39	1,38	1,37	1,36
16	1,42	1,51	1,51	1,50	1,48	1,47	1,46	1,45	1,44	1,44	1,43	1,41	1,40	1,39	1,38	1,37	1,36	1,35	1,34
17	1,42	1,51	1,50	1,49	1,47	1,46	1,45	1,44	1,43	1,43	1,41	1,40	1,39	1,38	1,37	1,36	1,35	1,34	1,33
18	1,41	1,50	1,49	1,48	1,46	1,45	1,44	1,43	1,42	1,42	1,40	1,39	1,38	1,37	1,36	1,35	1,34	1,33	1,32
19	1,41	1,49	1,49	1,47	1,46	1,44	1,43	1,42	1,41	1,41	1,40	1,38	1,37	1,36	1,35	1,34	1,33	1,32	1,30
20	1,40	1,49	1,48	1,47	1,45	1,44	1,43	1,42	1,41	1,40	1,39	1,37	1,36	1,35	1,34	1,33	1,32	1,31	1,29
21	1,40	1,48	1,48	1,46	1,44	1,43	1,42	1,41	1,40	1,39	1,38	1,37	1,35	1,34	1,33	1,32	1,31	1,30	1,28
22	1,40	1,48	1,47	1,45	1,44	1,42	1,41	1,40	1,39	1,39	1,37	1,36	1,34	1,33	1,32	1,31	1,30	1,29	1,28
23	1,39	1,47	1,47	1,45	1,43	1,42	1,41	1,40	1,39	1,38	1,37	1,35	1,34	1,33	1,32	1,31	1,30	1,28	1,27
24	1,39	1,47	1,46	1,44	1,43	1,41	1,40	1,39	1,38	1,38	1,36	1,35	1,33	1,32	1,31	1,30	1,29	1,28	1,26
25	1,39	1,47	1,46	1,44	1,42	1,41	1,40	1,39	1,38	1,37	1,36	1,34	1,33	1,32	1,31	1,29	1,28	1,27	1,25
26	1,38	1,46	1,45	1,44	1,42	1,41	1,39	1,38	1,37	1,37	1,35	1,34	1,32	1,31	1,30	1,29	1,28	1,26	1,25
27	1,38	1,46	1,45	1,43	1,42	1,40	1,39	1,38	1,37	1,36	1.35	1,33	1,32	1,31	1,30	1,28	1,27	1,26	1,24
28	1,38	1,46	1,45	1,43	1,41	1,40	1,39	1,38	1,37	1,36	1,34	1,33	1,31	1,30	1,29	1,28	1,27	1,25	1,24
29	1,38	1,45	1,45	1,43	1,41	1,40	1,38	1,37	1,36	1,35	1,34	1,32	1,31	1,30	1,29	1,27	1,26	1,25	1,23
30	1,38	1,45	1,44	1,42	1,41	1,39	1,38	1,37	1,36	1,35	1,34	1,32	1,30	1,29	1,28	1,27	1,26	1,24	1,23
40	1,36	1,44	1,42	1,40	1,39	1,37	1,36	1,35	1,34	1,33	1,31	1,30	1,28	1,26	1,25	1,24	1,22	1,21	1,19
60	1,35	1,42	1,41	1,38	1,37	1,35	1,33	1,32	1,31	1,30	1,29	1,27	1,25	1,24	1,22	1,21	1,19	1,17	1,15
120	1,34	1,40	1,39	1,37	1,35	1,33	1,31	1,30	1,29	1,28	1,26	1,24	1,22	1,21	1,19	1,18	1,16	1,13	1,10
∞	1,32	1,39	1,37	1,35	1,33	1,31	1,29	1,28	1,27	1,25	1,24	1,22	1,19	1,18	1,16	1,14	1,12	1,08	1,00

* Multiplicar por 100.

Tabla 5. Probabilidades asociadas con valores tan pequeños como los valores observados de *U* en el test de Mann-Whitney.

$n_2 = 3$

U \ n_1	1	2	3
0	0,250	0,100	0,050
1	0,500	0,200	0,100
2	0,750	0,400	0,200
3		0,600	0,350
4			0,500
5			0,650

$n_2 = 4$

U \ n_1	1	2	3	4
0	0,200	0,067	0,028	0,014
1	0,400	0,133	0,057	0,029
2	0,600	0,267	0,114	0,057
3		0,400	0,200	0,100
4		0,600	0,314	0,171
5			0,429	0,243
6			0,571	0,343
7				0,443
8				0,557

$n_2 = 5$

U \ n_1	1	2	3	4	5
0	0,167	0,047	0,018	0,008	0,004
1	0,333	0,095	0,036	0,016	0,008
2	0,500	0,190	0,071	0,032	0,016
3	0,667	0,286	0,125	0,056	0,028
4		0,429	0,196	0,095	0,048
5		0,571	0,286	0,143	0,075
6			0,393	0,206	0,111
7			0,500	0,278	0,155
8			0,607	0,365	0,210
9				0,452	0,274
10				0,548	0,345
11					0,421
12					0,500
13					0,579

$n_2 = 6$

U \ n_1	1	2	3	4	5	6
0	0,143	0,036	0,012	0,005	0,002	0,001
1	0,286	0,071	0,024	0,010	0,004	0,002
2	0,428	0,143	0,048	0,019	0,009	0,004
3	0,571	0,214	0,083	0,033	0,015	0,008
4		0,321	0,131	0,057	0,026	0,013
5		0,429	0,190	0,086	0,041	0,021
6		0,571	0,274	0,129	0,063	0,032
7			0,357	0,176	0,089	0,047
8			0,452	0,238	0,123	0,066
9			0,548	0,305	0,165	0,090
10				0,381	0,214	0,120
11				0,457	0,268	0,155
12				0,545	0,331	0,197
13					0,396	0,242
14					0,465	0,294
15					0,535	0,350
16						0,409
17						0,469
18						0,531

Fuente: H.B. Mann; D.R. Whitney. "On a test o whether one of two random variables is stochastically larger than the other". *Ann. Math. Stat.* (vol. 18).

