

JAIME FABIÁN VERA CHANG

Fundamentos de Bio estadística

aplicada a las ciencias agroalimentarias





**Fundamentos de Bio estadística
aplicada a las ciencias agroalimentarias**

Autor:

Jaime Fabián Vera Chang

Fundamentos de Bio estadística
aplicada a las ciencias agroalimentarias

Autor.
Jaime Fabián Vera Chang



Primera edición: noviembre 2018

© Universidad Técnica Estatal de Quevedo 2018

© Ediciones Grupo Compás 2018

ISBN: 978-9942-33-085-7

Diseño de portada y diagramación: Grupo
Compás

Este texto ha sido sometido a un proceso de
evaluación por pares externos con base en la
normativa del editorial.

Quedan rigurosamente prohibidas, bajo las
sanciones en las leyes, la producción o
almacenamiento total o parcial de la presente
publicación, incluyendo el diseño de la portada,
así como la transmisión de la misma por
cualquiera de sus medios, tanto si es electrónico,
como químico, mecánico, óptico, de grabación
o bien de fotocopia, sin la autorización de los
titulares del copyright.

Guayaquil-Ecuador 2018

Cita.

Vera, J. (2018) Fundamentos de Bio estadística aplicada a las ciencias agroalimentarias,
Editorial Grupo Compás, Guayaquil Ecuador, 155 pag

PRÓLOGO

Este libro es el resultado de experiencias en concursos en los galardones nacionales organizados por la senescyt en Ecuador, además de lecturas, apuntes, capacitaciones además de las experiencias del ejercicio de la cátedra de Estadística I del rediseño de carrera de Ingeniería en Alimentos, participaciones en congresos, eventos científicos, arbitraje de libros, y revistas científicas aproximadamente de una década de labores del profesor Jaime Fabián Vera Chang docente – Investigador de la Universidad Técnica Estatal de Quevedo (UTEQ). Ex Investigador de la Unidad de Investigación Científica y Tecnológica, División Agrícola, Área cacao.

La razón por la cual decidí escribir esta obra, en primer lugar, fue mi familia y mis amigos a quienes escuché, consideré sus consejos y empecé esta noble labor el poder aportar con una guía básica sobre el uso de la estadística aplicada, mi interés fue contar un texto guía para mis estudiantes de Ingeniería en Alimentos, Agropecuaria y carreras afines. Como también a académicos de otras instituciones, quienes me visitaban para que les ayude a estructurar su proyecto de investigación, en lo que respecta a como poder elaborar y presentar los resultados obtenidos, productos de los análisis experimentales, durante mi condición como miembro del Comité de Investigación Formativa (2012-2019), agradezco el aporte de mis amigos: José Aguilar Reyes, Kléber Estupiñan, Wilfrido Escobar Pavón, Diana Vasco Mora y Gregorio Vásquez Montufar, muy agradecido por colaboración y valiosos consejos en las revisiones de los proyectos de investigación entregados para dar directrices de cómo mejorar el rigor y la eficiencia de los trabajos de los aspirantes de Ingeniería.

Cuando empezamos la labor como comité era notable que existieran muchos vacíos sobre cómo aplicar la estadística, y los estudiantes se acercaban a preguntar en las tareas integradoras sobre que herramienta estadística era la más adecuada y como debían organizar los datos, y esto me llevó a reflexionar sobre cómo podríamos agradecer a nuestra institución, hasta que después de muchas horas y diálogos de mi padre el Ing. Jaime Set Vera Barahona me estimuló con cartas nemotécnicas permitiendo el inicio en horas de la madrugada de la construcción de esta guía básica la cual duró tres años de labor invisible.

Esta segunda edición contiene los fundamentos teóricos-prácticos la cual está dividida en ocho capítulos en contiene una información compilada cuyo objetivo fundamental recordar algunos conceptos básicos introductorios sobre la importancia de la estadística, técnicas estadísticas, principios de estadística los tipos, las gráficas de datos mediante un histogramas, medidas de tendencia central, y variabilidad, probabilidad, hipótesis, pruebas no paramétricas más empleadas y análisis de regresión y correlación con ejercicios resueltos, y muestra los cálculos matemáticos paso a paso de manera que se demuestra cómo se realizan los análisis desde el computador o con el uso de una calculadora, como también su interpretación y como se puede ejecutar un análisis y presentar resultados en proyectos, artículos, etc.

En esta segunda edición como todo trabajo humano está sujeta a mejoras por lo cual mi intención principal es contribuir a la Academia Latinoamericana, incluyendo fundamentos para entender las materias de titulación como Metodología de la Investigación y Diseño experimental.

Quiero agradecer primeramente a Dios, al rector El Dr. Eduardo Díaz Ocampo, a los estudiantes y compañeros docentes, a la decana Facultad de Ciencias Pecuarias- UTEQ, Ing. Jenny Torres Navarrete por su amistad y apoyo incondicional en beneficio de la Institución (periodo 2015-2019) como también al Director de Investigación Dr. Byron Oviedo Bayas por su constante apoyo y responsabilidad en el quehacer investigativo, y especialmente a la editorial Compas como contraparte editora, por su paciencia y dedicación para llevarla a cabo la publicación de esta obra.

Finalmente mis agradecimientos a Diana Casco, mi esposa y a mi hija Bianca Vera cuya paciencia para soportarme sentado durante horas frente al computador, me sirvieron de estímulo para seguir adelante y mejorar este segundo texto.

DEDICATORIA

A **DIOS** quien es mi fortaleza
mis padres, hijos, sobrinos, amigos
por ellos vivo, investigo y
trabajo

Homenaje póstumo a la
memoria de mi madre:

Blanca Chang Malats

A mis maestros universitarios:

Manuel Haz Álvarez

Anthony Álvarez Tubay

Aurelio Haón Jama

Alfredo Olvera Díaz

***Para Bianca y Diana
mis amores***

A las autoridades de la UTEQ:

Dr. Eduardo Díaz Ocampo

Ing. Roberto Pico Saltos

Ing. Guadalupe Murillo C.

TABLA DE CONTENIDO

CAPITULO I	7
INTRODUCCIÓN.....	8
1.1. CONCEPTOS GENERALES	8
1.1.1. RESEÑA HISTÓRICA	8
1.2. POR QUÉ ES IMPORTANTE LA ÉTICA Y LA ESTADÍSTICA	9
1.3. APLICACIÓN DE LA ESTADÍSTICA	10
1.4. CONCEPTOS BÁSICOS	10
1.4.1. LA ESTADÍSTICA COMO PARTE DEL METODO CIENTÍFICO.....	13
1.5. POBLACIÓN Y MUESTRA	15
1.6. VARIABLES	17
1.6.1. Otras clasificaciones	19
1.7. REDONDEO DE DATOS.....	20
1.8. NOTACIÓN SISTEMÁTICA	22
1.9. USO DE LA COMPUTADORA PARA PROCESAR DATOS	23
1.9.1. VENTAJAS Y DESVENTAJAS DE USAR UNA COMPUTADORA.....	23
1.9.2. RECOMENDACIONES.....	24
CAPITULO II.....	25
2.0 GRÁFICAS ESTADÍSTICAS USADAS.....	26
2.1. DISTRIBUCIÓN DE FRECUENCIAS	34
2.1.1. INTERVALOS DE CLASE.....	36
2.1.2. TAMAÑO O ANCHURA DE UN INTERVALO DE CLASE.....	36
HISTOGRAMA Y POLIGONOS DE FRECUENCIA	37
2.1.3. DISTRIBUCIÓN DE FRECUENCIAS RELATIVA	37
2.1.4. DISTRIBUCIÓN DE FRECUENCIAS ACUMULADA.....	38
2.1.5. DISTRIBUCIÓN DE FRECUENCIAS RELATIVAS Y ACUMULADAS.....	38
2.1.6. GRAFICAS APLICADAS AL DESARROLLO DE LA TEORÍA DE FRECUENCIAS	38
2.1.7. USANDO EXCEL E INFOSTAT.....	40
CAPITULO III.....	61
3.0. MEDIDAS DE TENDENCIA O POSICIÓN CENTRAL.....	62
3.1. NOTACIÓN MATEMÁTICA NECESARIA.....	62
3.2. LA MEDIA ARITMETICA	64
3.3. OTROS TIPOS DE MEDIA	65
LA MEDIA GEOMETRICA $M_g = M_o$	65
3.4. LA MEDIA PONDERADA.....	66
3.4.1. Propiedades de la media.....	67
3.5. MODO = MODA = VALOR MODAL: M_d	68
3.6. MEDIANA M_e	69
.....	72
3.7. RELACIÓN EMPIRICA ENTRE MEDIA, MEDIANA Y MODA	
ANÁLISIS DE DATOS	80
.....	84
CAPITULO IV	85
4.0 MEDIDAS O ESTADIGRAFOS DE DISPERSIÓN	85
4.1. DISPERSIÓN O VARIACIÓN	85

.....	85
4.2. RANGO O AMPLITUD	86
4.3. LA VARIANZA	88
4.4. DESVIACIÓN TÍPICA (Raíz de la varianza)	89
4.5. DESVIACIÓN MEDIA O PROMEDIO DE DESVIACIÓN	91
4.6 COEFICIENTE DE VARIACIÓN	
46.1. EJERCICIO PROPUESTOS:.....	92
USANDO EXCEL E INFOSTAT.....	93
<i>CAPITULO V</i>	97
5.0 MEDIDAS DE POSICIÓN NO CENTRALES.....	98
5.1. USANDO INFOSTAT	101
<i>CAPITULO VI</i>	105
6.0 PROBABILIDAD DE OCURRENCIA DE UN SUCESO	106
6.1. PRINCIPIOS DE PROBABILIDAD	106
6.1.1. ESPACIO MUÉSTRALES	106
6.1.2. EXPERIMENTO ALEATORIO	107
6.1.3. LA DISTRIBUCIÓN T DE STUDENT.....	107
Cálculo de probabilidades	111
<i>CAPITULO VII</i>	114
7.0. PRUEBA DE HIPÓTESIS.	115
<i>CAPITULO VIII</i>	125
<i>PRUEBAS NO PARAMÉTRICAS</i>	126
8.1 INTERVALOS DE CONFIANZA	126
8.2 PRUEBA DE HIPOTESIS	127
8.3. PRUEBA DE CHI CUADRADO	130
8.4. PRUEBA DE "t" STUDENT	131
8.5. KRUSKAL WALLIS PARA K MUESTRAS	135
<i>CAPITULO IX</i>	137
9.0 REGRESIÓN Y CORRELACIÓN	138
9.2 REGRESIÓN LINEAL SIMPLE.....	139
TIPOS REGRESIONES Y SUS RESPECTIVOS MODELOS:.....	140
9.3. COEFICIENTE DE CORRELACIÓN LINEAL SIMPLE	140
9.4. USANDO EL EXCEL E INFOSTAT	147
<i>Bibliografía</i>	154
<i>Tablas Estadísticas</i>	155

INTRODUCCIÓN

INTRODUCCIÓN

La estadística es un conjunto de técnicas para la colección, manejo, descripción y análisis de información, de manera que las conclusiones obtenidas de ella, tengan un rango de confiabilidad especificado entonces, esta ciencia junto con los Diseños Experimentales, son herramientas fundamentales para realizar investigación científica y en cualquier área de la ingeniería y en nuestro caso, para solucionar los problemas agro-alimentarios.

En el día ocurren sucesos en el ámbito económico, social, político, biológico, y educacional los que están presentes en un momento dado, lo que da a pensar que tan valiosa información si no se registra se pierde, por otra parte, el registro de estos datos nos puede ayudar el comportamiento de estos fenómenos y saber que sucede o que puede suceder usando métodos y técnicas.

La estadística o métodos estadísticos desempeñan un importante papel protagónico en casi todas las facetas del progreso humano. Anteriormente solo era aplicada a los asuntos del estado, de donde viene su nombre; pero ahora la influencia de la estadística se extiende a la agricultura, biología, negocios, química, física, electrónica, mecánica, comunicaciones, economía, enseñanza, medicina, ciencias políticas, psicología, y demás campos de la ciencia e ingeniería.

Por otras personas es un método para obtener, presentar y escribir grandes cantidades de datos indispensables para la toma de decisiones. El objetivo básico de este fascículo es aclarar los conceptos básicos utilizados con frecuencia y analizar las aplicaciones del concepto de "Estadística".

La palabra tiene su sentido más amplio para aquellas personas cuyo trabajo requiere de estos conceptos y técnicas juegan un papel importante en las actividades que realizan los profesionales agropecuarios, alimentos y todas las ciencias aplicadas.

1.1. CONCEPTOS GENERALES

1.1.1. RESEÑA HISTÓRICA

La palabra estadística, para algunos proviene de la palabra *status* cuyo significado es estado; para otros se deriva del vocablo griego *statera* que significa balanza, por otro lado en palabra alemana *stara* que significa estado, encargado de hacer estadísticas

La estadística se fundamenta sobre datos, es tan antigua como la propia civilización humana, ya que existe vestigio o evidencia fue aplicada desde antes de la venida de Cristo, en China el emperador Yao (2.238 a. C.) dispuso de la realización de un censo para determinar la situación de sus habitantes.

La estadística se fundamentada por el londinense Jhon Graunt (1604 - 1661) un mercader que redactó un pequeño libro "Natural and political Observations made upon the Bills of Mortality", esta obra trató de interpretar los fenómenos biológicos a partir de datos numéricos, treinta años más tarde formó parte de la Royal Society y publicó sobre su trabajo de mortalidad "Philosophical Transactions".

1.2. POR QUÉ ES IMPORTANTE LA ÉTICA Y LA ESTADÍSTICA

Porque los datos estadísticos y las conclusiones obtenidas aplicando metodología estadística ejercen una profunda influencia en casi todos los campos de la actividad humana. En particular, la estadística invade cada vez más cualquier investigación relativa a las ciencias Agro-Alimentarias. Este crecimiento, probablemente relacionado con el interés por aumentar la credibilidad y confiabilidad de las investigaciones, no garantiza que en todos los casos la metodología estadística haya sido correctamente utilizada, o peor aún, que sea válida, por lo cual es un asunto ético, es decir por ser una herramienta del investigador es importante no alterar la información original, por muchos errores es importante tener disciplina científica, planificar, organizar, defender la verdad de los hechos o fenómenos medidos.

Los conceptos y temas de la estadística se utilizan en la actualidad en un gran número de ocupaciones, las técnicas estadísticas constituyen una parte integral de las actividades de investigación en distintas áreas del conocimiento.

La persona que comprenda los conceptos estadísticos y su metodología obtendrá el insumo necesario para entender e interpretar estadística sobre todos sus campos de acción generando un eficiente desempeño profesional.

Con frecuencia escuchamos en los medios de comunicación local noticias que mencionan lo siguiente:

Se ha demostrado un incremento porcentual de las ventas de automóviles en el año.

La explotación de petróleo en el Ecuador Específicamente para el año 2014 nuestra meta es pasar de 325.000 (barriles diarios de promedio en 2013) a 327.000 barriles diarios"

Estadísticas demuestran que el consumo de huevo de gallina aumenta el colesterol en los mientras que el de codorniz tiene menor debido al contenido de omega 3.

Se ha comprobado estadísticamente que la pasta dental de mayor aceptación del público es la marca Colgate.

El CENSO Agropecuario tiene información de las distintas especies del sector agrícola y pecuario que se desarrollan en el Ecuador. Cuenta con catorce capítulos que recogen información relacionada a la superficie plantada y cosechada, ganado, aves y datos de 26 productos agrícolas específicos cuyos resultados son presentados en desagregados provinciales, regionales y nacionales.

1.3. APLICACIÓN DE LA ESTADÍSTICA

En el día ocurren sucesos en el ámbito económico, social, político, biológico, y educacional los que están presentes en un momento dado, lo que da a pensar que tan valiosa información si no se registra se pierde, por otra parte, el registro de estos datos nos puede ayudar el comportamiento de estos fenómenos y saber que sucede o que puede suceder usando métodos y técnicas.

La estadística o métodos estadísticos está despeñando un importante papel ascendente en casi todas las facetas del progreso humano. Anteriormente solo era aplicada a los asuntos del estado, de donde viene su nombre; pero ahora la influencia de la estadística se extiende a la agricultura, biología, negocios, química, física, electrónica, mecánica, comunicaciones, economía, enseñanza, medicina, ciencias políticas, psicología, y demás campos de la ciencia e ingeniería.

1.4. CONCEPTOS BÁSICOS

La estadística es una ciencia relativamente moderna, de veloz formación, que crece cada día, a medida que haya más respuesta los problemas que poseen los investigadores. En la aplicación de la estadística, los principios son siempre generales, aunque algunas veces las técnicas pueden diferir.

Buscar una definición de estadística es siempre difícil por su amplio alcance, pero como punto de partida podrían ser razonables y satisfactorias cualquiera de las siguientes definiciones:

El término “estadística” también se emplea para designar un área de un estudio: una disciplina: Como área de estudio la estadística proporciona los métodos que ayudan a resolver problemas correspondientes, para ello es necesario conocer conceptos básicos.

Según Steel y Torrie “la estadística es la ciencia pura y aplicada encargada de crear, desarrollar y aplicar técnica tal que las incertidumbres de inferencias inductivas pueden evaluarse”.

La estadística es parte del método científico que tiene por objeto, la recolección, la organización, el análisis, la interpretación y la presentación de datos.

La palabra *estadística* se refiere a un sistema o método usado en la recolección, organización, análisis, y descripción numérica de la información (Martínez), también se puede decir que la Estadística estudia el comportamiento de los fenómenos de grupo.

La estadística está ligada en los métodos científicos en la toma, organización, recopilación, presentación, análisis de datos, tanto para la deducción de conclusiones, análisis de datos, tanto para la deducción de conclusiones como para tomar decisiones razonables de acuerdo al análisis.

En un sentido más estricto, el término se realiza para denotar los mismos datos o números que se derivan de ellos, como por ejemplo los promedios para la muestra (\bar{x}) y (μ) para designar la media poblacional.

La estadística es la ciencia que se encarga de la recopilación, organización, presentación, análisis e interpretación de datos numéricos con el fin de tomar decisiones. Los métodos tratan de la presentación y resumen de datos, estimación de parámetros poblacionales, prueba de hipótesis con relación a los parámetros poblacionales, determinación de la exactitud de las estimaciones, estudio de variación, correlación y diseño de experimentos, la misma se clasifica o comprende técnicas descriptivas e inferenciales.

DESCRIPTIVA: La estadística Descriptiva es el estudio que incluye la obtención, organización, presentación y descripción de la información numérica.

INFERENCIAL: La inferencia estadística es una técnica mediante la cual se obtienen generalizaciones o se toman decisiones en base a una información parcial o completa obtenida mediante técnica descriptiva.

La inferencia hace referencia a un conjunto de métodos que permiten hacer predicciones acerca de características de un fenómeno sobre la base de información parcial acerca del mismo. Los métodos de la inferencia nos permiten proponer el valor de una cantidad desconocida (estimación) o decidir entre dos teorías contrapuestas cuál de ellas explica mejor los datos observados (test de hipótesis).

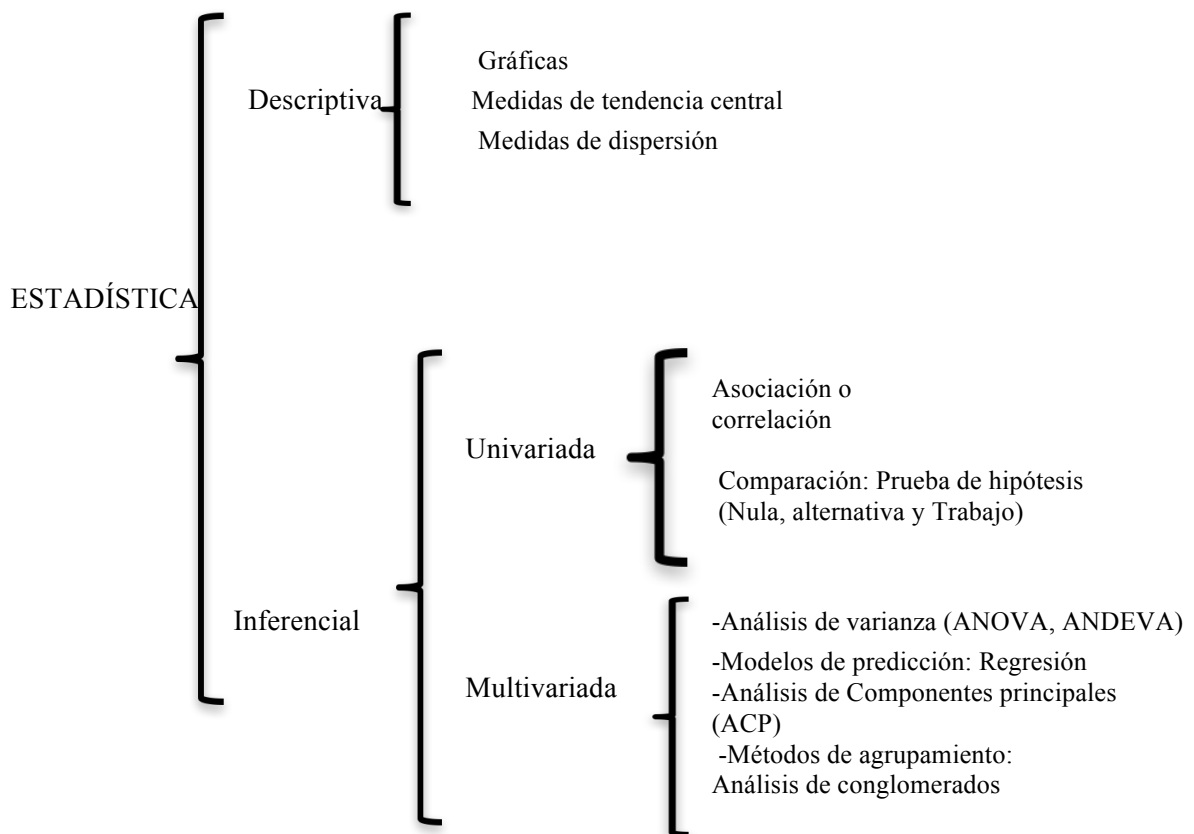
Para concluir diremos que existe otra gran división de las técnicas estadísticas:

a) Estadística Paramétrica.

b) Estadística No Paramétrica.

Los métodos paramétricos permiten hacer inferencias acerca de parámetros poblacionales de las distribuciones.

Los métodos no paramétricos es un conjunto de técnicas diseñadas para niveles de medición menores por ejemplo el nominal y ordinal, para efectuar estimaciones no habrá parámetros en estricto sentido, por lo cual existen varias ventajas para el investigador y analista de datos.



1.4.1. LA ESTADÍSTICA COMO PARTE DEL METODO CIENTÍFICO

La metodología estadística se emplea en muchos campos, se conoce que la estadística ayuda a diseñar y registrar información para describir y analizar con facilidad y mediante estimaciones, obtener conclusiones que enriquecen el conocimiento de todo el fenómeno que ocurren a nuestro alrededor, es decir cada día va ganando terreno en su aplicación en programas gubernamentales etc.

Ha quedado señalado como orienta la estadística al investigador en la toma de decisiones con respecto a sus puntos de vista, también es evidente que la estadística no es el conocimiento mismo, sino una ciencia que crea métodos para lograr conocimiento, los cuales forman parte método científico.

Describir el método científico general no es tarea fácil, pues no existe como tal; es decir no hay un método que siga un conjunto de pasos ordenados, sin embargo, por razones de exposición, es conveniente admitir pasos que puedan producirse continuamente en la solución lógica de los problemas y que forman el método científico general.

Observaciones. _ La ciencia comienza con la observación de determinadas partes de la naturaleza que son relevantes al conocimiento o solución de un problema. La observación conlleva a la solución y descripción, debiendo adoptar definiciones precisas de modo que una palabra específica sea entendida con el mismo significado y en forma cuantitativa.

Hipótesis. _ Para explicar los hechos observados, el científico debe traducir sus ideas en una hipótesis o "teoría tentativa" sobre la naturaleza y relación de las observaciones; y mientras la teoría no sea probada, no deberá confundirse con una ley. Muchas veces esta etapa comienza durante la observación.

Predicción._ La concepción de una hipótesis aplicada a los fenómenos observados hace posible la utilización de la lógica y permite deducir varias consecuencias. Si la teoría es satisfactoria, estas deducciones constituyen un nuevo conocimiento y toman el nombre de predicciones, NO EN SENTIDO DE AUGURIOS, sino más bien en el de pronósticos anticipando lo que se verá, si ciertas observaciones, (aún no producidas).

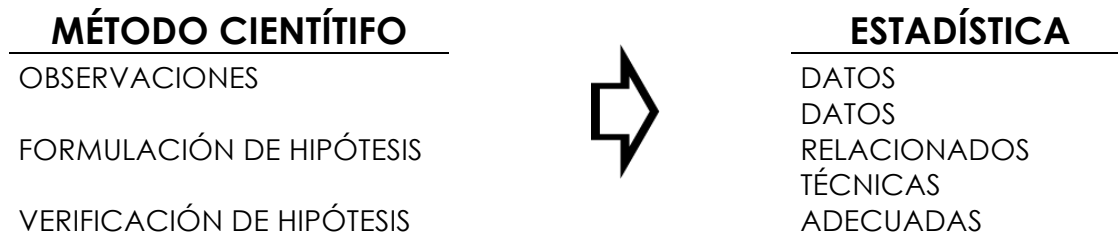
Verificación y experimentación._ En muchos casos las hipótesis son tan simples y sus consecuencias tan obvias, que hace posible probarlas directamente. Para esto el investigador recolecta nuevos hechos que le permiten probar las predicciones formuladas en base a la teoría. En esta etapa el círculo se cierra y el ciclo empieza nuevamente.

La verificación puede hacerse realizando un experimento que permita probar la hipótesis. En un experimento el observador puede inferir en algún grado con la naturaleza y crear condiciones favorables a sus propósitos

ETAPAS:

- a) Etapa de planeamiento
- b) Etapas de recolección
- c) Etapa de elaboración de Datos
- d) Etapa de presentación de datos
- e) Etapa de análisis e interpretación de datos
- f) Etapas de Publicación de Datos.

RELACIÓN DE LA BIOESTADÍSTICA CON EL MÉTODO CIENTÍFICO

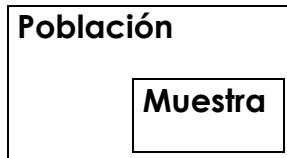


1.5. POBLACIÓN Y MUESTRA

Población es cualquier conjunto formado por elementos que pueden ser inequívocamente identificados y que además poseen uno o más atributos factibles de ser medidos. Por ejemplo, el conjunto de todas las vacas lecheras de una misma raza es una **población**, ya que los elementos de este conjunto son fácilmente identificables el uno del otro y además, poseen características medibles u observables, tales como, el peso, la edad, el nivel de producción, etc. Pero estudiar una característica cualquiera usando para ello todos los elementos de una población resulta, en la inmensa mayoría de los casos, una tarea imposible, y es por ello que en la práctica, solo sea factible analizar una parte de la misma. A cualquier subconjunto de elementos extraídos de una determinada población se le llama **Muestra**.

En una colección de datos que tienen características de un grupo de individuos u objetos, tales como número de porciones de maderita fina, producción de leche en una región, peso de los estudiantes de una universidad, porcentaje de testa en almendras de cacao almacenado, determinación de enfermedades en una población de cerdos, cantidad arroz consumido en una ciudad, es a menudo imposible o poco probable obtener los datos en su totalidad, sobre todo si la población es grande. En lugar de examinar el grupo entero llamado **población o universo**, se examina una pequeña parte del grupo llamado muestra.

Una muestra es un conjunto de medidas u observaciones tomadas a partir de una población dada; es un subconjunto de la población. Desde luego, el número de observaciones de una muestra será la población de la misma, las muestras se toman debido a que no es factible desde el punto de vista económico usar a toda la población.



Una población puede ser *finita* o *infinita*. Por ejemplo, la población consistente en todos los laureles (***Cordia alliodora***) que crecen en el cantón Quevedo, es finita; es decir el número de árboles es limitado; mientras que la población formada por todos los posibles sucesos como lanzamientos sucesivos de una moneda es infinita, sin embargo algunas veces, en una población finita con un número grande de elementos, el comportamiento de cálculo de la muestra resulta igual al de una población infinita, ya que existe un tamaño muestral óptimo, lo que quiere decir que no se altera por más que aumenten el tamaño poblacional.

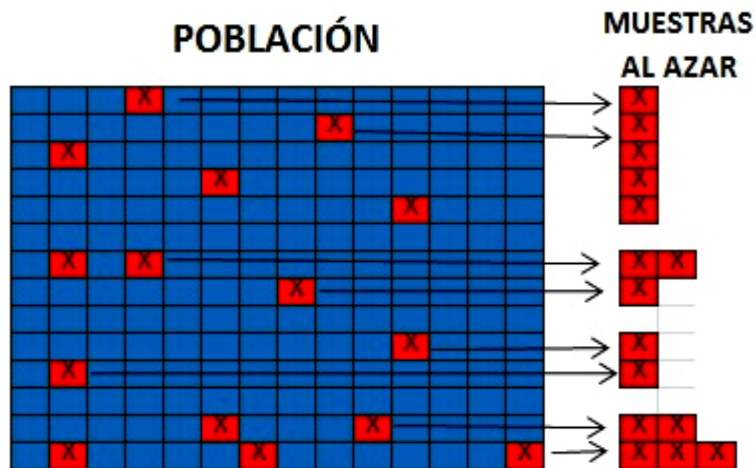


Figura 1. Esquemas de una muestra aleatoria.

Si una muestra es representativa de una población, se pueden deducir importantes conclusiones acerca de esta, a partir del análisis de la misma. La parte de la estadística que trata de las condiciones bajo las cuales tales inferencias son válidas se llama estadística inductiva o estadística inferencial. Al no poder estar absolutamente seguros de la veracidad de tales inferencias, se ha de utilizar con frecuencia conclusiones. La parte de la estadística que trata solamente de describir y analizar un conjunto de datos sin sacar conclusiones o inferencias de un grupo mayor se llama estadística descriptiva o deductiva.

Por lo tanto hay dos fases en el campo de la Estadística. En primer lugar está la fase que solo se limita a la descripción de una serie de datos sin llegar a conclusiones generalizadas de un grupo mayoritario. Esta fase se la conoce como Estadística deductiva o descriptiva. En un segundo lugar está la fase de análisis que trata de llegar a conclusiones acerca de un grupo mayor basado en la información de un grupo menor o muestra: es esta la Estadística inductiva o de inferencia.

Usamos una muestra para conocer o estimar características de la población, denominamos:

PARÁMETRO \Rightarrow una medida resumen calculada sobre la población

ESTADÍSTICO \Rightarrow una medida resumen calculada sobre la muestra

1.6. VARIABLES

Una variable es una característica observable o medible de un objeto en estudio que puede adoptar diferentes valores o expresarse en varias propiedades. Como también una característica observable ligada o una relación determinada a otros aspectos, estas relaciones son aspectos básicos del método científico ya que se fundamenta en el hecho de medir, analizar y concluir sobre un problema específico.

Es un símbolo tal como X , Y , H , X , que puede tomar un valor cualquiera de un conjunto determinado, llamado **dominio** de la variable. Si la variable puede tomar solamente un valor se llama **constante**.

Una variable que teóricamente puede tomar cualquier valor entre dos valores dados se llama **variable continua**, si no es así se llama **variable discreta**:

a) Discretos. La variable sólo puede tomar un cierto conjunto de valores posibles. En

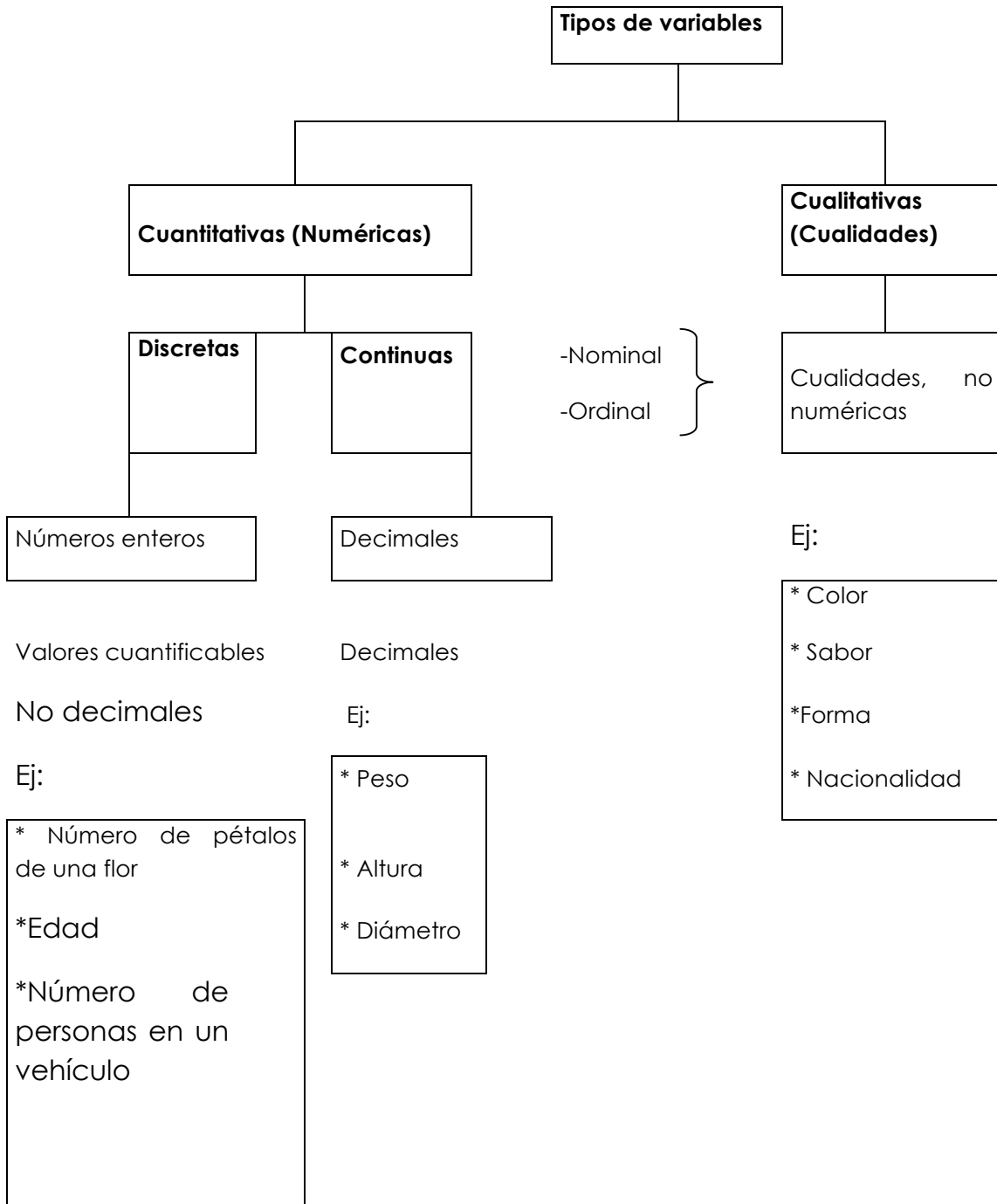
general, aparecen por conteo.

Ejemplo: número de miembros del hogar, número de intervenciones quirúrgicas,

número de casos notificados de una cierta patología.

b) Continuos. Generalmente son el resultado de una medición que se expresa en unidades. Las mediciones pueden tomar teóricamente un conjunto infinito de valores posibles dentro de un rango. En la práctica

los valores posibles de la variable están limitados por la precisión del método de medición o por el modo de registro.



CATEGORÍAS NOMINALES ⇒ No existe orden obvio entre las categorías.

Ejemplos: país de origen, estado civil, diagnóstico.

CATEGORÍAS ORDINALES ⇒ Existe un orden natural entre las categorías.

Ejemplos 1 variable discreta:

En una familia el número de hijos puede tomar cualquiera de los valores 0,1,2,3..... pero no puede ser 1.7, 2.5, 3.6...

En un parto en cuyes la camada es de 1,2,3,4,5..... pero no puede ser 1.4, 2.2, 3.7....

Numero de huevos en gallinas de postura en series cortas es de 2,3,4,5 .. pero no puede ser 1.3, 2.6, 3.6...

Numero de brotes de cedro (*Cedrela Odorata* L.)in vitroes de 2, 3, 4, 5 pero no puede ser 2.5, 3.8, 4.7.....

Número mínimo de muestras elementales en un lote de cacao seco almacenados en bodegas es de 2,3,4,7..... pero no puede ser 2.4, 3.7, 4.7, 6.9.....

Ejemplos 2 variable continua:

La altura de un individuo puede ser de 63.5 o 65. 843 pulgadas

La gravedad específica de la leche fresca es de 1.027, 1.030, 1.035

El peso seco de nódulos en raíces inoculadas con *Bradyrhizibiuma* los 25 días es de 0.5, 0.8, 1.2, 1.7 g.....

El porcentaje de proteína del rastrojo de maíz es de 4.7, 5.1, 5.5, 5.8,6.0 %

1.6.1. Otras clasificaciones

Existe otro tipo de clasificaciones de las variables las cuales se describen a continuación

1.6.1.1. Variables dependientes vs Independiente y la relación de la estadística con el método científico.

Las variables independientes son aquellos que son manipulados mientras que las variables dependientes se miden sólo o registradas. Esta distinción aparece terminológicamente confuso para muchos porque, ya que algunos estudiantes dicen: "todas las variables dependen de algo." Sin embargo, una vez que te acostumbras a esta distinción, se hace indispensable. Los términos variable dependiente e independiente se aplican principalmente a la investigación experimental en el que se manipulan algunas variables, y en este sentido son "independiente" de

los patrones iniciales de reacción, características, intenciones, etc de los sujetos. Se espera que algunos otras variables para ser "dependientes" de la manipulación o de las condiciones experimentales. Es decir, dependen de "lo que el sujeto va a hacer" como respuesta. Algo contrario a la naturaleza de esta distinción, estos términos también se utilizan en los estudios en los que no manipulamos literalmente variables independientes, pero sólo asignan los sujetos a los "grupos experimentales" sobre la base de algunas de las propiedades pre-existentes de los sujetos. Por ejemplo, si en un experimento, los machos se comparan a las mujeres con respecto a su conteo de células blancas (WCC), género que se podría llamar la variable independiente y la WCC la variable dependiente.

Variables dependientes (Y): reciben este nombre de las variables a ser explicadas, es decir el objeto de la investigación, la cual trata de explicar en función de otros elementos.

Variables independientes (X): son las variables explicativas, es decir los factores o elementos susceptibles de explicar las variables en (Y), en una investigación experimental son las variables que se conducen.

Variables intermedias o intervinientes: en algunos casos en el análisis de relación causa – efecto. Se introducen una o más variables de enlace interpretativo entre las variables dependientes o independientes.

Variables explicatorias: son las propiedades que interesan directamente al investigador en términos de su modelo.

Variables externas: son las que están fuera del interés teórico inmediato y que pueden afectar los resultados de una investigación empírica. Que una variable sea considerada explicatorias o externa depende de la investigación en particular.

Los datos que vienen definido por un conteo o enumeraciones se llaman variable discreta, mientras que los datos que llevan decimales se llaman continuos.

1.7. REDONDEO DE DATOS

Dentro de los procesos estadísticos hay ocasiones en la que los datos suministrados para el análisis y sobre todo aquellos que son resultados de un proceso matemático, en los que obtenemos datos con valoraciones decimales, lo más común es aplicar esta técnica, que consiste en aproximar lo más posible a un valor más cercano o uno que sea más fácil para trabajar las operaciones matemáticas, por ejemplo, si el valor

obtenido o recolectado es 72.3, se tiene la tendencia de redondear la cantidad a al valor más cercano, en este caso sería el 72 y se tiende a despreciar la tendencia decimal, por lo contrario si el valor obtenido es 72.8 se redondea hacia el valor más cercano que es el 73, en la práctica de estadística es preferible trabajar con los valores suministrados u obtenidos en un cálculo de tal manera que se evite el error de redondeos acumulados.

El resultado de redondear 72.8 al entero más próximo es 73 puesto que 72.8 está más cerca de 73 que de 72.

El redondeo de 72.465 a un decimal con aproximación de centésimas nos encontramos con el dilema que 72.465 está a la mitad del recorrido entre 72.46 y 72.47.

Se acostumbra en este caso redondear al número par más próximo que antecede al 5. Así; 72.465 será 72.5

Problemas Resueltos

Redondear cada uno de los siguientes números a la exactitud que se indica.

48.7	aproximando unidades	=	49.0
136.5	aproximando unidades	=	137.0
2.484	aproximando centésimas	=	2.48
0.0435	aproximando milésimas	=	0.044
450.001	aproximando unidades	=	450.0
143,95	aproximando unidades	=	144.0
399,579	aproximando centésimas	=	400.68
24,448	aproximando unidades	=	24,5
556.500	aproximando centésimas	=	556.50
199.501	aproximando unidad	=	200.0

Sumar los números 4,35; 8,65; 2,95; 12,45; 6,65; 7,55; 9,75 **a)** directamente, **b)** redondeando de acuerdo al criterio del “par más próximo” **c)**, redondeando todos los datos por exceso

	a) 4,35	b) 4,4	c) 4,4
	8,65	8,6	8,7
	2,95	3,0	3,0
	12,45	12,4	12,5
	6,65	6,6	6,7
	7,55	7,6	7,6
	9,75	9,8	9,8
Total	<hr/> 52,35	Total <hr/> 52,4	Total <hr/> 52,7

El método **(b)** es más exacto que el **(c)** puesto que la acumulación de error redondeo se minimiza en el método **(b)**.

1.8. NOTACIÓN SISTEMÁTICA

Son las diferentes formas de escribir algunas de las cantidades que implican anotaciones amplias, como son: notación logarítmica, notación del sistema binario, notación del sistema de cualquier base, notación científica, notación factorial, entre otros.

Se usa en aquellos casos con números con varios ceros antes o después del punto decimal, para su notación, usando la potencia 10.

Se usa en aquellos casos con números con muchos ceros antes o después del punto decimal, para su notación, usando la potencia 10.

Ejemplo 1:

$$10^1 = 10,$$

$$10^2 = 10 \times 10 = 100$$

$$10^5 = 10 \times 10 \times 10 \times 10 \times 10 = 100,000$$

$$10^8 = 100'000\ 000$$

Ejemplo 2:

$10^0 = 1$; $10^{-1} = 0,1$; $10^{-2} = 0,01$; $10^{-5} = 0,00001$

Ejemplo 3:

$864'000\ 000 = 8,64 \times 10^8$; $0,00003416 = 3,416 \times 10^{-5}$

También es utilidad el ejemplo de las reglas

$(10^p)(10^q) = 10^{p+q}$; $10^p/10^q = 10^{p-q}$

En 10^p , P= exponente

10= base

1.9. USO DE LA COMPUTADORA PARA PROCESAR DATOS

Las computadoras nos ahorran los aspectos tediosos del análisis estadístico y en principio producen cálculos correctos, pero no garantizan que obtendremos resultados válidos y correctos. Consideraremos brevemente las ventajas y desventajas de usar una computadora para procesar datos y consideraremos algunas formas de armar archivos de datos.

1.9.1. VENTAJAS Y DESVENTAJAS DE USAR UNA COMPUTADORA

a) Ventajas

- Exactitud y velocidad. Cuando el software es de calidad se obtienen resultados correctos rápidamente.
- Versatilidad. Se tiene acceso a un amplio rango de técnicas estadísticas. Muchas más de las que es posible describir en cualquier curso de estadística.
- Gráficos. Se pueden producir representaciones de los datos originales o de los resultados obtenidos que permiten una mejor visualización.
- Flexibilidad. Una vez que se ha construido la base de datos, se pueden realizar

pequeños cambios y repetir el análisis. Por ejemplo, es posible excluir algunos casos, hacer análisis por subgrupos o estratos, etc.

- Nuevas variables. Es simple generar nuevas variables. Ejemplo: diferencia entre mediciones antes y después de un tratamiento, calcular edad como diferencia de fechas,

crear variables categóricas a partir de variables numéricas, recategorizar variables cualitativas, realizar transformaciones, etc.

- Volumen de datos. Algunos programas pueden procesar un número de registros o de variables ilimitado.

b) Desventajas

- Errores en el software. Muchos paquetes estadísticos de uso corriente presentan errores en algunos procedimientos. Los más seguros son: EXCEL, SAS, S-PLUS, STATA, SPSS e INFOSTAT. Si no se tiene seguridad acerca de la calidad del software que se está usando debería chequearse comparando los resultados de cada procedimiento con ejemplos de libro o con software de primer nivel.

- Versatilidad. Esta ventaja se transforma en desventaja porque al haber tantos métodos estadísticos disponibles es fácil usar uno inapropiado. Es importante que el usuario tenga en claro sus limitaciones en conocimientos estadísticos y use sólo los métodos

que comprende. Si el problema parece requerir métodos que no son familiares, es aconsejable consultar a un estadístico.

1.9.2. RECOMENDACIONES

Antes de realizar cualquier análisis estadístico en cualquier software es importante verificar que los datos estén bien tabulados es transcendental revisar por segunda ocasión para estar seguro, por experiencia este es el error más común del investigador.

GRÁFICAS ESTADÍSTICAS USADAS

2.0 GRÁFICAS ESTADÍSTICAS USADAS

DIAGRAMA CIRCULAR

Es de especial utilidad para mostrar proporciones (porcentajes) relativas de una variable. Se crea marcando una porción del círculo correspondiente a cada categoría de la variable. Ejemplo Productos industrializados en Ecuador de enero a Marzo 2016.

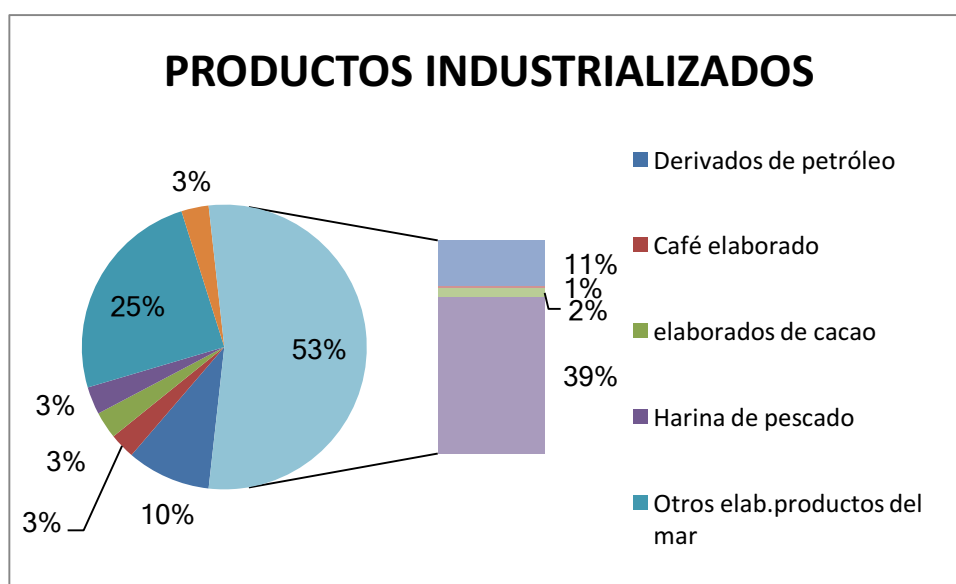
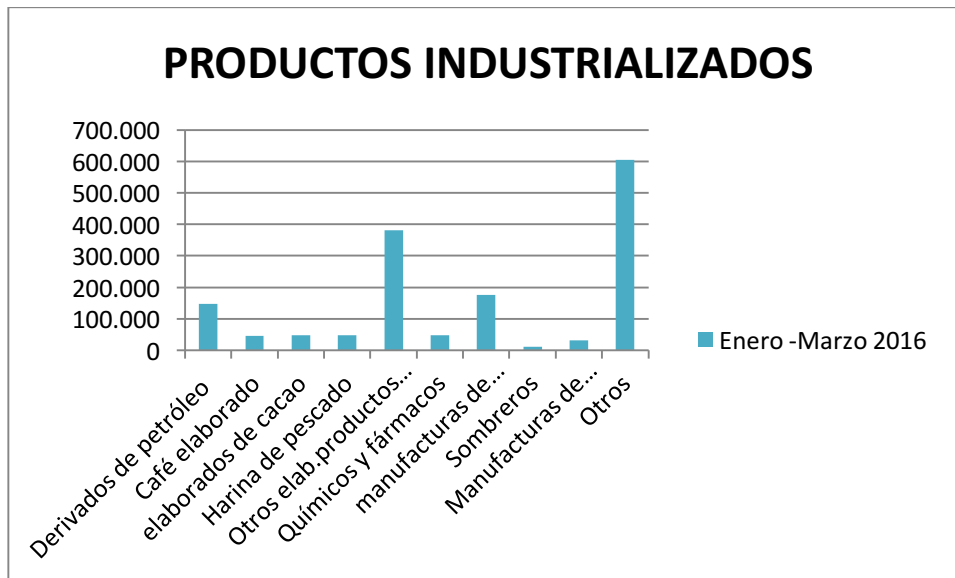


DIAGRAMA DE BARRAS

Es una forma gráfica de representar datos cualitativos que se han resumido en una distribución de frecuencias, de relativas o de porcentuales. Hay varios tipos de gráficos de barras, como son :

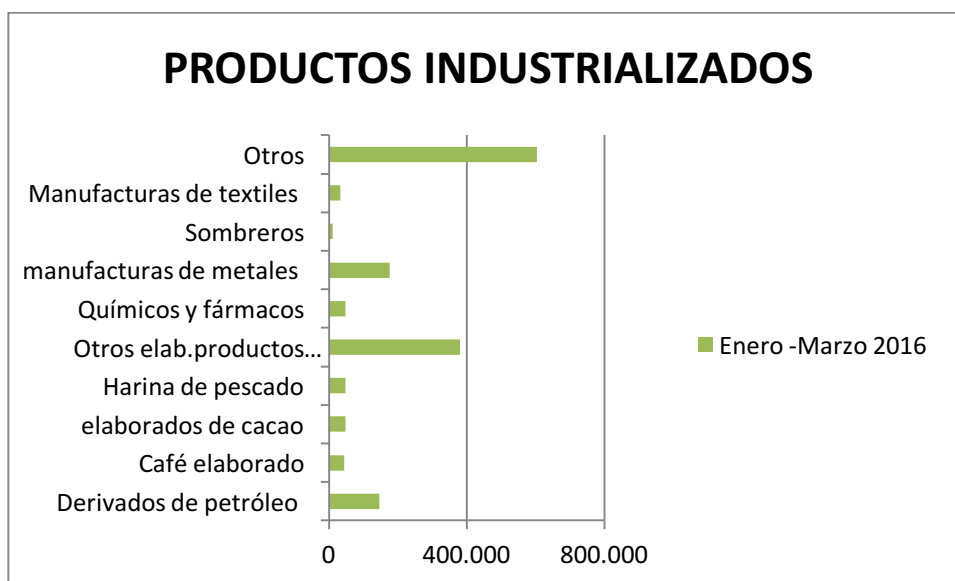
GRÁFICA SIMPLE DE BARRAS VERTICALES

Para respuestas categóricas cualitativas en el que solo interviene una barra para cada clase. Su trazo se realiza ubicando en el eje horizontal de la gráfica los nombres que identifican cada una de las clases. En el eje vertical se usa una escala de frecuencias, una de frecuencias relativas o una de porcentuales. Luego, con una barra de un ancho fijo trazada sobre cada indicador de clase llegamos a la altura que corresponde al tipo de frecuencia escogido. Las barras se separan a fin de señalar que cada clase es una categoría independiente. Los espacios entre las barras deben corresponder a la mitad del ancho de una barra.



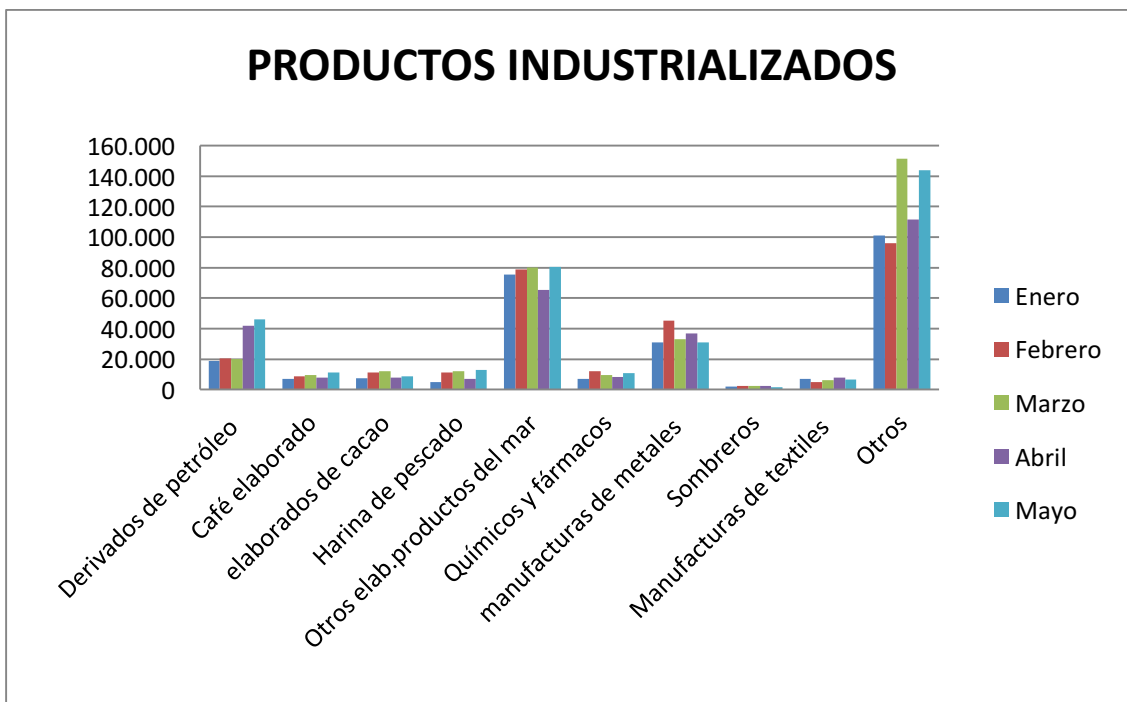
GRÁFICA SIMPLE DE BARRAS HORIZONTALES

Se utiliza principalmente para facilitar la comparación entre las diferentes clases que componen los datos categóricos. El trazo de la gráfica es muy similar a la gráfica de barras verticales, solo que éstas van en forma. De esta manera se logra una mejor visualización en las preferencias.