Tabla 5 (Continuación). Probabilidades asociadas con valores tan pequeños como los valores observados de *U* en el test de Mann-Whitney.

$n_2 = 7$

U \ n_1	1	2	3	4	5	6	7
0	0,125	0,028	0,008	0,003	0,001	0,001	0,000
1	0,250	0,056	0,017	0,006	0,003	0,001	0,001
2	0,375	0,111	0,033	0,012	0,005	0,002	0,001
3	0,500	0,167	0,058	0,021	0,009	0,004	0,002
4	0,625	0,250	0,092	0,036	0,015	0,007	0,003
5		0,333	0,133	0,055	0,024	0,011	0,006
6		0,444	0,192	0,082	0,037	0,017	0,009
7		0,556	0,258	0,115	0,053	0,026	0,013
8			0,333	0,158	0,074	0,037	0,019
9			0,417	0,206	0,101	0,051	0,027
10			0,500	0,264	0,134	0,069	0,036
11			0,583	0,324	0,172	0,090	0,049
12				0,394	0,216	0,117	0,064
13				0,464	0,265	0,147	0,082
14				0,538	0,319	0,183	0,104
15					0,378	0,223	0,130
16					0,438	0,267	0,159
17					0,500	0,314	0,191
18					0,562	0,365	0,228
19						0,418	0,267
20						0,473	0,310
21						0,527	0,355
22							0,402
23							0,451
24							0,500
25							0,549

Fuente: H.B. Mann; D.R. Whitney. "On a test o whether one of two random variables is stochastically larger than the other". *Ann. Math. Stat.* (vol. 18).

Tabla 5 (Continuación). Probabilidades asociadas con valores tan pequeños como los valores observados de *U* en el test de Mann-Whitney.

	$n_2 = 8$									
U \ n_1	1	2	3	4	5	6	7	8	t	*Normal*
0	0,111	0,022	0,006	0,002	0,001	0,000	0,000	0,000	3,308	0,001
1	0,222	0,044	0,012	0,004	0,002	0,001	0,000	0,000	3,203	0,001
2	0,333	0,089	0,024	0,008	0,003	0,001	0,001	0,000	3,098	0,001
3	0,444	0,133	0,042	0,014	0,005	0,002	0,001	0,001	2,993	0,001
4	0,556	0,200	0,067	0,024	0,009	0,004	0,002	0,001	2,888	0,002
5		0,267	0,097	0,036	0,015	0,006	0,003	0,001	2,783	0,003
6		0,356	0,139	0,055	0,023	0,010	0,005	0,002	2,678	0,004
7		0,444	0,188	0,077	0,033	0,015	0,007	0,003	2,573	0,005
8		0,556	0,248	0,107	0,047	0,021	0,010	0,005	2,468	0,007
9			0,315	0,141	0,064	0,030	0,014	0,007	2,363	0,009
10			0,387	0,184	0,085	0,041	0,020	0,010	2,258	0,012
11			0,461	0,230	0,111	0,054	0,027	0,014	2,153	0,016
12			0,539	0,285	0,142	0,071	0,036	0,019	2,048	0,020
13				0,341	0,177	0,091	0,047	0,025	1,943	0,026
14				0,404	0,217	0,114	0,060	0,032	1,838	0,033
15				0,467	0,262	0,141	0,076	0,041	1,733	0,041
16				0,533	0,311	0,172	0,095	0,052	1,628	0,052
17					0,362	0,207	0,116	0,065	1,523	0,064
18					0,416	0,245	0,140	0,080	1,418	0,078
19					0,472	0,286	0,168	0,097	1,313	0,094
20					0,528	0,331	0,198	0,117	1,208	0,113
21						0,377	0,232	0,139	1,102	0,135
22						0,426	0,268	0,164	0,998	0,159
23						0,475	0,306	0,191	0,893	0,185
24						0,525	0,347	0,221	0,788	0,215
25							0,389	0,253	0,683	0,247
26							0,433	0,287	0,578	0,282
27							0,478	0,323	0,473	0,318
28							0,522	0,360	0,368	0,356
29								0,399	0,263	0,396
30								0,439	0,158	0,437
31								0,480	0,052	0,481
32								0,520		

Fuente: H.B. Mann; D.R. Whitney. "On a test o whether one of two random variables is stochastically larger than the other". *Ann. Math. Stat.* (vol. 18).

Tabla 6. Valores críticos de *T*. Prueba de Wilcoxon

Tamaño de la muestra, *n*	Prueba de una cola 0,05	Prueba de una cola 0,01	Prueba de dos colas 0,05	Prueba de dos colas 0,01
5	1			
6	2		1	
7	4	0	2	
8	6	2	4	0
9	8	3	6	2
10	11	5	8	3
11	14	7	11	5
12	17	10	14	7
13	21	13	17	10
14	26	16	21	13
15	30	20	25	16
16	36	24	30	19
17	41	28	35	23
18	47	33	40	28
19	54	38	46	32
20	60	43	52	37
21	68	49	59	43
22	75	56	66	49
23	83	62	73	55
24	92	69	81	68
25	101	77	90	68
26	110	85	98	76
27	120	93	107	84
28	130	102	117	92
29	141	111	127	100
30	152	120	137	109