GRÁFICA DE BARRAS COMPONENTES

Este tipo de gráfica se usa cuando las diferentes categorías de datos se componen de otras clases, de tal forma que cada barra se pueda subdividir y representar cada una de estas clases. Así mismo, entre las categorías y sus componentes se compara valores. También se le conoce como barras agrupadas. Se puede hacer uso de barras horizontales o de barras verticales; su escogencia depende de lo que se pretenda ilustrar para que facilite su visualización.



GRÁFICA DE BARRAS SECCIONADAS

Esta gráfica compara entre categorías el aporte de cada valor al total, dando lugar a una columna apilada para cada clase. También se puede presentar de manera horizontal o vertical.

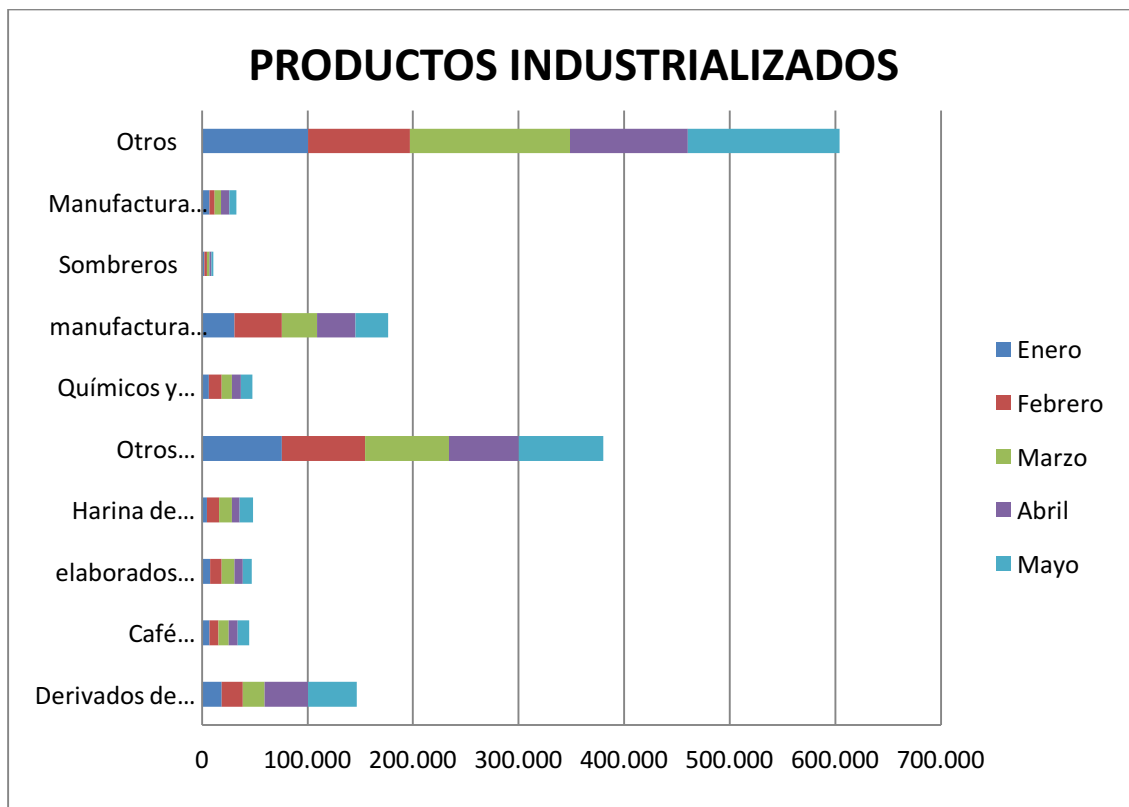


DIAGRAMA DE PARETO

Es un tipo especial de diagrama de barras verticales, donde las respuestas categóricas se grafican en orden descendente de frecuencias y se combinan con un polígono acumulado en la misma escala. El diagrama de Pareto se usa ampliamente en el control estadístico de procesos y el control estadístico de la calidad del producto.

Lo que se pretende con este tipo de gráfico es describir en donde se presenta el mayor porcentaje del problema y que factores lo afectan. Este concepto, se conoce como la regla de 80-20, considera que el 80% de la actividad se debe al 20% de los factores. Al concentrarse en el 20% de los factores, los gerentes pueden atacar el 80% del problema.

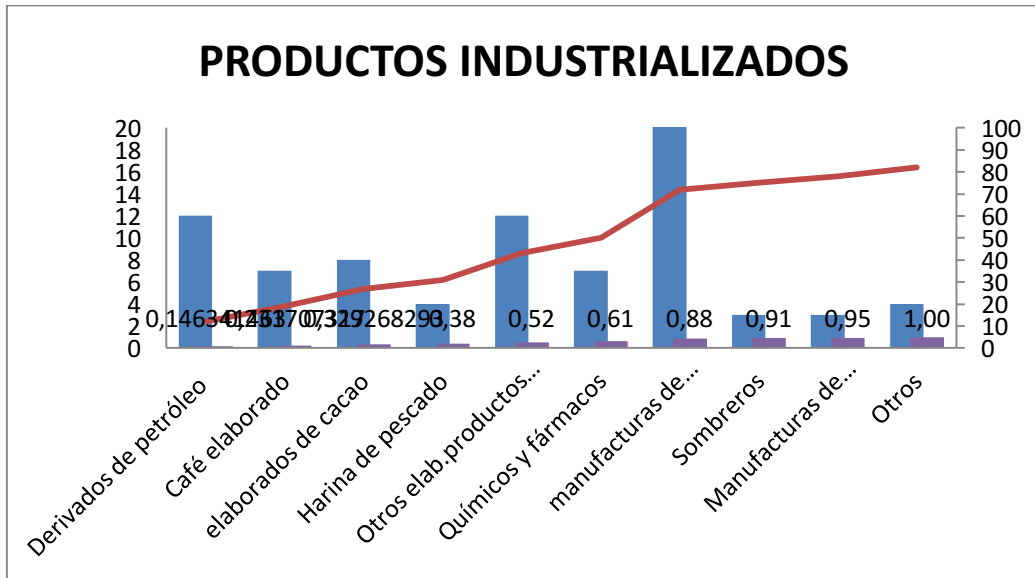
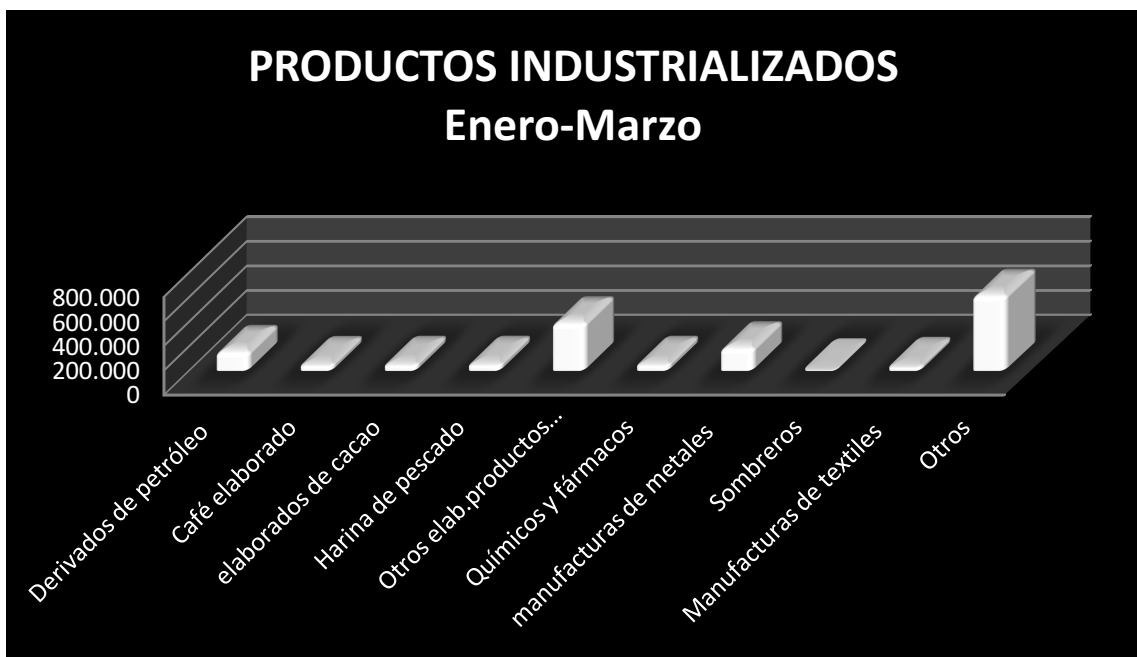


DIAGRAMA DE BARRAS

Tienen el mismo uso que los datos categóricos, solamente que intervienen dos variables, una que representa el tiempo y la otra cantidad (Ingresos, ventas, IPC, Costos, No. De unidades producidas, etc.). Dependiendo de lo que se quiera representar se ofrecen los diagramas de barras simples, de componentes, bidireccionales y seccionados.



GRÁFICAS DE LINEA

Se ilustra mediante segmentos de línea los cambios en cantidades con respecto al tiempo. Son especialmente útiles en el comercio y en los negocios.

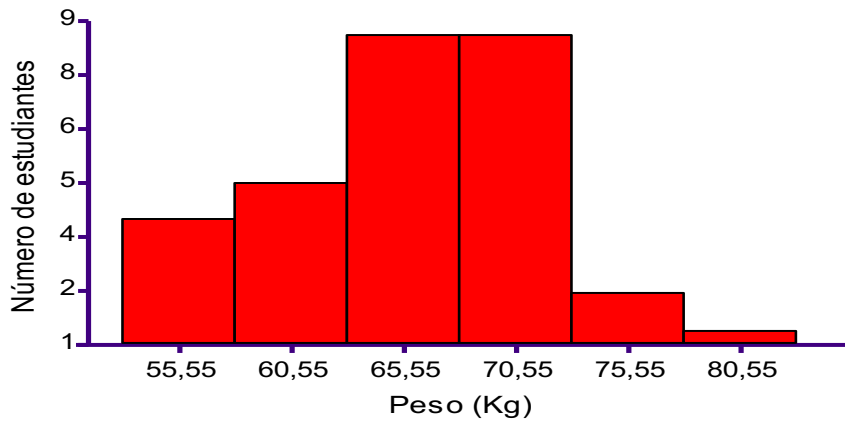


DATOS NUMERICOS

HISTOGRAMAS

Una de las maneras más comunes de representar una distribución de frecuencia. Su gráfica consiste en un conjunto de barras, en la que la base de cada barra representa una clase o intervalo, indicada en el eje horizontal, y la altura por su frecuencia, indicada en el eje vertical. Generalmente las barras se trazan adyacentes una a la otra.

HISOGRAMA DE FRECUENCIAS



POLÍGONO DE FRECUENCIA

De segmentos de línea que conectan los puntos formados por la intersección del punto medio de clase y la frecuencia de clase absoluta, relativa o porcentual.



OJIVA

Es un polígono acumulado de frecuencia absoluta ,relativa o porcentual y por lo tanto representa segmentos de línea que se origina al conectar los puntos formados por la intersección entre el límite inferior

de cada clase con la frecuencia acumulada. Es conocida como POLÍGONO DE FRECUENCIA ACUMULADA MENOR QUE, ya que muestra el número o porcentaje de observaciones menores a cierto valor. La ojiva es importante por que nos permite extrapolar información que la distribución de frecuencia oculta y así como calcular estadísticos como la mediana, cuartiles, deciles y percentiles, en forma aproximada. Para construir la ojiva se debe primero elaborar la distribución de frecuencia menor que.

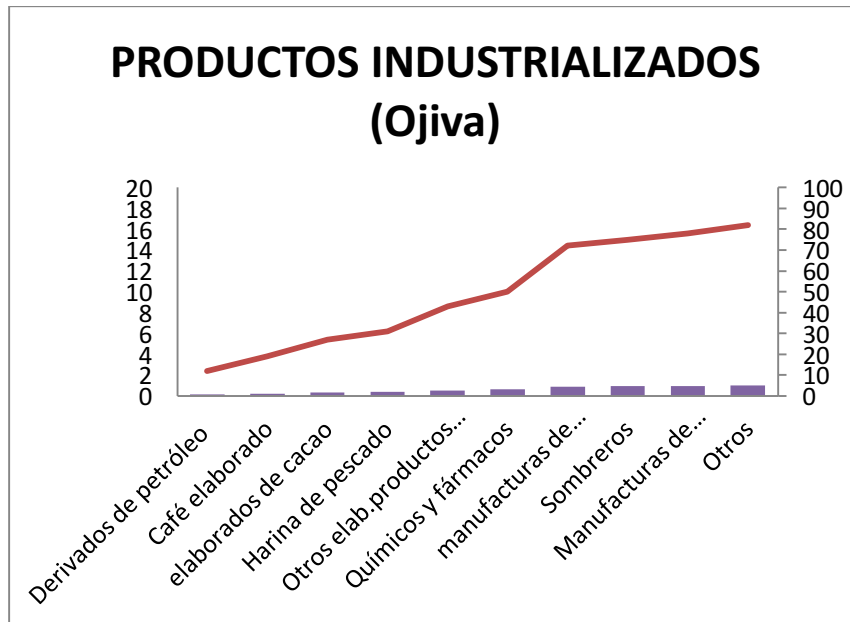


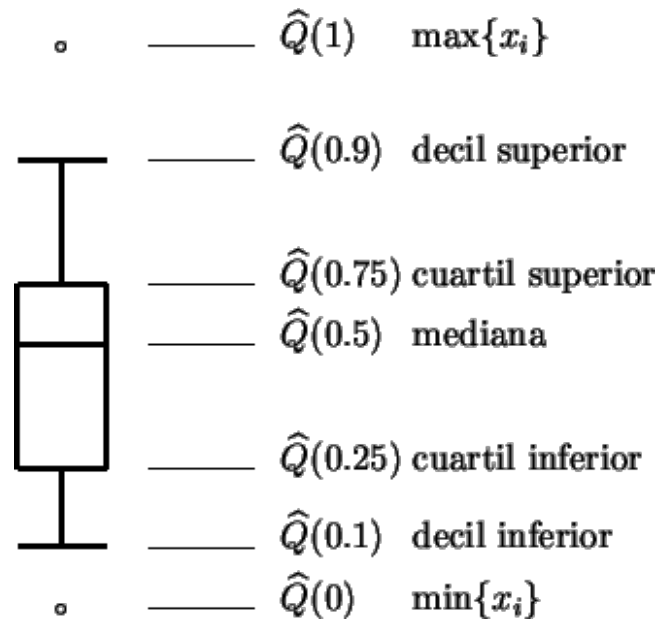
DIAGRAMA DE CAJA

Es una representación gráfica, basada en cinco números estadísticos: valor mínimo, Q1, la mediana, Q3 y valor máximo. Se utiliza como una técnica de ANÁLISIS EXPLORATORIO DE DATOS y tiene la ventaja de que no se requiere de la desviación estándar, ni de la media aritmética y así como de resumir los datos en una distribución de frecuencia, situación que si necesita el histograma, el polígono y la ojiva. Para su trazo, se lleva a cabo

los siguientes pasos:

1. Crear una escala apropiada a lo largo del eje horizontal
2. Dibujar una caja entre el Q1 y el Q3

3. Dentro de la caja trazamos una línea recta vertical que representa la mediana
4. Finalmente, trazamos líneas horizontales de la caja hasta el valor mínimo y de la caja hasta el valor máximo. A estas líneas horizontales fuera de la caja se les conoce como “bigotes”



2.1. DISTRIBUCIÓN DE FRECUENCIAS

La distribución de frecuencia es un método utilizado para organizar y resumir la información. Indicando en número de veces que se repite. Estos pueden presentarse en cuadros o en forma gráfica.

Es el resultado de ordenar una serie de valores obtenidos al medir una variable. La distribución de frecuencia es un cuadro de datos que básicamente se compone de dos columnas; en la primera aparecen los valores de la variable y en la segunda las frecuencias, es decir el número de veces que se ha presentado cada puntuación.

Si el número total de datos obtenidos es suficientemente grande, se puede simplificar los datos, con una distribución de frecuencia agrupadas en intervalos. También se utilizan las frecuencias acumuladas en las que las frecuencias de cada intervalo se suman todos los intervalos inferiores.

Toma de datos: Es la obtención de una colección de los mismos que no han sido ordenados numéricamente. Un ejemplo Una muestra aleatoria de 100 estudiantes de una universidad se tomó el peso (Kg)

Ordenación: Llamado también “Arreglo” es la disposición de los datos en cierto orden de magnitud, sea este creciente o decreciente. Su observación lleva de inmediato a tratar de averiguar cuál es la diferencia que existe entre los valores extremos del arreglo (rango)

Rango: Se lo conoce también como “recorrido o rango” es la diferencia entre el dato mayor menos el dato menor de su número, siguiendo con el ejemplo de los estudiantes tendremos:

$X_1= 54$	$X_2= 55$	$X_3= 55$	$X_4= 58$	$X_5= 60$	$X_6= 62$	$X_7= 63$	$X_8= 63$	$X_9= 63$	$X_{10}= 64$
$X_{11}= 65$	$X_{12}=65$	$X_{13}= 66$	$X_{14}= 66$	$X_{15}= 67$	$X_{16}= 67$	$X_{17}= 67$	$X_{18}= 67$	$X_{19}= 67$	$X_{20}=69$
$X_{21}= 69$	$X_{22}=72$	$X_{23}= 72$	$X_{24}= 73$	$X_{25}= 73$	$X_{26}= 73$	$X_{27}= 73$	$X_{28}= 76$	$X_{29}= 77$	$X_{30}= 80$

Ejercicio resuelto:

Rango menor = m (54)

Rango mayor = M (80)

Entonces para obtener el rango:

$R = M - m$

$R = 80 - 54$

$R = 26$

$M = R + m$ $M = 26 + 54$ $M = 80$ (Mayor)
--

$m = M - R$ $m = 80 - 26$ $m = 54$ (Menor)
--

Los datos extremos de este arreglo son 54 y 80 Kg respectivamente y el rango es igual a $80-54 = 26$

Cuando el número de datos que se dispone es grande es conveniente distribuirlo en clases o categorías y determinar el número de datos que pertenecen a cada categoría, número que constituye la frecuencia de clase. Una ordenación tabular de los datos en clases y con las frecuencias correspondientes a cada una, se conoce como una distribución de frecuencias o tabla de frecuencias.

Distribución de frecuencia de peso de estudiantes (kilogramos) de 30 estudiantes de una universidad. Los datos ordenados y resumidos como la distribución de frecuencias anterior, se suelen llamar datos agrupados en intervalos de clase.

2.1.1. INTERVALOS DE CLASE

Un símbolo que define una clase, tal como 54 a 59 de la tabla 2.1, se conoce como intervalo de clase. **Los números extremos de 54 y 59 son los límites de clase;** el número **54 es el límite inferior** de la clase y el mayor **59 es el límite superior**.

Nota.- Los términos clase e intervalo se utilizan a menudo indistintamente, aunque el intervalo de clase realmente es un símbolo.

Reglas para obtener las tablas de frecuencia:

Aplicando la fórmula de Sturges **$K = 1 + 3,323 (\log n)$ o K raíz de n**

Valor del intervalo **$W = R/K$**

Donde **R** es el rango

En este ejercicio se tendrá que $k = 1 + 3,323 (\log 30) = 5,90$ lo aproximamos a 6, pero se hubiera podido considerar que $m = 5$, sin embargo la aproximación no tiene importancia.

2.1.2. TAMAÑO O ANCHURA DE UN INTERVALO DE CLASE

Para obtener el intervalo de clase se lo realiza de la siguiente manera: $R = 26 / 6 = 4,33$ correspondería a la anchura de la clase o longitud de clase.

Entonces a partir de 54 se empieza el límite inferior para formar los distintos intervalos como se muestra a continuación en la siguiente tabla.

Tabla 2.1

Nº Clase	Alturas (pulgadas)	Número de estudiantes (FA)
1	54 - 58.3	4
2	58.3 - 62.6	2
3	62.667 - 66.9	13
4	66.9 - 71.2	2
5	71.2 - 75.5	6
6	75.5 - 80.0	3
	Σ	30

La primera clase o categoría, por ejemplo, comprende pesos de 54 a 59 kilogramos y viene indicada por el símbolo 54-59. Puesto que 4 estudiantes tienen una altura perteneciente a esta clase

Si los pesos se registraron con aproximación, el intervalo de clase 54 – 59 teóricamente se conoce como límites reales de clase o límites verdaderos

Marca de clase

La marca de clase es el punto medio del intervalo de clase y se obtiene sumando los límites inferior y superior de la clase y dividiendo para dos. Así la marca de clase del intervalo $(54+ 59)/2 = 56,5$ este valor también se lo llama punto medio de la clase.

HISTOGRAMA Y POLIGONOS DE FRECUENCIA

Son representaciones gráficas de las distribuciones de frecuencia

Un histograma o histograma de frecuencia consiste en una serie de rectángulos que se caracterizan por:

Sus bases sobre un eje horizontal (eje x) con centros en las marcas de clase y ancho igual al tamaño de los intervalos de clase.

Superficies proporcionadas a las frecuencias de clase

Un polígono de frecuencia es un gráfico de línea trazado sobre las marcas de clase de los rectángulos en el histograma. (Ver Capítulo II).

2.1.3. DISTRIBUCIÓN DE FRECUENCIAS RELATIVA

Si la frecuencia de cada una de las clases se divide por el total de frecuencias de todas las clases, y el resultado se multiplica por 100, de esta manera se obtiene la frecuencia relativa de la tabla 2.1 es $9/100 = 0.09 = 9 \%$. La suma de las frecuencias relativas de todas las clases es 1 o 100 %.

Si las frecuencias en la tabla anterior se sustituyen por las correspondientes frecuencias relativas, la tabla resultante distribución de frecuencias relativas, distribución porcentual o tabla de frecuencias relativas.

Las distribuciones de frecuencias relativas tienen también como representaciones gráficas: histogramas y polígonos. Para aquello se

siguen los mismos procedimientos dados; la diferencia está en el cambio de escala del eje de las ordenadas o frecuencias del histograma otros autores le denominan polígonos de frecuencias.

2.1.4. DISTRIBUCIÓN DE FRECUENCIAS ACUMULADA

Además de conocer el número de observaciones que hay para cada clase, es importante también conocer el total de observaciones que existe desde la primera clase hasta una categoría determinada. Por ejemplo la frecuencia acumulada del intervalo de clase 68.1 – 73 inclusive en la tabla 2.1 la suma de la frecuencia absoluta es = 30 significando que 30 estudiantes tienen pesos menores que 73 kilogramos.

Una tabla que represente las frecuencias acumuladas se llama distribución de frecuencias acumuladas, para la distribución del peso de los estudiantes.

2.1.5. DISTRIBUCIÓN DE FRECUENCIAS RELATIVAS Y ACUMULADAS

2.1.6. GRAFICAS APLICADAS AL DESARROLLO DE LA TEORÍA DE FRECUENCIAS

Diagrama de frecuencias. Son mayormente usados para representa las frecuencias relativas y absolutas. Se representan por líneas delgadas verticales u horizontales.

Ejemplo 1. Con los datos de la tabla 2.1 elaboraremos los respectivos diagramas de frecuencias.

Un histograma es la representación gráfica en el plano coordenado de las características concentradas en la tabla de frecuencias de una variable continua (Figura 2.)

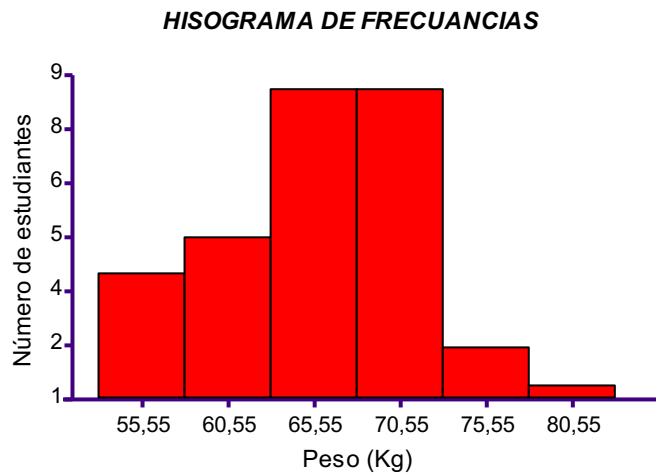


Figura 2. Histograma para el peso kg de estudiantes. En la tabla de frecuencias 2.1.

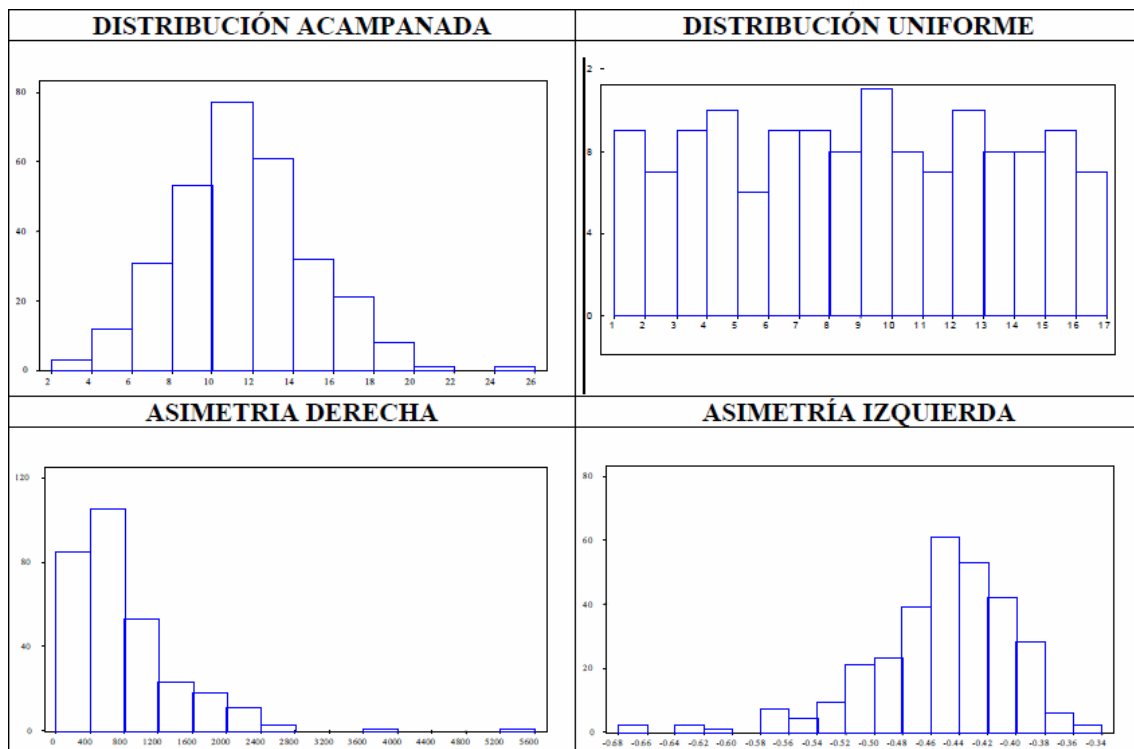
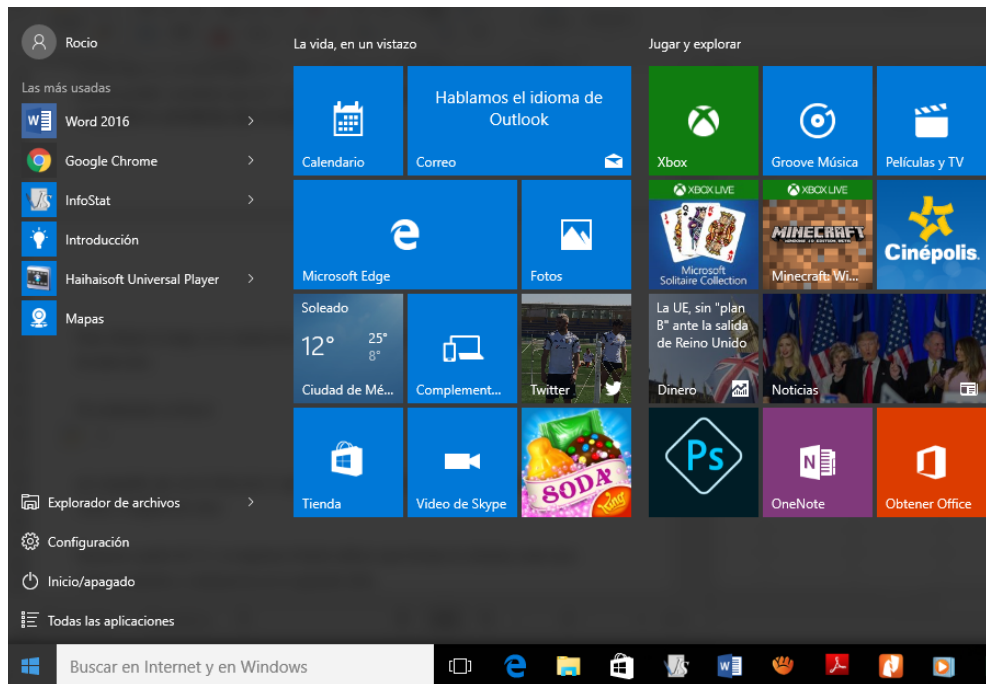


Figura 3. Modelos de distribución y asimetría del histograma.

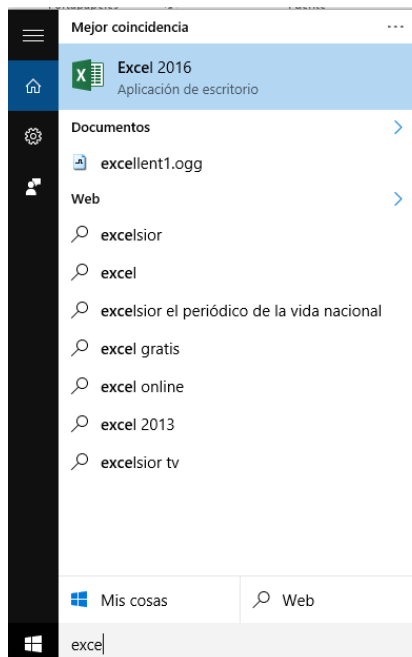
2.1.7. USANDO EXCEL E INFOSTAT

Procedimiento en Excel

Dar clic en inicio



Buscar Excel 2016



Ingresar los datos

	A	B	C	D
1	Peso (kg)			
2	54			
3	55			
4	55			
5	58			
6	60			
7	62			
8	63			
9	63			
10	63			
11	64			
12	65			
13	65			
14	66			
15	66			
16	67			
17	67			
18	67			
19	67			
20	67			
21	69			
22	69			
23	72			
24	72			
25	73			
26	73			
27	73			
28	73			
29	76			
30	77			
31	80			
32				

Contar la cantidad de datos

The screenshot shows an Excel spreadsheet with the following data in column A:

	A	B	C	D	E
1	Peso (kg)				
2	54				
3	55				
4	55				
5	58		SOLUCIÓN		
6	60		N° de Datos	=CONTAR(A2:A31)	
7	62		Valor max. o Rango Mayor	CONTAR(valor1, [valor2], ...)	
8	63		Valor Mín. o Rango menor		
9	63		Rango		
10	63		N° de Intervalos		
11	64		Amplitud de intervalo de clase		
12	65				
13	65				
14	66				
15	66				
16	67				
17	67				
18	67				
19	67				
20	67				
21	69				
22	69				
23	72				

54 y 80 kg respectivamente y el rango es igual a $80 - 54 = 26$.

Cuando el número de datos que se dispone es grande es conveniente distribuirlo en clases o categorías y determinar el número de datos que pertenecen a cada categoría, número que constituye la frecuencia de clase. Una ordenación tabular de los datos en clases y con las frecuencias correspondientes a cada una, se conoce como una distribución de frecuencias o tabla de frecuencias.

Distribución de frecuencia de peso de estudiantes (kilogramos) de 30 estudiantes de la universidad. Los datos ordenados y resumidos como la distribución de frecuencias anterior, se suelen llamar datos agrupados en intervalos de clase.

Procedimiento en Excel

Ingresar la fórmula para calcular el valor máximo o Rango Mayor

The screenshot shows the Excel interface with the formula bar containing `=MAX(A2:A31)`. The spreadsheet data is as follows:

	A	B	C	D
1	Peso (kg)			
2	54			
3	55			
4	55			
5	58			
6	60			
7	62			
8	63			
9	63			
10	63			
11	64			
12	65			
13	65			
14	66			
15	66			
16	67			
17	67			
18	67			
19	67			
20	67			
21	69			
22	69			
23	72			

A summary table titled "SOLUCIÓN" is located in the right side of the spreadsheet:

SOLUCIÓN	
Nº de Datos	30
Valor max. o Rango Δ	<code>=MAX(A2:A31)</code>
Valor Min. o Rango menor	
Rango	
Nº de Intervalos	
Amplitud de intervalo de clase	

Calcular el valor mínimo

The screenshot shows the Excel interface with the formula bar containing `=MIN(A2:A31)`. The spreadsheet data is identical to the previous image:

	A	B	C	D
1	Peso (kg)			
2	54			
3	55			
4	55			
5	58			
6	60			
7	62			
8	63			
9	63			
10	63			
11	64			
12	65			
13	65			
14	66			
15	66			
16	67			
17	67			
18	67			
19	67			
20	67			
21	69			
22	69			
23	72			

The summary table titled "SOLUCIÓN" is updated with the minimum value:

SOLUCIÓN	
Nº de Datos	30
Valor max. o Rango Mayor	80
Valor Min. o Rango μ	<code>=MIN(A2:A31)</code>
Rango	
Nº de Intervalos	
Amplitud de intervalo de clase	

Para calcular el Rango= Valor Max – Valor min

The screenshot shows an Excel spreadsheet with two tables. The first table, in columns A and B, lists weights in kilograms. The second table, in columns C and D, provides the solution for the frequency distribution.

	A	B	C	D
1	Peso (kg)			
2		54		
3		55		
4		55		
5		58	SOLUCIÓN	
6		60	N° de Datos	30
7		62	Valor max. o Rango Mayor	80
8		63	Valor Min. o Rango menor	54
9		63	Rango	=D6-D7
10		63	N° de Intervalos	
11		64	Amplitud de intervalo de clase	
12		65		
13		65		
14		66		
15		66		
16		67		
17		67		
18		67		
19		67		
20		67		
21		69		
22		69		
23		77		

INTERVALOS DE CLASE

Un símbolo que define una clase, tal como 54.00 a 58.33 de la tabla 2.1., se conoce como intervalo de clase. **Los números extremos de 54.00 y 58.33 son los límites de clase;** el número **54.00 es el límite inferior** de la clase y el mayor **58.33 es el límite superior**.

Nota. - los términos clase e intervalo se utilizan a menudo indistintamente, aunque el intervalo de clase realmente es un símbolo.

Reglas para obtener las tablas de frecuencia:

Aplicando la fórmula de Sturges $K=1+3.323(\log n)$ o K raíz de n

Valor de intervalo **$W = R/K$**

Donde **R** es el rango.

En este ejercicio se tendrá que $k= 1 + 3,323 (\log 30) = 5.90$ lo aproximamos a 6, pero se hubiera podido considerar que $m=5$, sin embargo la aproximación no tiene importancia.

Para calcular el N° de Intervalos aplicamos la fórmula de Sturges y redondear el resultado.

	A	B	C	D	E
1	Peso (kg)				
2	54				
3	55				
4	55		SOLUCIÓN		
5	58		N° de Datos	30	
6	60		Valor max. o Rango Mayor	80	
7	62		Valor Min. o Rango menor	54	
8	63		Rango	26	
9	63		N° de Intervalos	=1+3.323*(LOG10(30))	
10	63		Amplitud de intervalo de clase		
11	64				
12	65				
13	65				
14	66				
15	66				
16	67				
17	67				
18	67				
19	67				
20	67				
21	69				
22	69				
23	72				

TAMAÑO O ANCHURA DE INTERVALO DE CLASE

Para obtener el tamaño o amplitud del intervalo de clase se lo realiza de la siguiente manera Rango/N° de intervalos ejemplo: $26 / 6 = 4.333$.

Procedimiento en Excel:

Aplicar: Tamaño del Intervalo= Rango/N° de Intervalos

	A	B	C	D	E
1	Peso (kg)				
2	54				
3	55				
4	55				
5	58				
6	60				
7	62				
8	63				
9	63				
10	63				
11	64				
12	65				
13	65				
14	66				
15	66				
16	67				
17	67				
18	67				
19	67				
20	67				
21	69				
22	69				
23	72				

SOLUCIÓN	
N° de Datos	30
Valor max. o Rango Mayor	80
Valor Min. o Rango menor	54
Rango	26
N° de Intervalos	6
Amplitud de intervalo de clase	=D8/E9

Entonces a partir de 54 empieza el límite inferior para formar los distintos intervalos como se muestra a continuación en la siguiente tabla.

La primera clase o categoría, por ejemplo, comprende de 54 a 58.333 kilogramos y viene indicada por el símbolo 54 – 58.333. cuatro estudiantes tienen el peso perteneciente a esta clase.

Para el cálculo del límite inferior para formar los distintos intervalos con el ejercicio anterior se lo realiza de la siguiente forma:

Crear una tabla e ingresar el límite inferior “54”

	A	B	C	D	E	F	G
1	Peso (kg)						
2	54						
3	55						
4	55						
5	58						
6	60						
7	62						
8	63						
9	63						
10	63						
11	64						

SOLUCIÓN	
N° de Datos	30
Valor max. o Rango Mayor	80
Valor Min. o Rango menor	54
Rango	26
N° de Intervalos	6
Amplitud de intervalo de clase	4.333

Intervalos de Clase	
Lim Inferior	54
Lim Superior	

Con el valor mínimo que en este caso es 54 sumar el la amplitud o tamaño del intervalo

	A	B	C	D	E	F	G
1	Peso (kg)						
2		54					
3		55					
4		55					
5		58	SOLUCIÓN			Intervalos de Clase	
6		60	N° de Datos	30		Lim Inferior	Lim Superior
7		62	Valor max. o Rango Mayor	80		54	=F6+D10
8		63	Valor Min. o Rango menor	54			
9		63	Rango	26			
10		63	N° de Intervalos	6	6		
11		64	Amplitud de intervalo de clase	4.333			

Agregar el segundo límite inferior y sumar la amplitud del intervalo "4.333"

	A	B	C	D	E	F	G
1	Peso (kg)						
2		54					
3		55					
4		55					
5		58	SOLUCIÓN			Intervalos de Clase	
6		60	N° de Datos	30		Lim Inferior	Lim Superior
7		62	Valor max. o Rango Mayor	80		54	58.333
8		63	Valor Min. o Rango menor	54		=G6	
9		63	Rango	26			
10		63	N° de Intervalos	6	6		
11		64	Amplitud de intervalo de clase	4.333			

	A	B	C	D	E	F	G
1	Peso (kg)						
2		54					
3		55					
4		55					
5		58	SOLUCIÓN			Intervalos de Clase	
6		60	N° de Datos	30		Lim Inferior	Lim Superior
7		62	Valor max. o Rango Mayor	80		54	58.333
8		63	Valor Min. o Rango menor	54		58.333	62.667
9		63	Rango	26		62.667	=F8+D10
10		63	N° de Intervalos	6	6		
11		64	Amplitud de intervalo de clase	4.333			

Realizar el mismo procedimiento con las demás clases o categorías

	A	B	C	D	E	F	G
1	Peso (kg)						
2		54					
3		55					
4		55	SOLUCIÓN			Intervalos de Clase	
5		58	Nº de Datos	30		Lim Inferior	Lim Superior
6		60	Valor max. o Rango Mayor	80		54	58.33
7		62	Valor Min. o Rango menor	54		58.33	62.67
8		63	Rango	26		62.67	67
9		63	Nº de Intervalos	6	6	67	71.33
10		63	Amplitud de intervalo de clase	4.333		71.33	75.67
11		64				75.67	80

Si los pesos se registraron con aproximación, el intervalo de clase 54 – 59 teóricamente se conoce como límites reales de clase o límites verdaderos.

MARCA DE CLASE

La marca de clase es el punto medio del intervalo de clase y se obtiene sumando los límites inferior y superior de la clase y dividiendo para 2. Así la marca de clase del Intervalo $(54 + 58.33) / 2 = 55.2$ este valor también se lo llama punto medio de la clase.

Colocar $= (\text{lím inferior} + \text{lím superior}) / 2$ y arrastrar la fórmula para los demás valores

	A	B	C	D	E	F	G	H
1	Peso (kg)							
2		54						
3		55						
4		55	SOLUCIÓN			Intervalos de Clase		Marca de clase
5		58	Nº de Datos	30		Lim Inferior	Lim Superior	
6		60	Valor max. o Rango Mayor	80		54	58.33	$=((F6+G6)/2)$
7		62	Valor Min. o Rango menor	54		58.33	62.67	
8		63	Rango	26		62.67	67	
9		63	Nº de Intervalos	6	6	67	71.33	
10		63	Amplitud de intervalo de clase	4.333		71.33	75.67	
11		64				75.67	80	

Arrastrar la fórmula para los demás datos

	A	B	C	D	E	F	G	H
1	Peso (kg)							
2		54						
3		55						
4		55						
5		58						
6		60						
7		62						
8		63						
9		63						
10		63						
11		64						

SOLUCIÓN			Intervalos de Clase		
Nº de Datos	Valor max. o Rango Mayor	Lim Inferior	Lim Superior	Marca de clase	
30	80	54	58.33	56.17	
Valor Mín. o Rango menor	54	58.33	62.67	60.50	
Rango	26	62.67	67	64.83	
Nº de Intervalos	6	67	71.33	69.17	
Amplitud de intervalo de clase	4.333	71.33	75.67	73.50	
		75.67	80	77.83	

HISTOGRAMAS Y POLIGONOS DE FRECUENCIA

Es la representación gráfica de las frecuencias agrupadas de una variable continua sobre intervalos. A diferencia de los diagramas de barras, los histogramas dibujan rectángulos unidos entre sí, lo que significa que existe continuidad en la variable cuyos valores se representan en el eje horizontal que se haya dividido en intervalos de igual amplitud. Las áreas de los rectángulos son proporcionales a las frecuencias que representan.

Son distribuciones graficas de las distribuciones de frecuencias consiste en una serie de rectángulos que se caracterizan por:

Sus bases sobre un eje horizontal (eje x) con centros en las marcas de la clase y ancho igual al tamaño de los intervalos de clase.

Superficies proporcionadas a las frecuencias de clase.

Un polígono de frecuencia es un gráfico de línea trazado sobre las marcas de clase de los rectángulos en el histograma.

FRECUENCIAS ABSOLUTAS

Intervalos de Clase		Marca de clase	ni
Lim Inferior	Lim Superior		
54	58.33	56.17	4
58.33	62.67	60.50	2
62.67	67	64.83	13
67	71.33	69.17	2
71.33	75.67	73.50	6
75.67	80	77.83	3
			30
			F. Abs

FRECUENCIAS ABSOLUTAS ACUMULADAS

La primera será la misma que la frecuencia absoluta

Intervalos de Clase		Marca de clase	ni	h	hi
Lim Inferior	Lim Superior				
54	58.33	56.17	4	0.13	4
58.33	62.67	60.50	2	0.07	
62.67	67	64.83	13	0.43	
67	71.33	69.17	2	0.07	
71.33	75.67	73.50	6	0.20	
75.67	80	77.83	3	0.10	
			30		
			F. Abs	F. Relativ. (%)	

La segunda será el valor acumulado de la primera frecuencia absoluta más la segunda "4 + 2 = 6"

Intervalos de Clase		Marca de clase	ni	h	hi		
Lim Inferior	Lim Superior						
54	58.33	56.17	4	0.13	4		
58.33	62.67	60.50	2	0.07	6		
62.67	67	64.83	13	0.43			
67	71.33	69.17	2	0.07			
71.33	75.67	73.50	6	0.20			
75.67	80	77.83	3	0.10			
			30				
			F. Abs	F. Relativ. (%)	F. A. Acum.		

Para las siguientes frecuencias arrastrar la formula y observar que la última frecuencia es igual al número de datos, es un indicativo que se está realizando correctamente el proceso.

Intervalos de Clase		Marca de clase	ni	h	hi
Lim Inferior	Lim Superior				
54	58.33	56.17	4	0.13	4
58.33	62.67	60.50	2	0.07	6
62.67	67	64.83	13	0.43	19
67	71.33	69.17	2	0.07	21
71.33	75.67	73.50	6	0.20	27
75.67	80	77.83	3	0.10	30
			30		
			F. Abs	F. Relativ. (%)	F. A. Acum.

DISTRIBUCIÓN DE FRECUENCIAS RELATIVAS

La frecuencia relativa acumulada es el **cociente** entre la **frecuencia acumulada** de un determinado **valor** y el **número total de datos**.

Si la frecuencia de cada una de las clases se divide por el total de frecuencias de todas las clases, y el resultado se multiplica por 100, de esta manera se obtiene la frecuencia relativa es $9/100 = 0,09 = 9\%$. La suma de las frecuencias relativas de todas las clases es 1 o 100%.

Si las frecuencias en la tabla anterior se sustituyen por las correspondientes frecuencias relativas, la tabla resultante distribución de frecuencias relativas, distribución porcentual o tabla de frecuencias relativas.

Las distribuciones de frecuencias relativas tienen como representaciones gráficas: histogramas y polígonos. Para ello se siguen los mismos procedimientos dados, la diferencia está en el cambio de la escala del eje de las ordenadas o frecuencias del histograma otros autores le denominan polígonos de frecuencias.

Procedimiento de la frecuencia relativa

Para el calcula de la frecuencia relativa es = frecuencia absoluta / número de datos y presionar F2 y luego F4 para evitar cambios en los resultados al arrastrar la fórmula.

Intervalos de Clase	Marca de clase	ni	h	hi	Hi
---------------------	----------------	----	---	----	----

Intervalos de Clase		Marca de clase	ni	h
Lim Inferior	Lim Superior			
54	58.33	56.17	4	0.13
58.33	62.67	60.50	2	
62.67	67	64.83	13	
67	71.33	69.17	2	
71.33	75.67	73.50	6	
75.67	80	77.83	3	
			30	
			F. Abs	

Arrastrar la formula.

Intervalos de Clase		Marca de clase	ni	h
Lim Inferior	Lim Superior			
54	58.33	56.17	4	0.13
58.33	62.67	60.50	2	0.07
62.67	67	64.83	13	0.43
67	71.33	69.17	2	0.07
71.33	75.67	73.50	6	0.20
75.67	80	77.83	3	0.10
			30	
			F. Abs	F. Relativ. (%)

FRECUENCIAS RELATIVAS ACUMULADAS

La primera frecuencia relativa acumulada será igual a la primera frecuencia relativa (%).

Intervalos de Clase		Marca de clase	ni	h	hi	Hi
Lim Inferior	Lim Superior					
54	58.33	56.17	4	0.13	4	=6
58.33	62.67	60.50	2	0.07	6	
62.67	67	64.83	13	0.43	19	
67	71.33	69.17	2	0.07	21	
71.33	75.67	73.50	6	0.20	27	
75.67	80	77.83	3	0.10	30	
			30			
			F. Abs	F. Relativ. (%)	F. A. Acum.	F. R. Acum

La segunda frecuencia relativa acumulada será la suma de la primera frecuencia relativa más la segunda

Lim. Inferior o rango menor	Lim. Superior o rango Mayor					
54	58.3	56.17	4	0.13	4	0.13
58.3	62.6	60.50	2	0.07	6	0.20
62.67	67	64.83	13	0.43	19	0.63
67	71.33	69.17	2	0.07	21	0.70
71.33	75.67	73.50	6	0.20	27	0.90
75.67	80	77.83	3	0.10	30	1.00

30

F. Abs F. Relativ. (%) F. Acum. A. F. Acum. R.

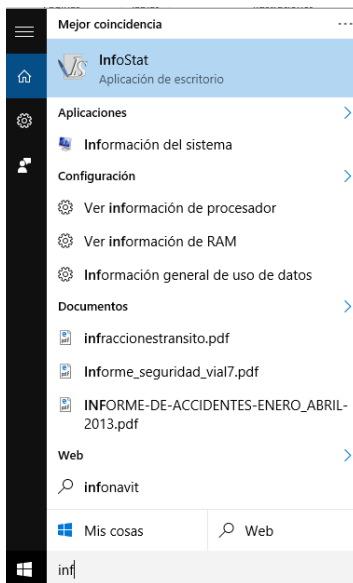
Intervalos de Clase		Marca de clase	ni	h	hi	Hi
Lim Inferior	Lim Superior					
54	58.33	56.17	4	0.13	4	0.13
58.33	62.67	60.50	2	0.07	6	=L6+J7
62.67	67	64.83	13	0.43	19	
67	71.33	69.17	2	0.07	21	
71.33	75.67	73.50	6	0.20	27	
75.67	80	77.83	3	0.10	30	
			30			
			F. Abs	F. Relativ. (%)	F. A. Acum.	F. R. Acum

Arrastrar la fórmula para las demás categorías.

Intervalos de Clase		Marca de clase	ni	h	hi	Hi
Lim Inferior	Lim Superior					
54	58.33	56.17	4	0.13	4	0.13
58.33	62.67	60.50	2	0.07	6	0.20
62.67	67	64.83	13	0.43	19	0.63
67	71.33	69.17	2	0.07	21	0.70
71.33	75.67	73.50	6	0.20	27	0.90
75.67	80	77.83	3	0.10	30	1.00
			30			
			F. Abs	F. Relativ. (%)	F. A. Acum.	F. R. Acum

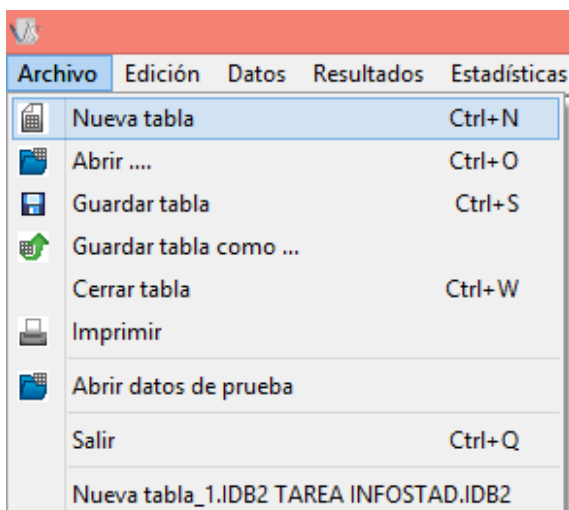
USANDO EL INFOSTAT:

En el menú inicio buscar Infostat



Ingresar a Infostat.

Seleccionar Archivo y Nueva Tabla.



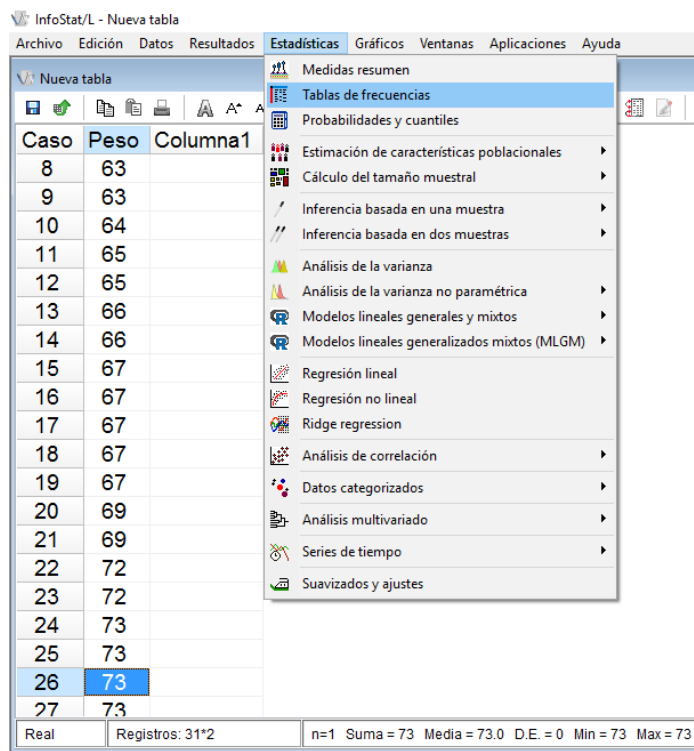
Copiar los datos de Excel

	A	B	C	D
1	Peso (kg)			
2	54			
3	55			
4	55			
5	58			
6	60			
7	62			
8	63			
9	63			
10	63			
11	64			
12	65			
13	65			
14	66			
15	66			
16	67			
17	67			
18	67			
19	67			
20	67			
21	69			
22	69			
23	72			
24	72			
25	73			
26	73			
27	73			
28	73			
29	76			
30	77			
31	80			
32				

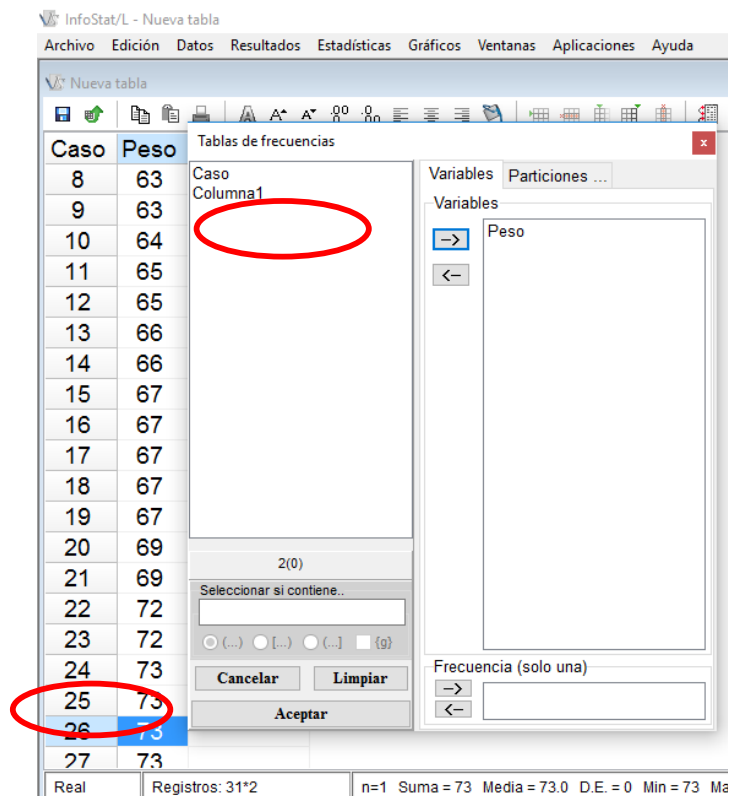
Clic derecho sobre la columna 1 en InfoStat y seleccionar Pegar Incluyendo nombre de columnas.

Caso	Columna1
1	54
2	55
3	55
4	58
5	60
6	62
7	63
8	63
9	63
10	64
11	65
12	65
13	66
14	66
15	67
16	67
17	67
18	67
19	67
20	69
21	69
22	72
23	72
24	73
25	73
26	73
27	73
28	76
29	77
30	80
31	

Ir a Estadística y seleccionar tabla de frecuencias



Pasar a variable "PESO" y aceptar.



Seleccionar las frecuencias Absolutas acumuladas y Frecuencia relativa acumulada, en número de clases personalizado 6 y aceptar.

InfoStat/L - Nueva tabla

Archivo Edición Datos Resultados Estadísticas Gráficos Ventanas Aplicaciones

Nueva tabla

Caso Peso

8	63
9	63
10	64
11	65
12	65
13	66
14	66
15	67
16	67
17	67
18	67
19	67
20	69
21	69
22	72
23	72
24	73
25	73
26	73
27	73

Tablas de frecuencias

Opciones de la tabla de frecuencias Ajustes

Variable: Peso (Real)

LI LS MC FAA FRA

FA FR E(FA) E(FR)

Frecuencias absolutas acumuladas

Usar cálculo automático en todas las variables

Número de clases

Calcular

Automáticamente Personalizado 6

Valor mínimo 54 Intervalo de clase

Valor máximo 80 4.33333333333333

Tratar a las variables enteras como conteos

Intervalos cerrados por derecha

Aceptar Cancelar Ayuda

Real | Registros: 31*2 | n=1 Suma = 73 Media = 73.0 D.E. = 0 M

InfoStat/L - Nueva tabla

Archivo Edición Datos Resultados Estadísticas Gráficos Ventanas Aplicaciones Ayuda

Nueva tabla

Caso Peso

2	55
3	55
4	58
5	60
6	62
7	63
8	63
9	63
10	64
11	65
12	65
13	66
14	66

Resultados

Nueva tabla : 25/02/2016 - 11:17:23 p. m. - [Versión : 01/11/2014]

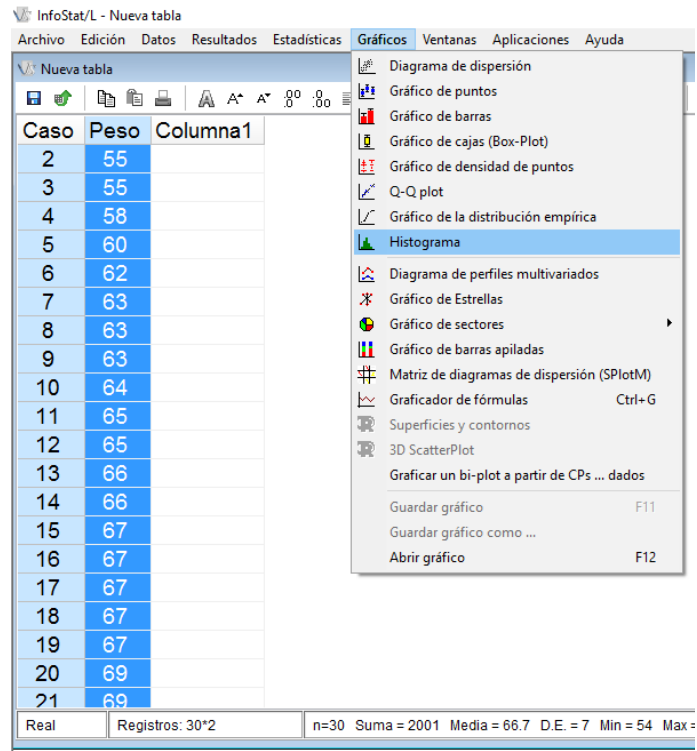
Tablas de frecuencias

Variable	Clase	LI	LS	MC	FA	FR	FAA	FRA
Peso	1	[54.00	58.33]	56.17	4	0.13	4	0.13
Peso	2	(58.33	62.67]	60.50	2	0.07	6	0.20
Peso	3	(62.67	67.00]	64.83	13	0.43	19	0.63
Peso	4	(67.00	71.33]	69.17	2	0.07	21	0.70
Peso	5	(71.33	75.67]	73.50	6	0.20	27	0.90
Peso	6	(75.67	80.00]	77.83	3	0.10	30	1.00

TablaFrec TablaFrec TablaFrec

HISTOGRAMAS

Seleccionar gráficos e histograma



Elegir la variable "Calificaciones" y aceptar.

InfoStat/L - Nueva tabla

Archivo Edición Datos Resultados Estadísticas Gráficos Ventanas Aplicaciones Ayuda

Nueva tabla

Caso	Peso	Columna1
2	55	
3	55	
4	58	
5	60	
6	62	
7	63	
8	63	
9	63	
10	64	
11	65	
12	65	
13	66	
14	66	
15	67	
16	67	
17	67	
18	67	
19	67	
20	69	
21	69	

Histograma

Caso
Columna1

Variables Particiones ...

Variables

> Peso

< -

2(0)

Seleccionar si contiene..

(...) [...] (...) {g}

Cancelar Limpiar

Aceptar

Real Registros: 30*2 n=30 Suma = 2001 Media = 66.7 D.E. = 7 Min = 54 Max = 80 P05 = 54 P95 = 76

Numero de clases "6"

InfoStat/L - Nueva tabla

Archivo Edición Datos Resultados Estadísticas Gráficos Venta

Nueva tabla

Caso	Peso
2	55
3	55
4	58
5	60
6	62
7	63
8	63
9	63
10	64
11	65
12	65
13	66
14	66
15	67
16	67
17	67
18	67
19	67
20	69
21	69

Herramientas gráficas

Series EjeX EjeY

Histograma-Peso

Seleccionar si contiene.. Tamaño 0

Opciones de histograma

Ajuste Ninguno

Frec. Frec. rel.

Clases 6

M. clases

Poligono

Bordes

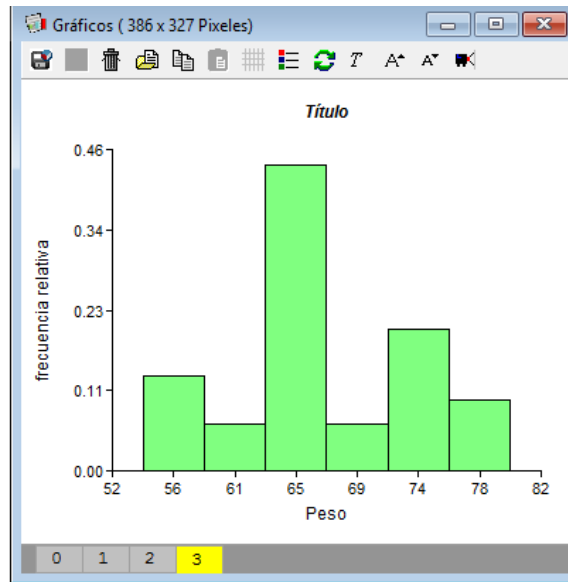
Cuerpo

Título visible

Título

Real Registros: 30*2 n=30 Suma = 2001

Grafico



CAPITULO III

MEDIDAS DE TENDENCIA O POSICIÓN CENTRAL

3.0. MEDIDAS DE TENDENCIA O POSICIÓN CENTRAL

En el capítulo anterior se estudió la manera de cómo agrupar tubularmente y analizar de una manera simple un conjunto de datos obtenidos por medio de una investigación. En un arreglo de datos que ofrece una información acerca de la variable en estudio, sin embargo no todo se ha dicho acerca de la distribución; tal vez sería conveniente hallar otras descripciones fundamentales de la variable y de su distribución; por ejemplo describir simplificada mente y representar a toda la variable o al conjunto de datos recolectados de una variable con el promedio o medidas de tendencia central.

También es necesario afirmar que existen varias medidas que describen el centro de una distribución que ofrece un conjunto muy grande de datos que por razones de exposición se han aproximado mediante una curva de trazo continuo.

Estas medidas de posición o tendencia central, denominados también como promedios, nos permiten determinar la posición de un valor respecto a un conjunto de datos, el cual se considera representativo o típico del total de las observaciones, sin embargo estas medidas se les denomina también estimadores o estadígrafos determinados a partir de un conjunto de datos de la población (Martínez, 2007).

3.1. NOTACIÓN MATEMÁTICA NECESARIA

Antes de estudiar las medidas de tendencia central es conveniente describir y que el estudiante conozca el uso de ciertos símbolos y notación para las variables y sus valores particulares. Ya que si una variable se denota por x , los valores particulares que toma la variable o sea las observaciones, pueden expresarse como:

$$X_1, X_2, X_3, \dots, X_i, \dots, X_N.$$

En esta notación, un valor cualquiera se representa como X_i (léase "X sub i"), generalmente coincide como la i – esima observación, y si hay N observaciones se dice que el valor de i varía hasta N .

Nota: Cuando todos los elementos que forman un conjunto son muchos y conocidos, no es necesario escribir, bastará con algunos de ellos, generalmente tres, seguido con una coma y después tres puntos seguidos.

Muchas veces es necesario indicar la suma de las observaciones de X , si N es pequeña; pero en caso contrario si N es muy grande es necesario usar la letra griega **sigma mayúscula**, Σ que significa "la suma de".

$$\sum_{i=1}^n X_i$$

Este símbolo se lee "la suma de X -sub- i , donde i va de 1 hasta N observaciones".

La letra colocada debajo del símbolo de suma se llama **índice** de suma.

Ejemplo:

$$\text{Sea } X_1 = 4 ; X_2 = -2 ; X_3 = 8$$

Solución

$$\sum_{i=1}^3 X_i = X_1 + X_2 + X_3 = 4 - 2 + 8 = 10$$

$$\sum_{i=1}^3 3X_i = 3X_1 + 3X_2 + 3X_3 = 12 - 6 + 24 = 30$$

$$\sum_{j=2}^3 X_j = X_2 + X_3 = -2 + 8 = 6$$

$$\sum_{k=1}^2 X_k = X_1 + X_2 = 4 - 2 = 2$$

Problema

Para los siguientes datos del ejercicio **a)** halle el valor de $\sum X_i^2$.

Solución:

$$\begin{aligned} \sum X_i^2 &= X_1^2 + X_2^2 + X_3^2 = 4^2 + (-2)^2 + 8^2 \\ &= 16 + 4 + 64 \\ &= 84. \end{aligned}$$

TEOREMA

Un factor constante, que afecta a una variable, puede sacarse del símbolo de suma. Así, si (a) es una constante y X una variable, entonces:

$$\sum_{i=1}^n aX_i = a \rightarrow \sum_{i=1}^n X_i$$

Prueba:

$$\begin{aligned} \sum_{i=1}^n aX_i &= aX_1 + aX_2 + \dots + aX_n \\ &= a(X_1 + X_2 + \dots + X_n) \\ &= a \sum X_i \end{aligned}$$

3.2. LA MEDIA ARITMETICA

La media aritmética o simplemente llamada media de un conjunto de n número tales como X_1, X_2, \dots, X_n , se denota por \bar{X} (léase X barra) y se define como "la suma de los valores de n números divididos en n ".



Representación gráfica del concepto de media

Usando los símbolos explicados, puede describirse como:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \sum_{i=1}^n X_i / n \text{ muestra; } \mu = \sum_{i=1}^n X_i / N \text{ población}$$

La media aritmética se conoce también con el nombre de promedio. Sin embargo, este término se usa para señalar a otras medidas de tendencia central.

Ventajas:

Es fácil de calcular y comúnmente usada

Es estable en el muestreo

Es altamente sensible a cualquier cambio de los datos

Desventajas:

Sus valores pueden ser distorsionados por valores extremos y por consiguiente puede no resultar muy representativa para la población o muestra.

Cuando una distribución es asimétrica puede que la media difiera de la moda, y mediana haciéndola menos representativa

Ejemplo 1:

Hallar la media de 11, 14, 17, 20, 16 y, 20 con aproximación de las cifras decimales.

$$\bar{X} = \frac{11+14+17+20+16+10}{6} = \frac{88}{6} = 14,67$$

Ejemplo 2:

La media aritmética de 8, 3, 5, 12, 10 es:

$$\bar{X} = \frac{8+3+5+12+10}{5} = \frac{38}{5} = 7,60$$

3.3. OTROS TIPOS DE MEDIA

LA MEDIA GEOMETRICA $M_g = M_o$

La media geométrica se expresa como "raíz n-ésima del producto de los números que forman los datos de la variable"

Fórmula

$$\text{Datos sin agrupar; } M_o = \sqrt{X_1.X_2.X_3.....X_N} ; \quad M_o = \sqrt{\pi.X_1}$$

$$\text{Datos agrupados; } M_o = \sqrt{f_1.f_2.f_3.....f_N}$$

Las fórmulas algunas veces es conveniente expresarla en forma logarítmica como:

$$\text{Log } \bar{X}_g = \frac{1}{N} (\text{Log } X_1 + \dots + \text{log } X_N)$$

Ejemplo 1:

Hallar la media geométrica de los números 2, 4 y, 8

$$M_o = \sqrt[3]{(2).(4).(8)} = \sqrt[3]{64} = 4 \text{ respuesta}$$

Problema:

Hallar la media geométrica en una ganadería bovina con los siguientes datos:

Cabezas de ganado	Año	Relación
1500	1940	..
3000	1950	2
24000	1960	8

Solución: $M_o = \sqrt{2 \times 8} = \sqrt{16} = 4$

$$4 \times 1500 = 6000$$

$$4 \times 6000 = 24000 \text{ (Es más real)}$$

Usando \bar{X} sería:

$$\bar{X} = (2+8)/2 = 5$$

entonces: $5 \times 1500 = 7500$

$$5 \times 7500 = 375000 \text{ (No es cierto)}$$

3.4. LA MEDIA PONDERADA

Para comprender el concepto de media ponderada pongamos el siguiente ejemplo: supongamos que en una tienda de agroquímicos se venden tres marcas de bioles a \$4.40, \$ 4.20 y, \$ 4,30 en día cualquiera se vendieron los productos con los tres precios. El precio medio de venta sería:

$$\bar{X} = \frac{4.10 + 4.20 + 4.30}{3} = 4.20$$

Pero si además se conoce, que vendieron 23 bioles marca A de 4.20, 72 bioles marca B de 4.10 y, 5 bioles marca C de 4,30

$$\bar{X} = \frac{(23 \times 4.20) + (72 \times 4.10) + (5 \times 4.30)}{23 + 72 + 5} = \frac{413.30}{100} = 4.13$$

Cantidad que es más representativa, que el simple valor medio hallado inicialmente.

Definición:

Dado un conjunto de observaciones tales como $X_1, X_2, \dots + X_N$ y un conjunto de valores P_1, P_2, \dots, P_N , asociados con cada observación X_i , respectivamente, recibiendo el nombre de factores de ponderación.

Media ponderada; $\bar{X}_p = \frac{X_1P_1 + X_2P_2 + \dots + X_NP_N}{P_1 + P_2 + \dots + P_N} = \frac{\sum X_iP_i}{\sum P_i}$

3.4.1. Propiedades de la media

Es necesario conocer y manejar correctamente las propiedades que tiene la media, ya que ayuda mucho a comprender ciertas operaciones matemáticas.

La suma de las desviaciones respecto a la media (simple y ponderada) es cero

En datos no agrupados $\sum_1^n (X_i - \bar{X}) = 0$

Ejemplo 1

Sean los números 2, 5, 7, 9 y, 12 con $\bar{X} = 7$

1, 3, 5, 7 y, 9 con $\bar{X} = 5$

Las desviaciones con respecto a la media, se simboliza como $d_j = X_i - \bar{X}$, y la suma de estas desviaciones de la siguiente manera:

$d_1 = 2 - 7 = -5$

$d_2 = 5 - 7 = -2$

$d_3 = 7 - 7 = 0$

$d_4 = 9 - 7 = 2$

$d_5 = 12 - 7 = 5$

Si las desviaciones de N números X_1, X_2, \dots, X_N de un número cualquiera A son: $d_1 = X_1 - A, d_2 = X_2 - A, d_N = X_N - A$ que:

$\bar{X} = A + \sum d_i / N = \sum d / N$ puesto que $d_j = X_j - A$ y $X_j = A$

+ $d_j: \bar{X} = \sum d_j / N = \sum A + d_j / N = \sum NA + \text{Sumatoria } d_j / N =$

$$\sum_{i=1}^5 d_j = 0$$

$$A + \sum d_j / N$$

En datos agrupados

$$\sum Z_i n_i = \sum (Y_i - \bar{Y}) n_i = 0$$

Aplicación:

X_i	F_i	$X_i - f_i$	$X_i - \bar{X}$	$d_i f_i$
2	3	6	3,9	11,7
4	4	16	1,9	7,6
6	6	36	0,1	0,6
8	5	40	2,1	10,5
10	2	20	4,1	8,2
Σ	20	118	0

$$\bar{X} = 5,9$$

La media de una constante es igual a la constante $M_k = K$ Si un técnico poda 4 parcelas de cacao y en todos ellos alcanzó podar 4 su promedio será cuatro.

$$\bar{X} = \frac{4+4+4+4}{4} = \frac{16}{4} = 4$$

Si cada valor de un conjunto X_1, X_2, \dots, X_N se multiplica (o se divide) por una constante, la media aritmética del nuevo conjunto es igual a la media aritmética del conjunto original multiplicado (o dividido) por la constante. En una industria azucarera se procesa 20 sacos con un peso en kilos de 1.190. Cuántas libras corresponderá.

$$\bar{X} = \frac{1.190}{20} = 59,5 \text{ kilos}$$

3.5. MODO = MODA = VALOR MODAL: M_d

La moda se define como el valor o número que ocurre con más frecuencia, o también el valor que más se repite. Aunque pueden existir

casos de dos modas o más tienen la denominación bimodal o plurimodal.

Ejemplo 1:

En el grupo siguiente 1, 5, 5, 9, 12, 12, 12, 14

En este ejemplo la moda será 12 en este caso es unimodal

Ejemplo 2:

En un conjunto de datos se registró los siguientes valores 2, 9, 21, 33, 72

En este ejemplo no existe valor modal.

La moda en datos agrupados

En el caso de datos agrupados donde se ha construido una curva de frecuencias para ajustar los datos, la moda es el valor de X correspondiente al máximo de la curva. Este valor de X se representa muchas veces con \hat{X} .

De una distribución de frecuencias o un histograma, la moda puede obtenerse con la siguiente fórmula:

$$\text{Moda: } M_d = L_1 + \left(\frac{\Delta_1}{\Delta_2 + \Delta_1} \right) c$$

Donde:

L_1 = Límite real inferior de clase (es decir la clase que contiene la moda)

Δ_1 = Exceso de la frecuencia modal sobre la frecuencia de la clase contigua inferior

Δ_2 = Exceso de la frecuencia modal sobre la frecuencia de la clase contigua superior

c = Tamaño de intervalos de clase modal

3.6. MEDIANA M_e

Es considerada como una medida de tendencia central, su importancia que la media aritmética, sin embargo para evitar dejarse llevar por valores extremos. Es estable, y se la describe como un punto medio de

un conjunto de datos, a este valor se le denomina mediana de n valores (Freund).

3.6.1. Mediana en datos no agrupados

Dado un conjunto de datos en orden creciente de magnitud, la mediana es el número colocado en el centro del arreglo, de tal forma que una mitad de las observaciones esté por encima y la otra por debajo de dicho valor. Si el número de observaciones es par, la mediana es la media de los valores centrales.

Ejemplo1:

	X_1 ,	X_2 ,	X_3 ,	X_4 ,	X_5 ,	X_6 ,	X_7 ,	X_8 ,		
Sean los números	3	4	4	5	6	8	8	10	$n = 8$	La mediana = 5.5
Sena los números	5	5	7	9 11		12	15	18	$n = 8$	La mediana = 14.5

$$M_e = \frac{9+11}{2} = 14,5$$

Fórmula

$$\left\{ \begin{array}{l} = M_e = \frac{X(n+1)}{2}, \text{impar} \\ = M_e = \frac{X_{n/2} + X_{n/2+1}}{2}, \text{impar} \end{array} \right.$$

3.6.2. Problemas:

Hallar la mediana de siguientes datos: 10, 11, 14, 16, 17, 20, calcular la media aritmética y la mediana:

Solución:

$$n = 6$$

$$\bar{X} = \frac{10+11+14+16+17+20}{6} = \frac{88}{6} = 14.67$$

Los valore corresponden a X_3 y X_4 entontes la mediana será:

$$M_e = \frac{14+16}{2} = \frac{30}{2} = 15$$

En este caso el valor de la media y mediana se ajustan, sin embargo la mediana es más representativa respecto al punto medio o valor central.

3.6.1. La mediana con datos agrupados

Si se tiene una distribución de frecuencia, la mediana es igualmente ese valor que tiene 50% de las observaciones por debajo y 50% por encima. Geométricamente, la mediana es el valor de X sobre el eje de las abscisas correspondiente a la ordenada que divide un histograma en dos partes iguales.

Para hallar el valor de la mediana en el caso de datos agrupados, debe encontrarse primero la clase mediana, la que define como la clase más baja para la cual la frecuencia acumulada excede $N/2$ siendo $N = \sum f_i$. Encontrada esta clase, la siguiente fórmula servirá para hallar el valor de la mediana de la siguiente manera:

$$\text{Mediana} = L_m + \left(\frac{\frac{N}{2} - FL}{F_m} \right) c$$

Donde:

L_m : es el límite real más bajo de la clase en la cual la mediana debe caer

F_m : es la frecuencia de esa clase

FL : es la frecuencia total de todas las clases debajo de esa clase

c : es el tamaño del intervalo de la clase.

Ejemplo2:

Nº	Límites de clases (Kg)	Frecuencia
1	52.5 - 57,5	8
2	57.5 - 62.5	9
3	62,5 - 67,5	6
4	67.5 - 72.5	4
5	72.5 - 77.5	2
6	77.5 - 82.5	1
		30

En la primera clase se encuentra 8 observaciones, debidamente hallar la mediana en la segunda clase, que indica que 15 por encima y 15 por abajo.

$$Lm = 57.5$$

$$N = \sum f_i = 30$$

$$FL = 8$$

$$Fm = 9$$

$$c = 5$$

$$\text{Mediana} = 57.5 + \left(\frac{15 - 8}{9} \right)^5 = 57.5 + 3.9 = 61.4$$

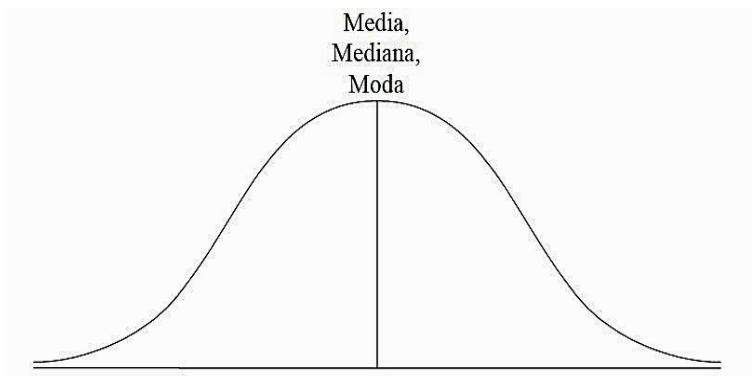
3.7. RELACIÓN EMPÍRICA ENTRE MEDIA, MEDIANA Y MODA

Para curvas de frecuencias unimodales que sean moderadamente sesgadas (asimétricas) se tiene la relación empírica:

$$\text{Media} - \text{Moda} = 3 (\text{Media} - \text{Mediana})$$

$$\text{Moda} = \text{Mediana} - 3 (\text{Media} - \text{Mediana})$$

$$\text{Entonces: Mediana} = \frac{2\text{Media} + \text{Moda}}{3} \quad \text{y} \quad \text{Media} = \frac{3\text{Mediana} - \text{Moda}}{2}$$



Por último es interesante señalar que en distribuciones totalmente simétricas, la media, la moda y, la mediana coinciden, localizándose en un mismo valor.

RELACIÓN ENTRE MEDIAS ARITMÉTICA(X), GEOMÉTRICAS (M_g) Y ARMÓNICAS (M_a).

La media geométrica de una serie de números positivos X_1, X_2, \dots, X_N es igual que su media aritmética, pero es mayor ó igual que su media armónica.

En símbolos $H \leq M_g \leq X$

La igualdad de los signos se presentan solamente cuando todos los números X_1, X_2, \dots, X_N son idénticas.

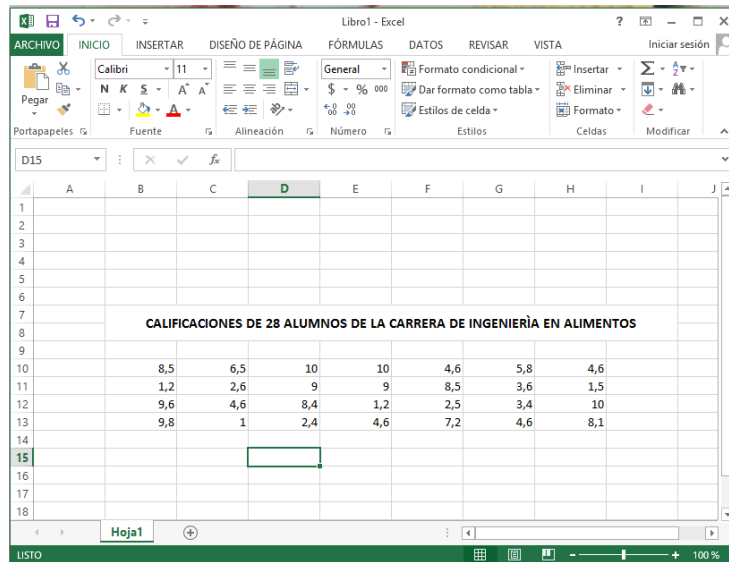
Ejemplo 1: los números 2, 4, y, 8 tienen de media aritmética 4.67; media geométrica 4.00 y de media armónica 3.43.

Entonces $3.43 < 4.00 < 4.67$

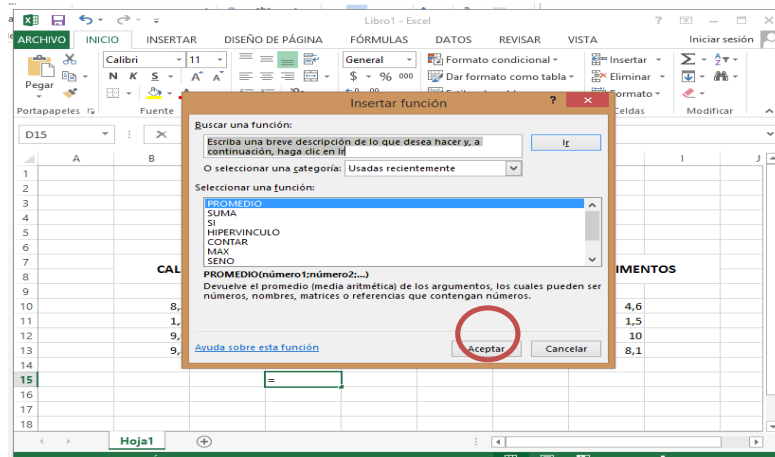
3.8. USANDO EL EXCEL E INFOSTAT

También podemos calcular la media aritmética en Excel siguiendo estos pasos:

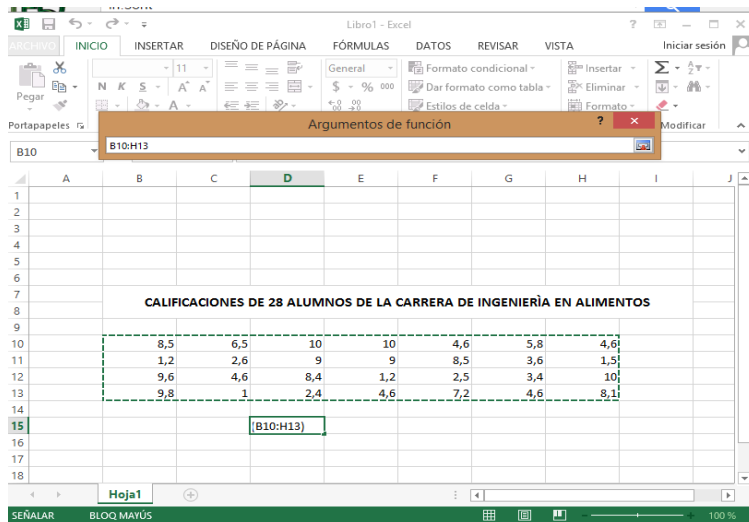
Colocamos los datos agrupados en Excel y seleccionamos el índice de formula (círculo rojo):



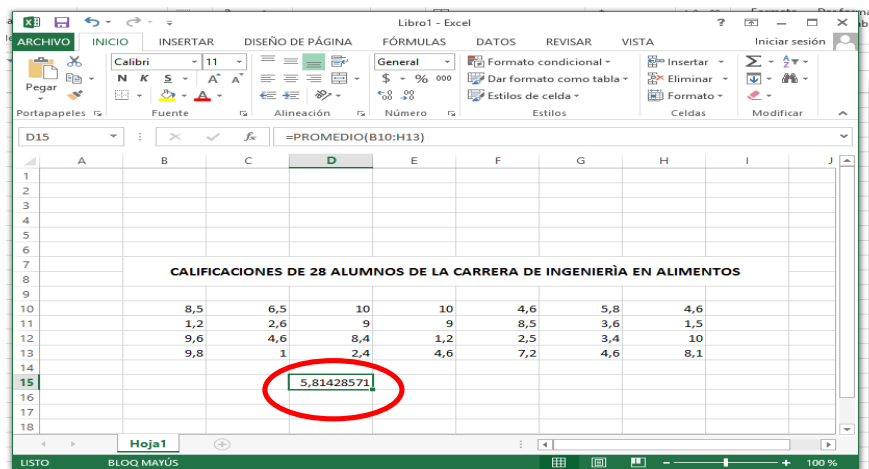
Se desplegará una ventana donde seleccionamos la opción promedio, clic en aceptar.



Posteriormente seleccionamos todos los puntajes de los cuales necesitamos la media aritmética.



Clic en Aceptar

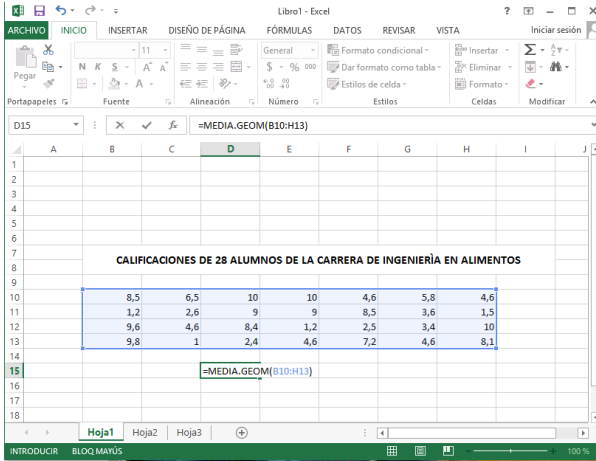


OTROS TIPOS DE MEDIA

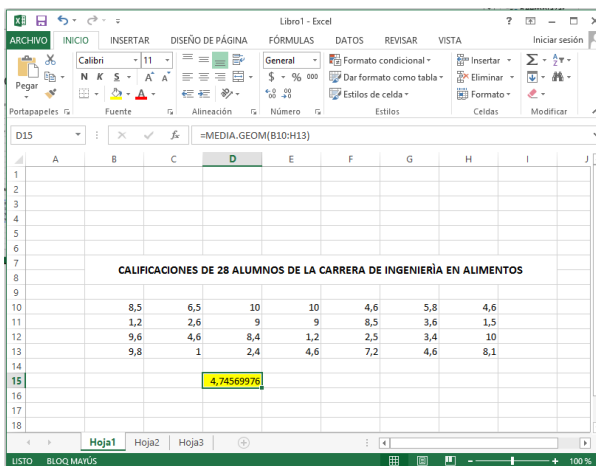
LA MEDIA GEOMETRICA $M_g = M_o$

MEDIA GEOMETRICA EN EXCEL

Colocamos todos los datos a examinar y colocamos en la función el comando =MEDIA.GEOM()



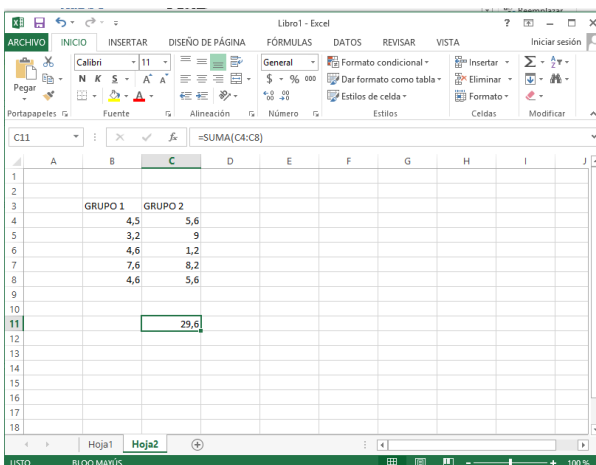
Clic en Enter y obtenemos la media geométrica



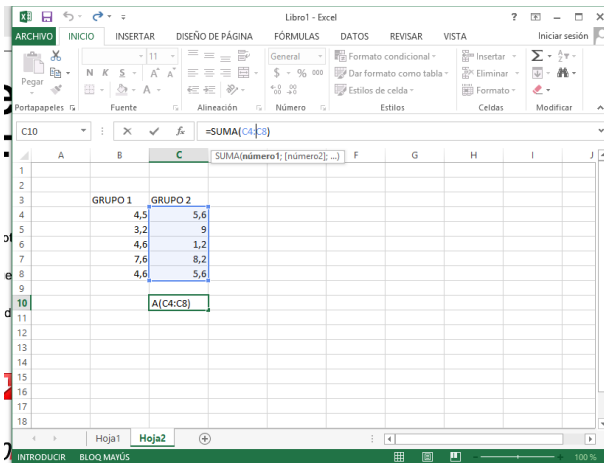
LA MEDIA PONDERADA

Para calcular la media ponderada en Excel procedemos de la siguiente manera:

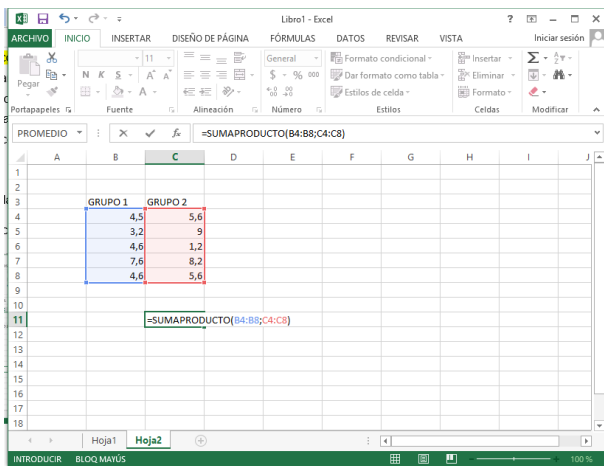
Agrupamos los datos en columnas



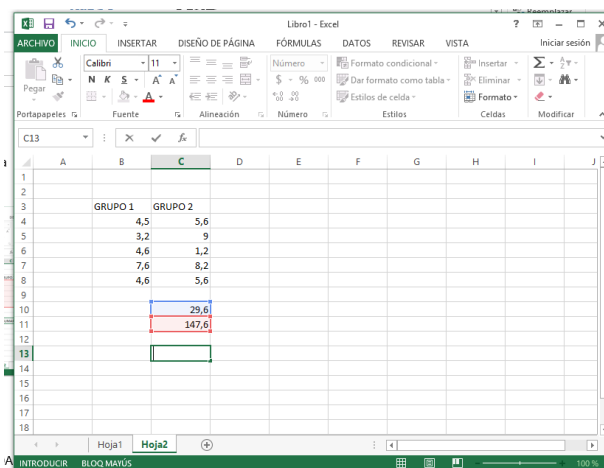
Realizamos la suma de la segunda columna:



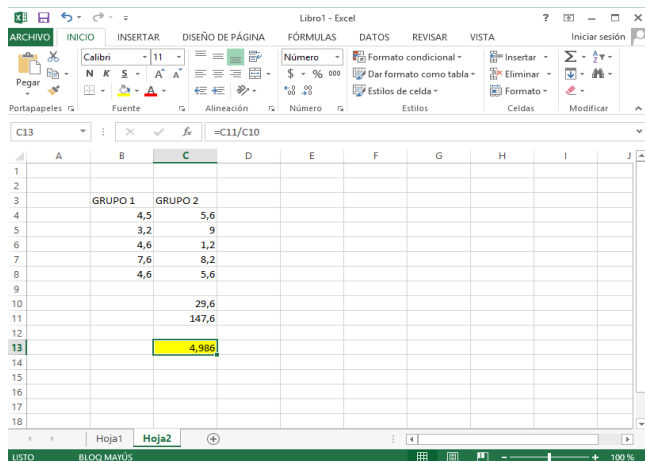
Colocamos la fórmula para la suma total de las dos columnas =SUMAPROMEDIO()



Una vez que tenemos los dos promedios, realizamos una división siempre tomando en cuenta que se dividirá primero la cantidad mayor para la menor.



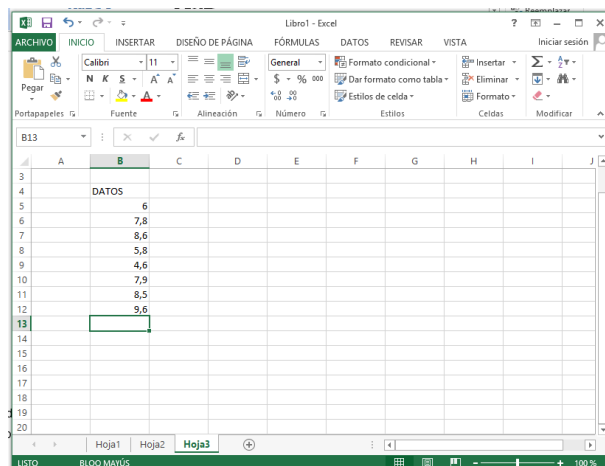
Damos clic en acepta y ese valor será la Media Ponderada



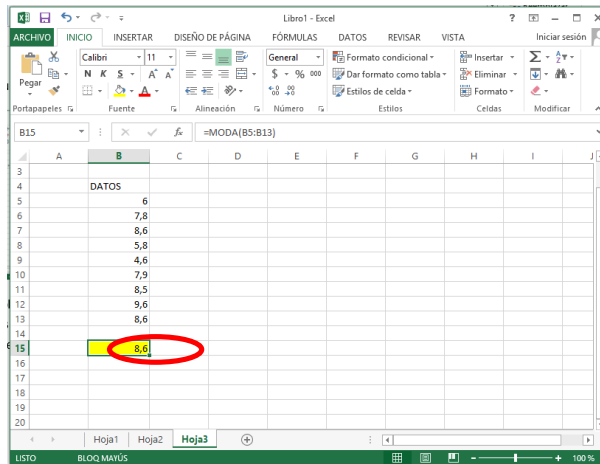
MODO= MODA= VALOR MODAL: Md

CALCULAR MODA EN EXCEL

Colocamos los datos en una columna



En el área de fórmulas colocamos sin las comillas “=MODA()”, entre paréntesis pondremos las columnas desde el primer dato hasta el ultimo separado por un “:”. Clic en Enter

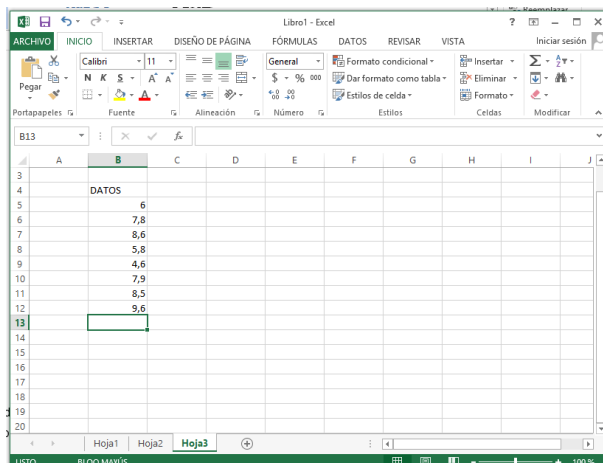


Ese número será la Moda (Número que con mayor frecuencia se repite entre datos)

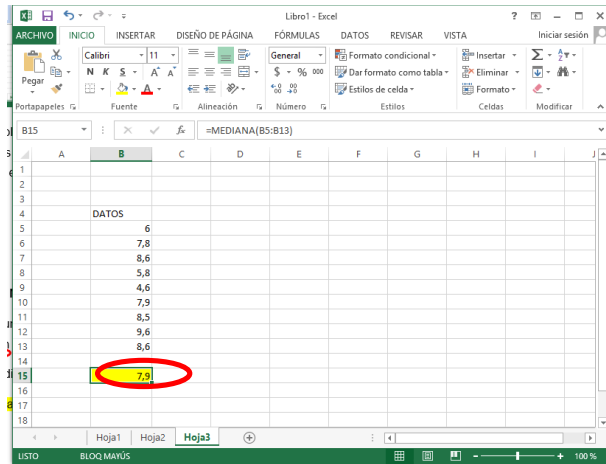
MEDIANA Me

CALCULAR MEDIANA EN EXCEL

Colocamos los datos en una columna

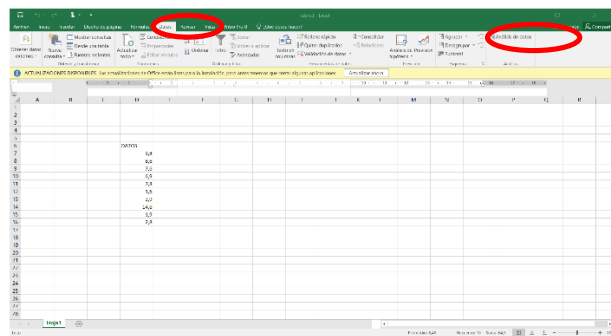


En el área de fórmulas colocamos sin las comillas “=MEDIANA()”, entre paréntesis pondremos las columnas desde el primer dato hasta el ultimo separado por un “:”. Clic en Enter

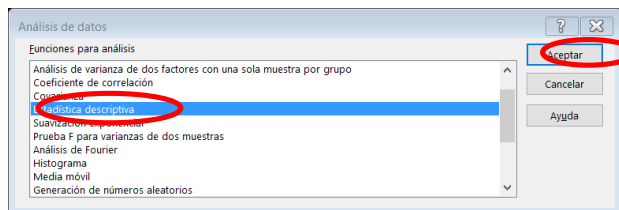


ANALISIS DE DATOS

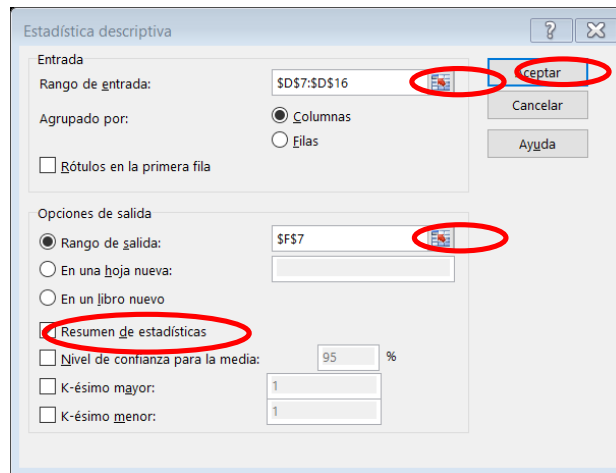
Clic en datos, análisis de datos.



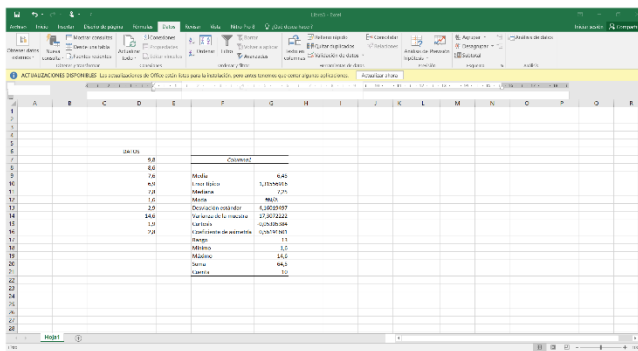
Clic en estadística descriptiva y aceptar.



Clic en rango de entrada y en rango de salida seleccionando respectivamente los datos, resumen de estadísticas, aceptar.

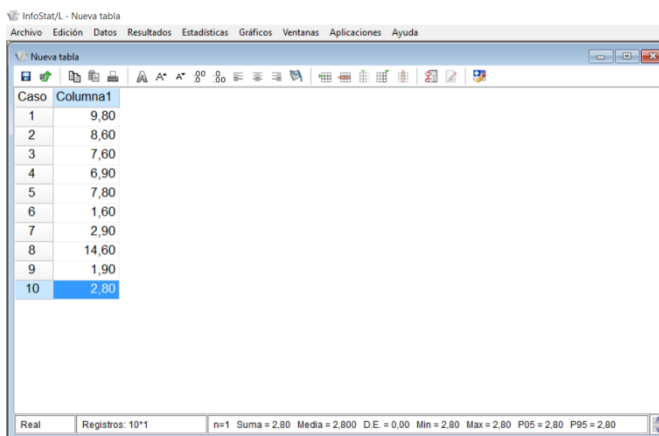


Obtenemos los datos.

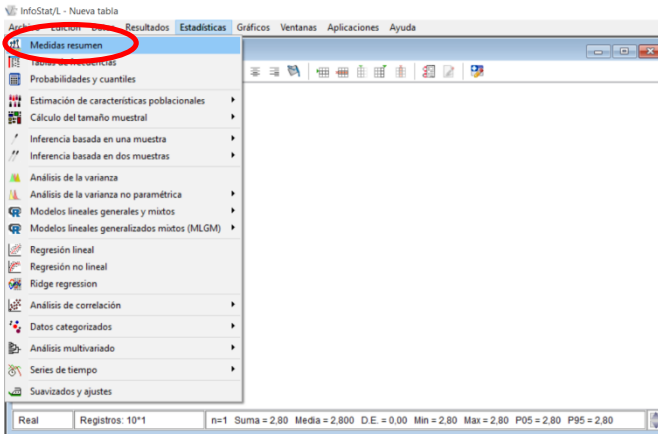


USANDO INFOTAT

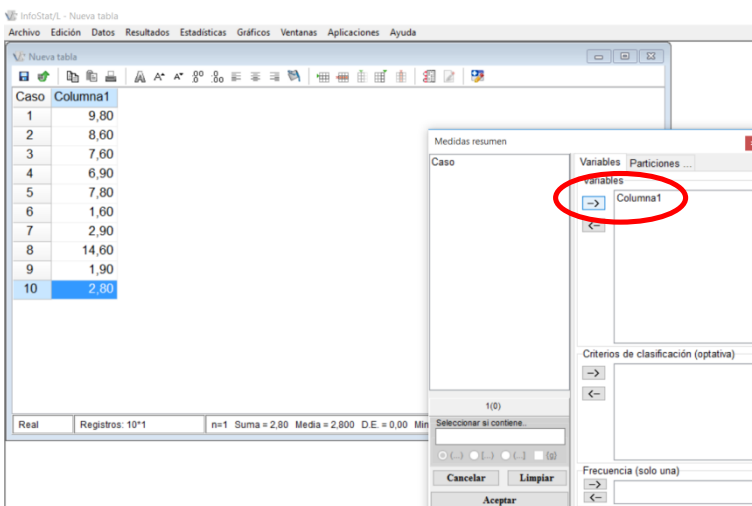
Colocamos los datos en Infostat



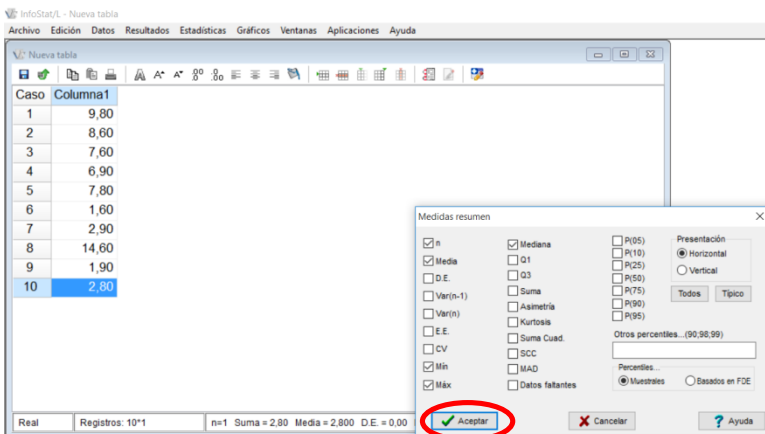
Clic en Estadísticas, Medidas de resumen.



En variable, colocamos el Item de Columna 1



Seleccionamos los indicadores que necesitamos analizar y aceptar



Obtenemos el resultado

InfoStat/L - Nueva tabla

Archivo Edición Datos Resultados Estadísticas Gráficos Ventanas Aplicaciones Ayuda

Nueva tabla

Resultados

Nueva tabla : 2/3/2016 - 10:29:57 - [Versión : 1/11/2014]

Medidas resumen

Variable n	Media	Min	Máx	Mediana
Columna1 10	6,45	1,60	14,60	7,25

MEDIDAS O ESTADIGRAFOS DE DISPERS

4.0 MEDIDAS O ESTADIGRAFOS DE DISPERSIÓN

4.1. DISPERSIÓN O VARIACIÓN

Es el grado en que los datos numéricos tienden a extenderse alrededor de un valor medio a esto se le denomina variación o dispersión de datos. Se utiliza distintas medidas de dispersión, las más empleadas son el rango, la desviación media, y las desviación típica y coeficiente de variación.

4.2. RANGO O AMPLITUD

El rango o amplitud es un estadígrafo simple que permite medir la variabilidad de un conjunto de números, en una distribución se toman los datos extremos, es decir la diferencia entre el mayor y menor valor de todos ellos.

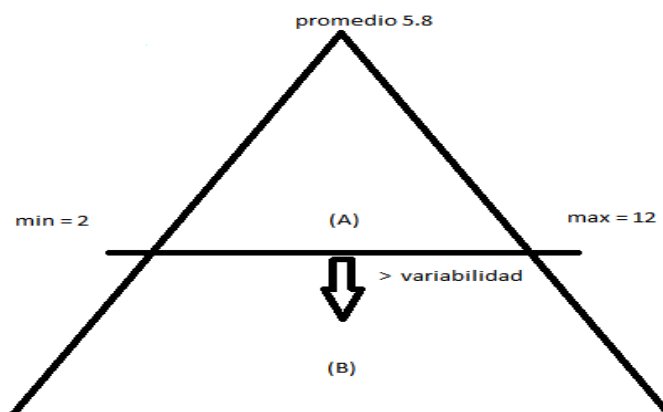
Ejemplo 1:

El rango de los números 2, 3, 3, 5, 5, 5, 8, 10 y, 12

$$R = M - m$$

$$12 - 2 = 10$$

en este caso la variabilidad es 10



Obsérvese que 10 – 12 representan (A), por lo tanto si los valores fueran más extensos representados en (B) nos indicaría una mayor amplitud por lo tanto se expresaría con una mayor variabilidad.

4.3. LA VARIANZA

Al estudiar las medidas de tendencia central, se expresó que "La suma de los cuadrados de las desviaciones de un conjunto de números con respecto a su media aritmética es mínimo". Si esta suma se divide entre el número de observaciones, se obtendrá una cantidad que representará la desviación cuadrada del promedio o media, llamada varianza, lo que es también un mínimo, que juega un papel importante en el muestreo y en la inferencia estadística. La raíz cuadrada de la varianza llamada varianza estándar.

Definición:

Dado un conjunto de observaciones, tales como $X_1, X_2, X_3, \dots, X_N$, la varianza, denotada por la letra griega sigma al cuadrado, se le define como "el cuadrado medio", de las desviaciones con respecto a su media aritmética.

La varianza de la población es definida como:

$$V(X) = \sum_1^n (X_i - \bar{u})^2 / N; \quad u = \frac{X_1 + X_2 + \dots + X_N}{N}; \quad u = \sum_1^n X_i / N$$

donde N: número de variables en la población.

Nota: cuando n es menor que <30, la varianza (varianza muestral) se define como (formula de definición):

$$S^2 = \sum_1^n (X_i - \bar{x})^2 / n-1$$

donde n es el número de variables en la muestra (n-1) es conocido como grados de libertad. Si se conoce el valor de u, el mejor estimado de sigma al cuadrado para una muestra es:

$$S^2 = \sum_1^n (X_j - \bar{u})^2 / n$$

Peso seco en gramos de cinco hojas

g/hoja X_i	$X - \bar{X}$	$(X - \bar{X})^2$	X_i	X_i^2
3	0	0	3	9
4	1	1	4	16
5	2	4	5	25
2	-1	1	2	4
1	-2	4	1	1
$\sum X = 15$	$\sum X_i = 3$	$\sum (X - \bar{X})^2 = 10$	$\sum X_i = 15$	$\sum X_i^2 = 55$

$$s^2 = \sum_{i=1}^n (X_i - \bar{x}) / n - 1$$

$$= (3-3)^2 + (4-3)^2 + (5-3)^2 + (2-3)^2 + (1-3)^2 / 5-1$$

$$= (0)^2 + (1)^2 + (2)^2 + (-1)^2 + (-2)^2 / 4$$

$$= 0 + 1 + 4 + 1 + 4 / 4$$

$$s^2 = \frac{10}{4} = 2.5$$

Para un número mayor de muestra (<30) hay un método corto que es más fácil de hacer, llamada fórmula de trabajo.

$$s^2 = \sum X_i^2 - (\sum X_i)^2 / n -$$

lo que se representa con:

$$s^2 = \sum x_i^2 - \sum \frac{(X_i^2)}{n} / n - 1$$

donde ; $\sum \frac{(X_i^2)}{N} =$ Factor de conversión

$n - 1 =$ Grados de libertad

$$s^2 = \frac{\sum (3^2 + 4^2 + 5^2 + 2^2 + 1^2) - \frac{(\sum (3 + 4 + 5 + 2 + 1))^2}{5}}{5 - 1} = \frac{55 - (\frac{15}{5})^2}{4} = 2.5$$

4.4. DESVIACIÓN TÍPICA (Raíz de la varianza)

De lo expuesto es fácil determinar que la varianza está dada en unidades que son los cuadrados de las unidades originales. La necesidad de tener una medida de dispersión que esta expresada en las mismas unidades originales, ha llevado a utilizar ampliamente la raíz cuadrada de la varianza, la que toma el nombre de desviación estándar, desviación típica o standard deviation:

$$(\sqrt{s^2})$$

Cuando es necesario distribuir la desviación típica de una población de la desviación típica de una muestra:

Definición:

La desviación típica de una serie de N números X_1, X_2, \dots, X_N se representa por s y se define por:

$$s = \sqrt{\sum (xi - \bar{x})^2 / N} ; \sqrt{\sum (x - \bar{x})^2 / N} ; \sqrt{\sum \bar{x}^2 / N} ; \sqrt{(x - \bar{x})^2}$$

donde x representa las desviaciones de cada uno de los números X_i de la media.

Así; s es la raíz cuadrada del cuadrado medio de las desviaciones de la media, o como a veces se le llama raíz del cuadrado medio de las desviaciones.

4.5. DESVIACIÓN MEDIA O PROMEDIO DE DESVIACIÓN

De una serie de N números de datos X_1, X_2, \dots, X_N es posible sacar submuestras las que tendrán su media y su desviación: Dichas estadísticas de \bar{X} y S naturalmente, se hallan sujetas constituyen una población a la variación. Se espera que esta variación va a ser menor que la existente entre observaciones individuales, las cuales representan.

Para la población,

$$\sigma_u^2 = \frac{\sigma^2}{N}; \sigma_u = \sqrt{\sigma^2 / N} = \frac{\sigma}{\sqrt{N}}$$

Para las muestras,

$$s_x^2 = \frac{s^2}{n}; s_x = \sqrt{s^2 / n} = \frac{\sigma}{\sqrt{n}}$$

Supongamos que tenemos la muestra formada por los números 4,6,8,y 10 para los que $S^2 = 6,66$ y $S = 2,58$

$$s_x = \frac{s}{\sqrt{n}}; = \frac{2,58}{\sqrt{4}} = \frac{2,58}{2} = 1,29$$

o también $s_x = \sqrt{s^2 / n} = \sqrt{6,66 / 4} = \sqrt{1,64} = 1,29$

$$\text{Desviación media} = \text{M.D.} = \sum_{i=1}^N [X_i - \bar{X}] / N = \sum [X - \bar{X}] / N$$

Donde \bar{X} es la media aritmética de los números y $\frac{X - \bar{X}}{N}$ es el valor absoluto de las desviaciones de las diferentes X_i de \bar{X} . (El valor absoluto de un número es el mismo número sin asociarle el signo alguno, representado por dos barras verticales a ambos lados del número. Así $[-4] = 4, [+3] = 3, [6] = 6, [-0,84] = 0,84$.

Ejemplo:

$$\bar{X} = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$$

$$\bar{X} = 6$$

$$\begin{aligned} \text{M.D.} = \bar{X} &= \frac{|2-6|+|3-6|+|6-6|+|8-6|+|11-6|}{5} \\ &= \frac{|-4|+|-3|+|0|+|+2|+|+5|}{5} = 2,8 \end{aligned}$$

Si X_1, X_2, \dots, X_k se presentan con frecuencias $f_1, f_2, f_3, \dots, f_k$ respectivamente, la desviación media puede escribirse así;

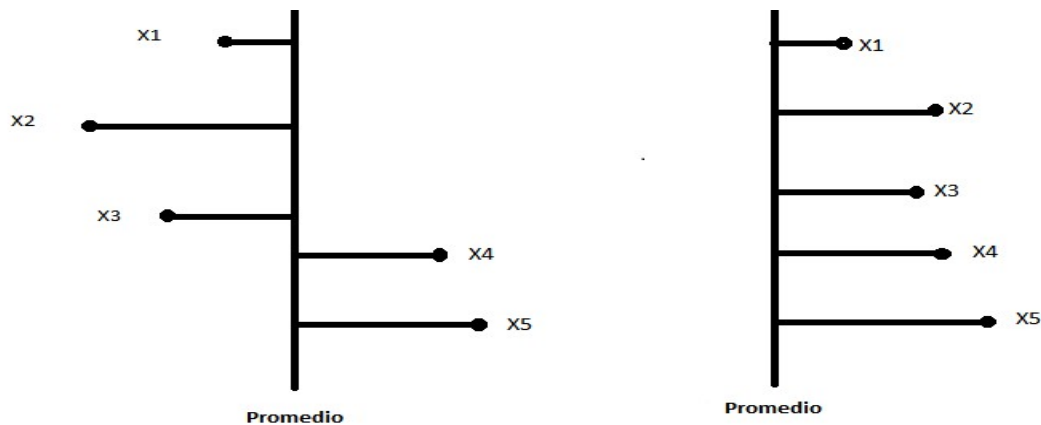
$$\text{M.D.} = \sum_{j=1}^k f_j \frac{|X_j - \bar{X}|}{N} = \sum f \frac{|X - \bar{X}|}{N} = \text{donde } \sum_{j=1}^k f_j = \sum f. \text{ Esta forma es útil}$$

para datos agrupados donde las diferencias X_j representan las marcas de clase. Sin embargo la desviación media se usa muy poco.

Una propiedad interesante de la suma $\sum |x - a|$ es que su valor es

mínimo cuando a es la mediana, es decir, la desviación media respecto de la mediana es mínima. Siendo más apropiado utilizar el término de desviación media absoluta que el de desviación media (Caballero).

Representación gráfica.



Desviaciones de las variables B) Desviaciones absolutas de las variables con respecto a la media con respecto a la media

Figura 4. Representación Gráfica De Las Desviaciones : A) Normal B) Absoluta

4.6 COEFICIENTE DE VARIACIÓN

Conocido también como variación relativa, es el grado de dispersión de dos distribuciones que no vienen dadas en las mismas unidades o que las medias no son iguales se utiliza el coeficiente de variación de Pearson, por lo tanto no se podrán comparar sus varianzas o sus desviaciones típicas entre la media. A medida que se expresen en la misma unidad, pero es importante determinar la variación existente respecto a la base, por lo tanto es multiplicado por cien, para ser expresado en porcentaje. Esto fue propuesto por Karl Pearson (1895), con el objetivo de comparar la variabilidad entre varias distribuciones de frecuencias (Martínez, 2006).

$$CV(\%) = \frac{s}{Media} \times 100$$

Asumiendo que se tiene una desviación estándar de $s = 4.69$ u una media o promedio de la muestra de $\bar{x} = 16.89$

$$CV(\%) = \frac{4.69}{16.89} \times 100 = 27.77$$

Para un número mayor de muestra (>30) hay un método corto que es más fácil de hacer, llamada fórmula de trabajo.

$$s^2 = \frac{\sum Xi^2 - (\sum Xi)^2 / n}{n-1}$$

lo que se representa con:

$$s^2 = \sum xi^2 - \frac{(\sum Xi^2)}{n} / n-1$$

donde ; $\frac{(\sum Xi^2)}{N} =$ Factor de conversión

$n - 1 =$ Grados de libertad

$$s^2 = \frac{(3^2 + 4^2 + 5^2 + 2^2 + 1^2) - \frac{\sum (3+4+5+2+1)^2}{5}}{5-1} = \frac{55 - \left(\frac{15}{5}\right)^2}{4} = 2.5$$

46.1. EJERCICIO PROPUESTOS:

El rango de los numeros 28,2 -19,2- 20,3- 19,4- 20,2- 17,8- 19,8

Para sacar el rango lo primero que debemos identificar los valores Maximos y Minimo de los datos.

M=28,2 m= 10,4

$$\boxed{R = M - m}$$

R= 28,2 - 17,8= 10,4 ► En este caso la variabilidad es 10.4

4.7. LA VARIANZA

Al estudiar las medidas de tendencia central, se expreso que "La suma de los cuadrados de las observaciones de un conjunto de numeros con respecto a su medida aritmetica es minimo" si esta suma se divide entre el numero de observaciones , se obtendra una cantidad que representara la desviacion cuadrada de la media , llamada varianza , lo que tambien es minimo que juega un papel mu importante en el muestreo y en la inferencia estadistica . La raiz cuadrada de la varianza llamada varianza estandar.

Definicion.- dado un conjunto de observaciones a tales como $X_1, X_2, X_3, \dots, X_N$, LA Varaianza , donotada por la letra griega sigma al cuadrado , se lo define como el "el cuadrado medio" de las desviaciones con respecto a su media aritmetica .

La varianaza de la pobblacion es definida como:

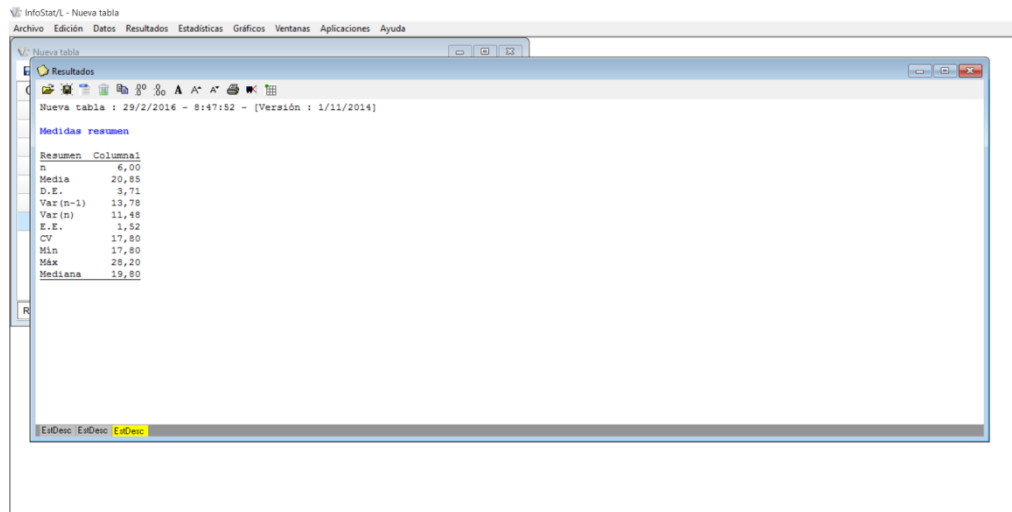
$$V(X) = a_0 + \sum_1^n (x_1 - u)^2$$

Donde N es el numero de variables en la poblacion.

Nota : cuan n es mayor que < 30 , la variaanza (varianaza muestral) se define como (formula definicion)

4.8. USANDO EL EXCEL E INFOTAT

1.-L o primero que realizamos es ingresar los datos en el excel

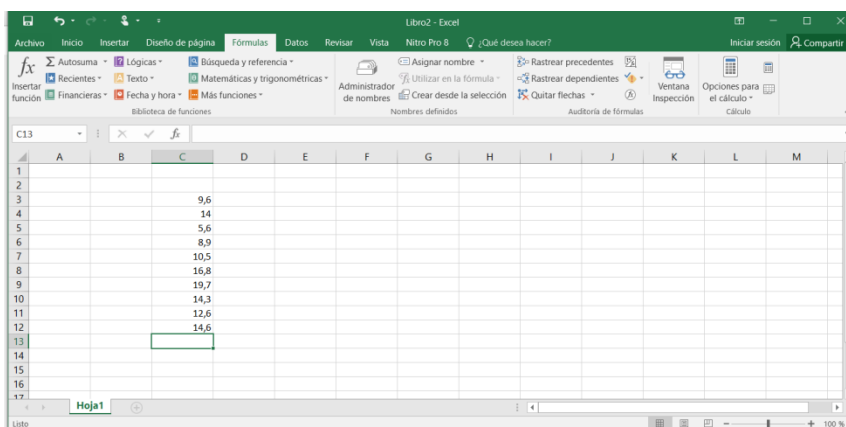


COEFICIENTE DE VARIACION

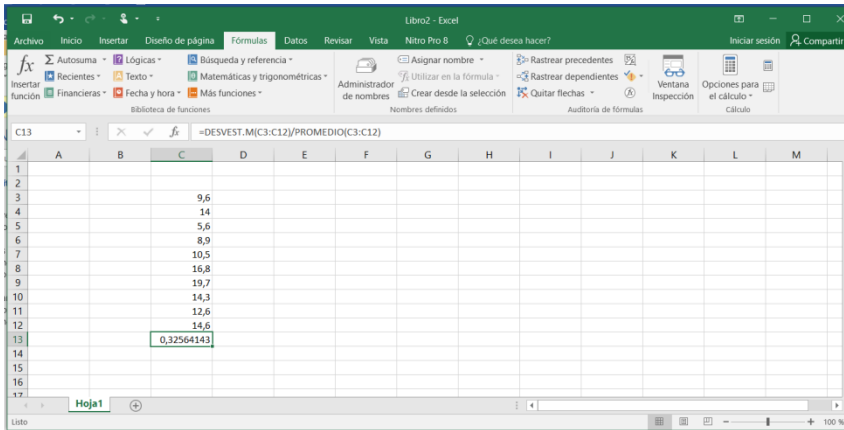
La fórmula para el coeficiente de variación sería

$$CV(\%) = \frac{s}{Media} \times 100$$

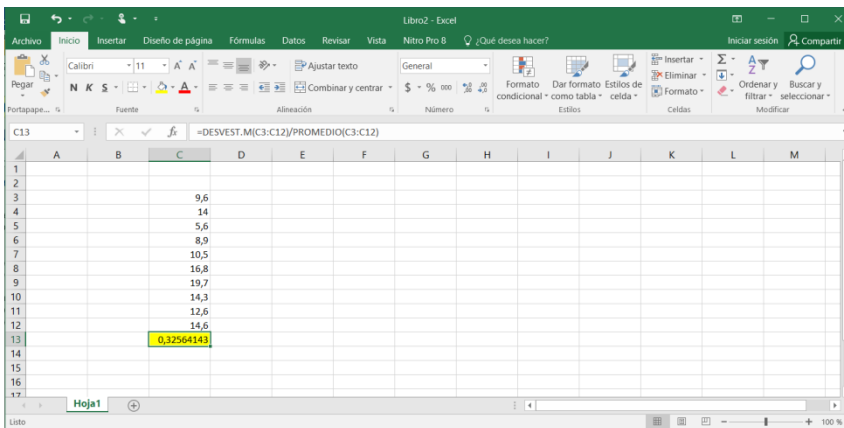
USANDO EXCEL E INFOTAT



En Excel no existe una función directa para sacar el coeficiente de variación por lo tanto debemos colocar la siguiente formula:

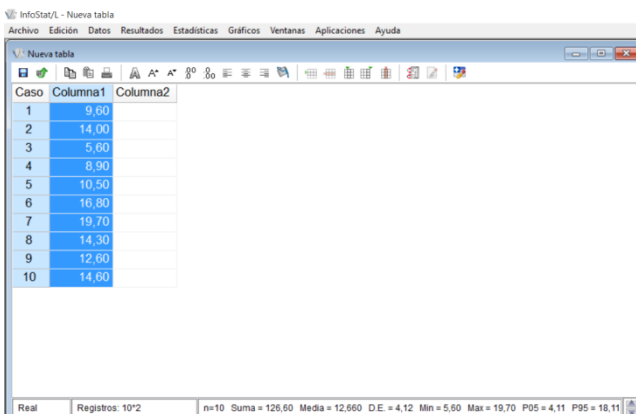


Clic en Aceptar y ese resultado será el Coeficiente de Variación

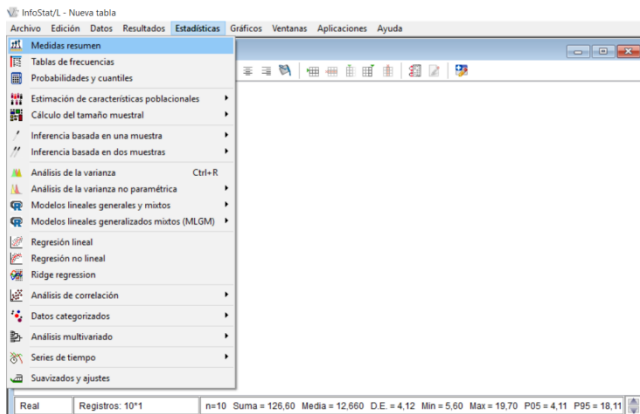


AHORA EN INFOSTAT....

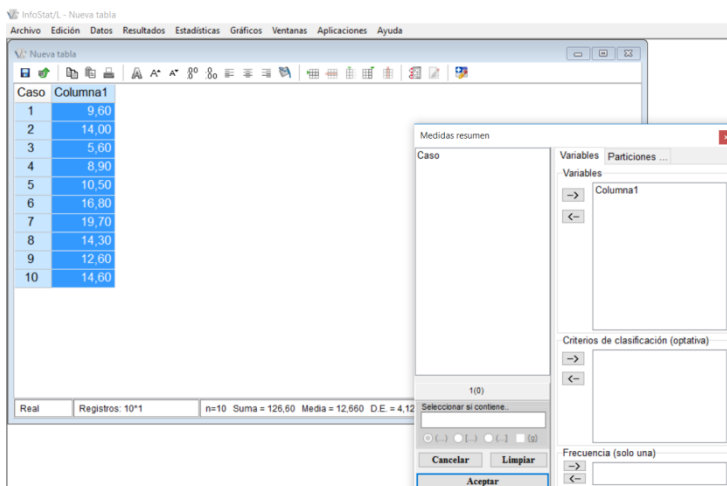
Colocamos Nueva Tabla y agregamos los datos.



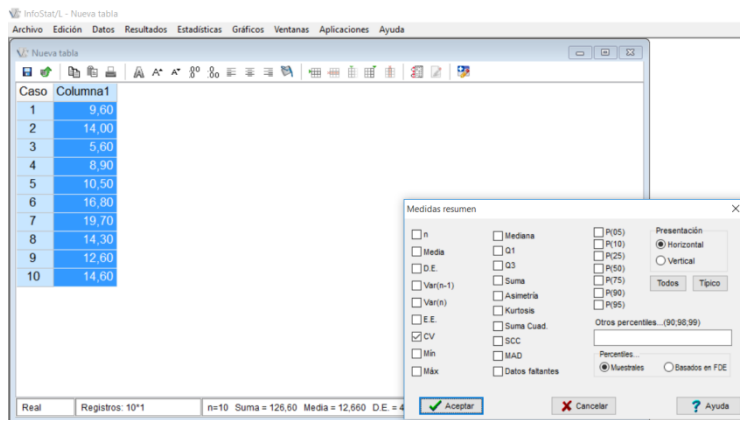
Clic en Estadística, posteriormente seleccionamos Analisis de Varianza



Colocamos en la parte de Variable Los Datos de la Columna 1, clic en aceptar



Seleccionamos el Item que deseemos realizar en este caso CV (Coeficiente de Variación)



Clic en aceptar, y ese resultado será el CV

MEDIDAS DE POSISCIÓN NO CENTRALES

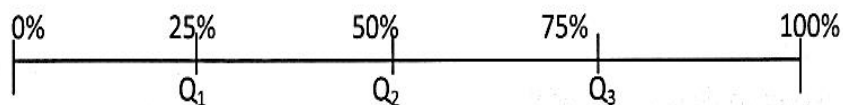
5.0 MEDIDAS DE POSICIÓN NO CENTRALES

Las medidas de posición no central permiten conocer otros puntos característicos de la distribución que no son los valores centrales. Entre otros indicadores, suelen utilizar que se dividen la muestra en partes iguales.

Cuartiles, Deciles y Percentiles

1. Cuartiles (Q)

Son 3 valores que distribuyen la serie de datos, ordenada de forma creciente o decreciente, cuatro tramos iguales en los que cada uno de ellos concentra el 25% de los datos.



Donde:

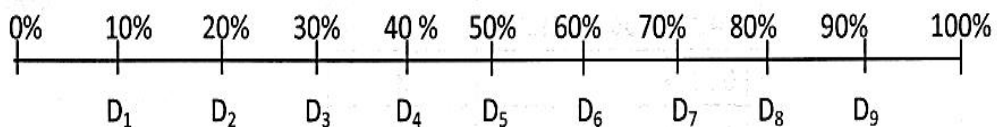
Q₁: 1º Cuartil

Q₂: 2º Cuartil

Q₃: 3º Cuartil

2. Deciles (D)

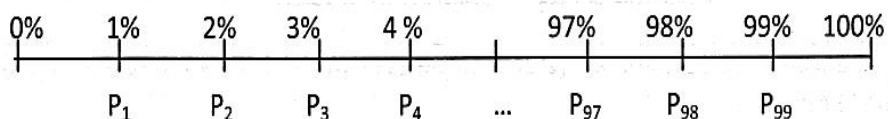
Son 9 valores que distribuyen la serie de datos, ordenada de forma creciente o decreciente, en tramos iguales en los que cada uno de ellos concentra el 10% de los resultados.



3. Percentiles (P).

Son 99 valores que distribuyen la serie de datos, ordenada de forma creciente o decreciente, en tramos iguales en los que cada uno de ellos concentra el 1% de los resultados.

Datos agrupados:



Si una serie de datos se colocan en orden de magnitud, el valor medio (o la media aritmética de los valores medios) que divide al conjunto de datos en dos partes iguales en la mediana. Por extensión, de esta idea se puede pensar en aquellos valores que dividen a los datos en cuatro partes iguales. Estos valores que dividen a los datos en cuatro partes iguales. Estos valores, representados por Q1, Q2 y Q3 se llaman primero, segundo y tercer cuartil, respectivamente; el valor de Q2 es igual al de la mediana. Análogamente, los valores que dividen los datos en diez partes iguales se llaman DECILES y se representan por D1, D2, D3, D4, D5, D9, mientras que los valores que dividen los datos en cien partes iguales se llaman PERCENTILES y se representan por P1, P2, P3, P4, P5, P99. El quinto decil y el quimenogésimo percentil se corresponde con la mediana. Y las percentiles P25 y P75 pueden con el primer y tercer cuartil respectivamente.

En conjuntos, cuartiles, deciles, percentiles y otros valores obtenidos por subdivisiones análogas de los datos se llaman cuarteles.

44. hallar a los cuartiles Q1, Q2 y Q3 de la siguiente distribución de frecuencias.

Salarios (Dólares)	Número de empleados
\$ 50 -- 59,99	8
\$ 60 -- 69,99	10
\$ 70 -- 79,99	16
\$ 80 -- 89,99	14
\$ 90 -- 99,99	10
\$ 100 - 109,99	5
\$ 110 - 119,99	2
	Total 65

El primer cuartil. Es aquel valor que se obtiene de $N/4 = 65/4 = 16,25$ de los casos comenzando por la clase primera. Puesto que la primera clase contiene 8 casos, tomaremos 8,25 casos ($16,25 - 8$) de los 10 que tiene la segunda clase. Por el método de interpolación lineal se tiene:

$$Q1 = 59,995 + 8,25 (\$ 10,00) = 10$$

$$Q1 = \$ 68,25$$

El segundo cuartil Q2 es aquel valor que se obtiene para el $2N/4 = N/2 = 65/2 = 32,5$ de los casos. Puesto que las dos primeras clases comprenden 18 casos tomaremos $32,5 - 18 = 14,5$ de las 16 casos de la tercera clase; entonces

$$Q2 = 69,945 + 14,5 (\$ 10,00) = \$ 79,06.$$

16

Este cuartil Q2 es realmente la mediana.

El tercer cuartil Q3 es aquel valor que obtiene para $3N/4 = 3 (65) = 48,75$ de los 4 casos. Puesto que las cuatro primeras clases comprenden 48 casos tomaremos $48,75 - 48 = 0,75$ de los 10 casos de la quinta clase; entonces:

$$Q3 = \$ 89,995 + 0,75 (\$ 10,00) = \$ 90,75.$$

10

De aquí que el 25% de los empleados ganan \$ 68,25 o menos, el 50% gana \$ 79,06 o menos y el 75% gana \$ 90,75 o menos.

b). Hallar los deciles. El primero, segundo,... Noveno deciles se obtienen para $N/10, 2N/10, \dots, 9N/10$ comenzando por la primera clase.

$$D1 = \$ 49,995 + 6,5 (\$ 10,00) = \$ 58,12 \rightarrow 10\%$$

8

$$D2 = \$ 59,995 + 5 (\$ 10,00) = \$ 65 \rightarrow 20\%$$

10

$$D3 = \$ 69,995 + 1,5 (\$ 10,00) = \$ 70,94 \rightarrow 30\%$$

16

$$D4 = \$ 69,995 + 8 (\$ 10,00) = \$ 75,00 \rightarrow 40\%$$

16

$$D5 = \$ 69,995 + 14,5 (\$ 10,00) = \$ 79,06 = \text{Mediana}$$

16

$$D6 = \$ 79,995 + 5 (\$ 10,00) = \$ 83,57 \rightarrow 60\%$$

14

$$D7 = \$ 79,995 + 11,5 (\$ 10,00) = \$ 88,21$$

14

$$D8 = \$ 89,995 + 4 (\$ 10,00) = \$ 94,00$$

10

$$D9 = \$ 79,995 + 5 (\$ 10,00) = \$ 83,57 \rightarrow 90\%$$

14

45. Determinar (a) el percentil (b) el percentil 60 para la distribución del problema anterior.

$$a). 35N/100 = (35) (65)/100 = 22,75 (22,75 - 18 = 4,75).$$

$P35 = \$ 69,995 + 4,75 (\$ 10,00) = \$ 72,95$ esto significa que el 35% de los

16

Empleados gana \$ 72,97 o menos.

$$b). 60N/100 = 65 \times 65/100 = 39,0 (39 - 34 = 5).$$

$P60 = \$ 79,995 + 5 (10,00) = \$ 83,57$ este valor es el mismo que se obtiene con el 14 sexto decil o tercer quintil

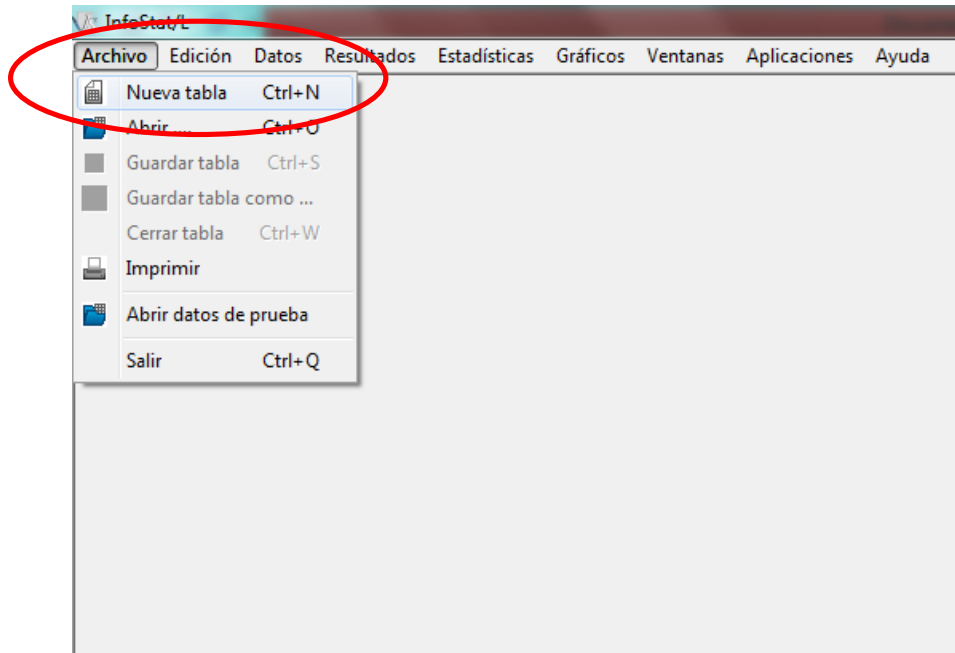
5.1. USANDO INFOSTAT

1er paso, en una hoja de Excel plasmamos nuestros datos a evaluar.

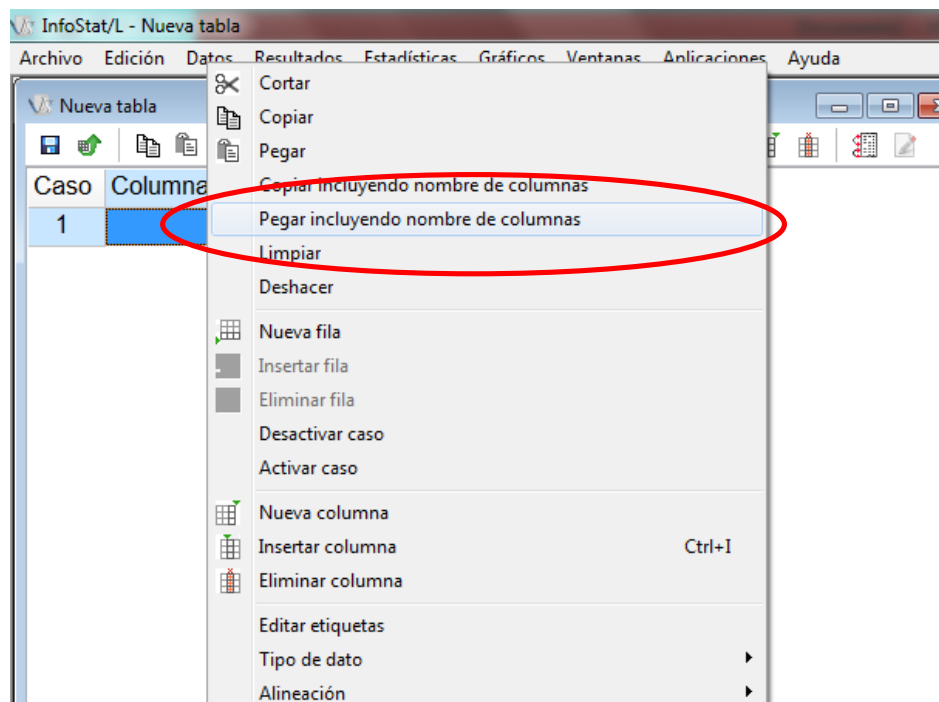
The screenshot shows the Microsoft Excel interface. The ribbon includes 'Archivo', 'Inicio', 'Insertar', 'Diseño de página', 'Fórmulas', 'Datos', 'Revisar', 'Vista', and 'Prueb'. The 'Inicio' ribbon is active, showing options for 'Portapapeles' (Cut, Copy, Paste, Copy format), 'Fuente' (Font: Calibri, size 11, bold, italic, underline, color, background color), and 'Alineación' (Text alignment, Orientation, Merge cells). The active cell is A1, containing the text 'Variable'. The data table is as follows:

	A	B	C	D	E	F	G
1	Variable						
2	1,40						
3	1,41						
4	1,42						
5	1,43						
6	1,44						
7	1,45						
8	1,46						
9	1,47						
10	1,48						
11	1,49						
12	1,50						
13							

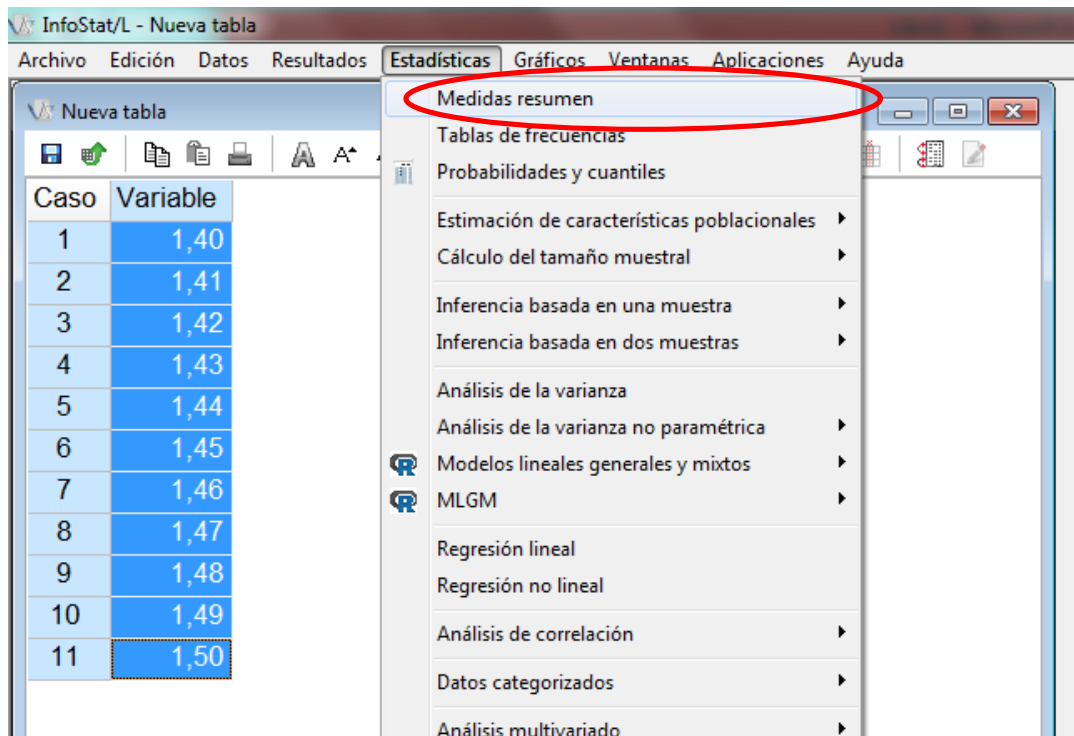
2do Paso, ingresamos al Programa Infostat y damos clic en archivo y posterior en nueva tabla.



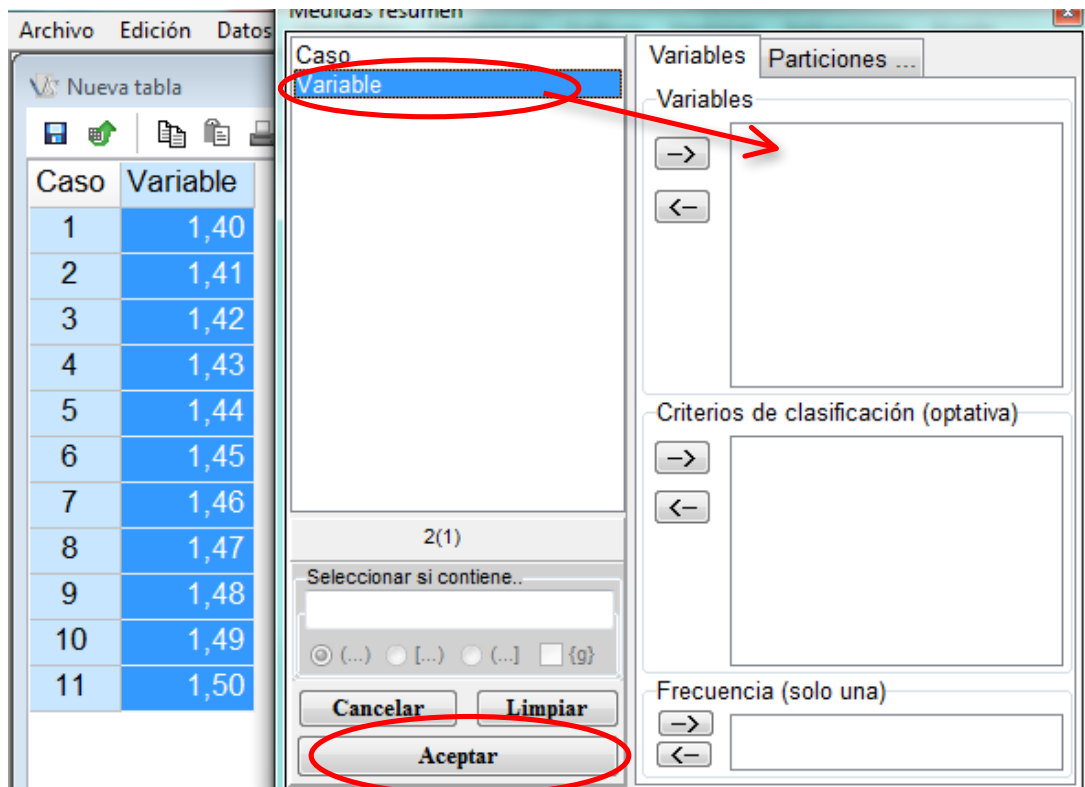
3er paso, copiamos y pegamos nuestros datos en la tabla de infostat.



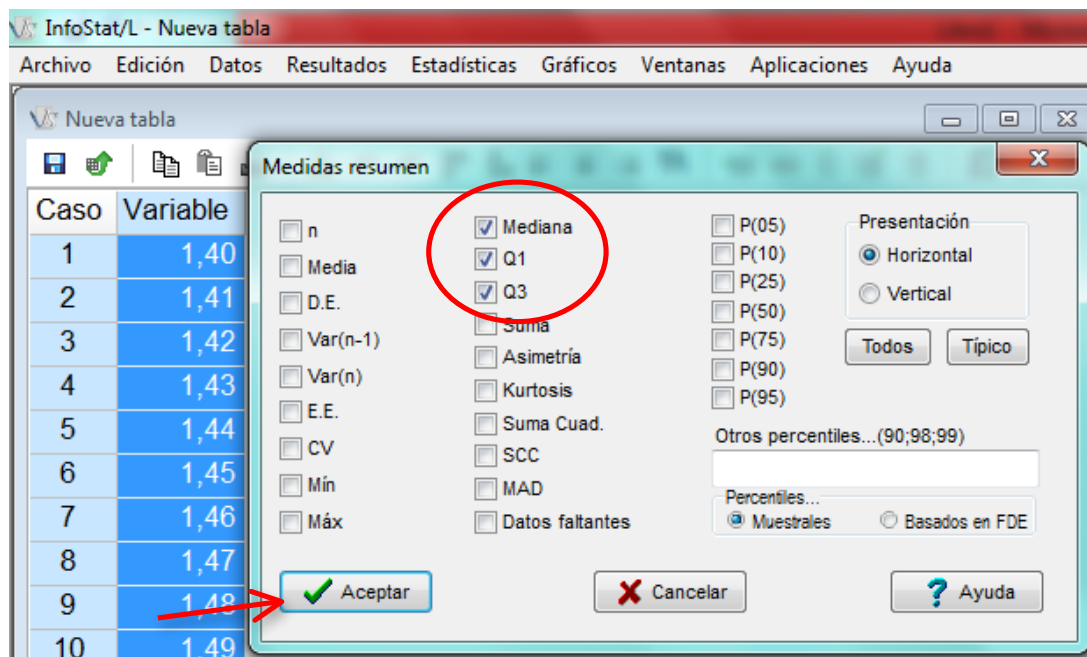
4to paso, damos clic en la opción estadística y posteriormente en medidas de resumen.



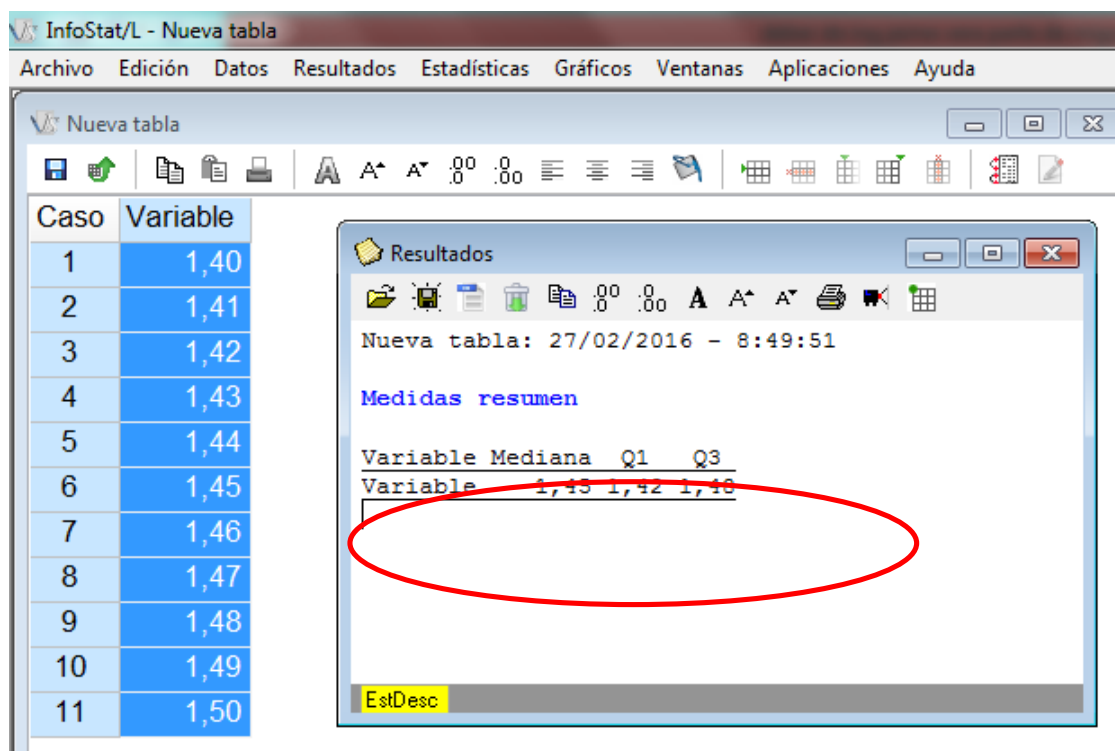
5to paso, dentro de la ventana medidas de resumen pasamos nuestras variables al cuadro perteneciente de variables y damos clic en aceptar.



6to paso, ponemos en selección lo que deseamos y damos clic en aceptar.



7mo paso, finalmente obtenemos los resultados, donde **Q1** significa Cuartil 1, **Mediana** significa Cuartil 2, **Q3** significa Cuartil 3.



PROBABILIDAD DE OCURRENCIA DE UN SUCESO

6.0 PROBABILIDAD DE OCURRENCIA DE UN SUCESO

Cuando atendemos el parto de una vaca, el nacimiento de un ternero es un suceso, número de frutos sanos cosechados, mientras que el nacimiento de una ternera es otro suceso. En la teoría de probabilidades, un suceso es uno o más de los posibles resultados de una actividad cualquiera.

Todos estamos familiarizados de una manera empírica con el concepto de probabilidad de ocurrencia de un suceso. Por ejemplo, si hemos planificado disfrutar con nuestra familia de un hermoso domingo en una agradable playa, seguramente cambiaremos nuestros planes luego de escuchar un parte meteorológico que pronostica un 70 % de probabilidad de lluvia para ese día.

6.1. PRINCIPIOS DE PROBABILIDAD

En cualquier experimento aleatorio existe siempre incertidumbre respecto a si un determinado acontecimiento ocurrirá o no. Como una medida de la posibilidad, o probabilidad, con la cual puede esperarse que ocurra el evento suele emplearse un número entre 0 y 1, pero si se está seguro de que ocurra un determinado evento, se dice que su probabilidad es de 100%, o bien 1, pero si se está seguro de que no ocurra, se dice que su probabilidad es cero. Si por ejemplo, la probabilidad es $\frac{1}{4}$, se dice que hay una posibilidad de 25% de que ocurra y una posibilidad de 75% de que no ocurrirá. De manera equivalente, puede decirse que las posibilidades en contra de su ocurrencia son de 75% contra 25%, o bien 3 a 1.

6.1.1. ESPACIO MUESTRALES

A un conjunto S que consta de todos los resultados posibles de un experimento aleatorio se le llama espacio muestral, y a cada resultado se le llama punto muestral. A menudo habrá más de un espacio muestral que describa los resultados de un experimento, pero, en general, solamente habrá uno que proporcione la mayor información.

Entonces un espacio muestral es el conjunto formado por todos los posibles resultados de un experimento aleatorio. El espacio muestral se denota con la letra S y será finito o infinito en dependencia de que el conjunto tenga un número finito o infinito de elementos.

Ejemplo

Si se lanza un dado, un espacio muestral, o conjunto de todos los resultados posibles, será $\{1, 2, 3, 4, 5, 6\}$, mientras que otro será {par, non}. Sin embargo, es claro que el último no será adecuado para determinar, por ejemplo, si un resultado es divisible entre 3.

A menudo es útil representar gráficamente un espacio muestral. En tales casos, siempre que sea posible, es deseable utilizar números en vez de letras

6.1.2. EXPERIMENTO ALEATORIO

Para todos es conocida la importancia de los experimentos en la ciencia y en la ingeniería. La experimentación es útil porque permite suponer que si se realiza un determinado experimento bajo condiciones esencialmente idénticas, se llegara a resultados básicamente iguales. En estas circunstancias, puede controlarse el valor de la variable que afecta el resultado de un experimento.

Sin embargo, en algunos experimentos no es posible conocer o controlar el valor de ciertas variables, de modo que los resultados variaran de una a otra realización del experimento, aun cuando la mayor parte de las condiciones sean iguales. Estos experimentos se describen como aleatorios.

La actividad que da lugar a la ocurrencia de un suceso se llama experimento. Entonces un experimento se considera aleatorio cuando su resultado no se puede predecir con toda exactitud. Como ejemplo de algunos experimentos aleatorios podemos citar, el estudio de la precipitación ocurrida en una etapa, el análisis del rendimiento de un cultivo, el estudio del comportamiento de las ventas de un alimento, etc.

Una característica importante de los experimentos aleatorios es la capacidad de los mismos para poder ser repetidos. Si un experimento aleatorio se repite un gran número de veces, la frecuencia de un resultado tiende a ser constante, y decimos entonces que el resultado muestra una regularidad estadística.

6.1.3. LA DISTRIBUCIÓN T DE STUDENT

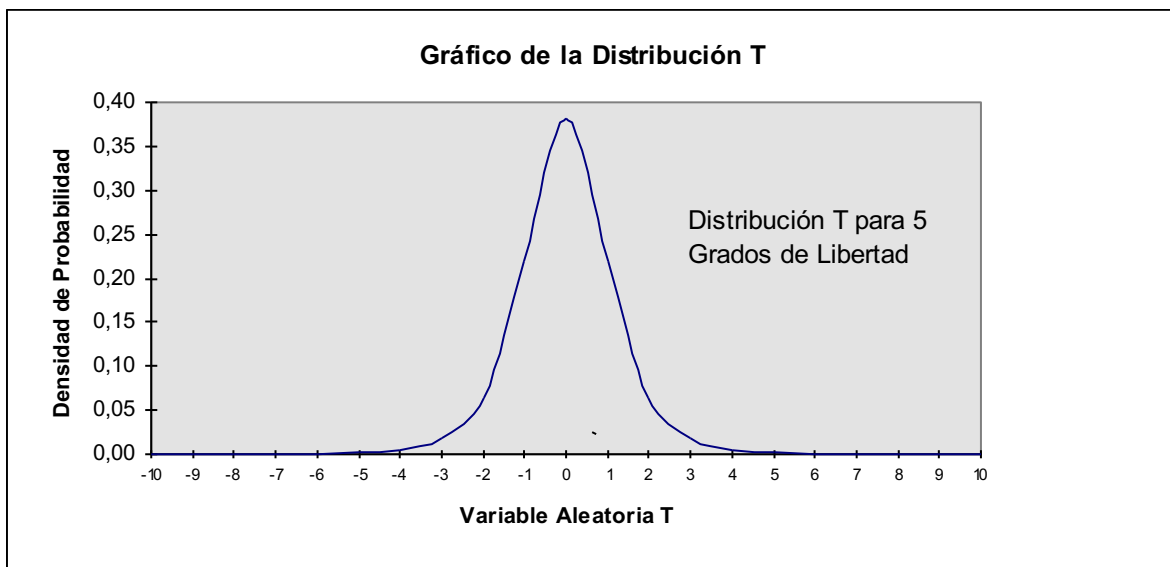
En la generalidad de los casos, no disponemos de la desviación standard de la población, sino de una estimación calculada a partir de una muestra extraída de la misma y por lo tanto no podemos calcular Z. En estos casos calculamos el estadístico T:

$$T = \frac{\mu - x}{S}$$

donde S es la desviación standard muestral, calculada con $n-1$ grados de libertad:

$$S = \sqrt{\frac{\sum (\bar{X} - X_i)^2}{n-1}}$$

El estadístico T tiene una distribución que se denomina *distribución T de Student*, que está tabulada para 1, 2, 3, ... etc. grados de libertad de la muestra con la cual se calculó la desviación standard. La distribución T tiene en cuenta la incertidumbre en la estimación de la desviación standard de la población, porque en realidad la tabla de T contiene las distribuciones de probabilidades para distintos grados de libertad:



Para un número de grados de libertad pequeño, es más ancha que la distribución normal tipificada. Cuando los grados de libertad tienden a infinito, la distribución T tiende a coincidir con la distribución normal standard. Es decir, en la medida que aumentemos el número de observaciones de la muestra, la desviación standard calculada estará más próxima a la desviación standard de la población y entonces la distribución T correspondiente se acerca a la distribución normal standard. El uso de la distribución T presupone que la población con que estamos trabajando tiene una distribución normal.

Distribución de Promedios Muestrales

Para comprender que significa distribución de promedios muestrales, vamos a suponer que realizamos un experimento con bolilleros como los usados en la lotería. Colocamos un número muy grande de bolillas blancas en un bolillero blanco, en cada una de las cuales figura un dato X . Este bolillero representa la población de observaciones X , y tiene media μ y varianza σ^2 . Supongamos que a continuación hacemos lo siguiente:

- 1) Tomamos una muestra de $n = 10$ bolillas blancas
- 2) Calculamos la media \bar{X} y la anotamos en una bolilla azul.
- 3) Colocamos la bolilla azul en un segundo bolillero de color azul.
- 4) Devolvemos las bolillas blancas a su bolillero y le damos vueltas.
- 5) Repetimos toda la operación muchas veces hasta que el bolillero azul esté lleno de bolillas.

Entonces, los números del bolillero azul forman una *población de promedios muestrales* \bar{X} . Esta es una población derivada de la anterior, y tiene la misma media o promedio que la distribución original, pero su varianza es un enésimo de la varianza de la distribución original:

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

En el caso del bolillero azul, si denominamos σ_m^2 a la varianza y μ_m a la media tenemos:

$$\mu_m = \mu$$

$$\sigma_m^2 = \frac{\sigma^2}{10}$$

La distribución de medias muestrales está situada en el mismo lugar (alrededor de la misma media) que la distribución original, pero es mucho más angosta, porque su varianza es la décima parte de la varianza original. La distribución original de observaciones representada por el bolillero blanco se denomina comúnmente *distribución madre* o *base*. Es obvio que, dada una distribución cualquiera, podemos obtener una distribución de promedios de muestras de 2 observaciones, de 3 observaciones, ...etc.

Al construir la población de promedios muestrales, realizábamos extracciones de 10 bolillas blancas después de dar vueltas al bolillero. Es decir, que estábamos realizando un *muestreo aleatorio* de la población madre, porque cada una de las bolillas blancas tenía la misma posibilidad de ser elegida para integrar la muestra. Aunque la población original no sea de distribución normal, si el muestreo es aleatorio, la población de promedios muestrales *se aproximará a la normalidad*, es decir, será *casi de distribución normal*. Este efecto se debe a un teorema de estadística matemática denominado Teorema Central del Límite. En resumen, si se cumple la hipótesis de muestreo aleatorio, tenemos:

	Distribución madre de X	Distribución muestral de \bar{X}
Media	μ	μ
Varianza	σ^2	$\frac{\sigma^2}{n}$
Desviación Standard	σ	$\frac{\sigma}{\sqrt{n}}$
Forma de la distribución	cualquiera	más cercana a la distribución normal que la distribución madre

En general, en los problemas que se presentan habitualmente, existe una población de observaciones cualesquiera, de la cual tomamos una muestra aleatoria, es decir, un subconjunto de observaciones elegidas al azar, por medio de la cual intentamos conocer todo lo que sea posible acerca de la población de la cual fue extraída. El promedio de la muestra de n elementos pertenece a la distribución de promedios muestrales de la población original. Es decir, que el promedio de la muestra que obtuvimos es uno de los muchos promedios muestrales que se distribuyen alrededor de μ con desviación standard $\frac{\sigma}{\sqrt{n}}$. Por lo

tanto, si la muestra es más grande (n mayor), estaremos en una distribución de promedios con desviación standard mas pequeña, por lo cual, el promedio de la muestra estará mas cerca del promedio del universo. Es por esto que es razonable pensar que el promedio de la muestra es una *estimación* del promedio del universo.

Distribución muestral de la suma y de la resta:

Muchas veces es importante conocer la distribución de la suma Y de dos variables aleatorias independientes y_A e y_B . Supongamos que:

y_A tiene una distribución con media μ_A y varianza σ_A^2

y_B tiene una distribución con media μ_B y varianza σ_B^2

¿Qué se puede decir de la media y la varianza de la distribución de $Y = y_A + y_B$? De nuevo se puede ilustrar el problema considerando dos bolilleros, cada uno con su población apropiada de bolillas. Imaginemos que hacemos una extracción aleatoria del bolillero A, para obtener y_A , y del bolillero B, para obtener y_B , sumamos los valores, escribimos la suma $Y = y_A + y_B$ en una bolilla roja y la introducimos en un tercer bolillero. Después de repetir esto muchas veces ¿Qué puede decirse de la distribución de las sumas que están en las bolillas rojas del tercer bolillero? Se puede demostrar que la media de la suma Y es:

$$\mu_S = \mu_A + \mu_B$$

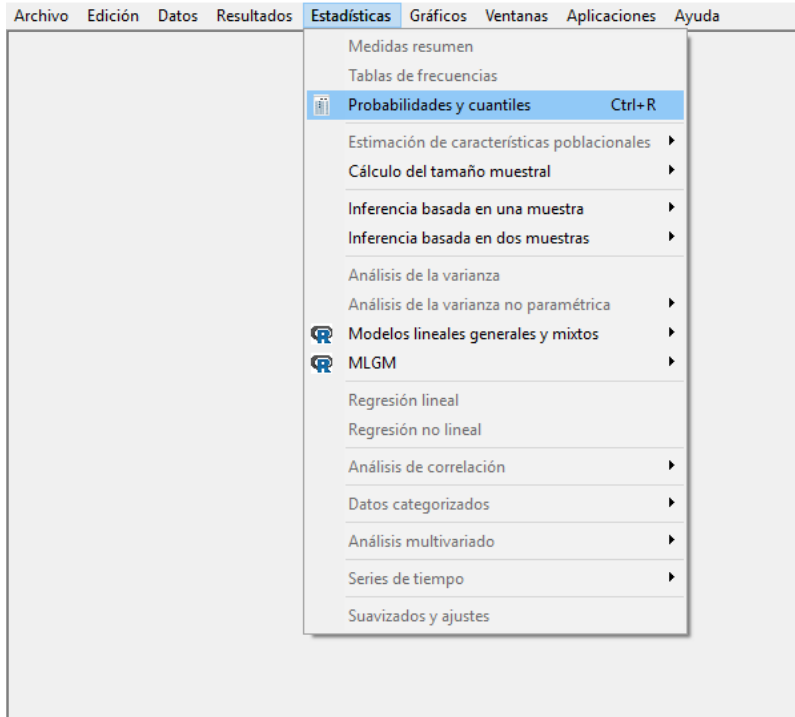
y la varianza de la suma Y es:

$$VAR(Y) = VAR(y_A + y_B) = VAR(y_A) + VAR(y_B) = \sigma_A^2 + \sigma_B^2$$

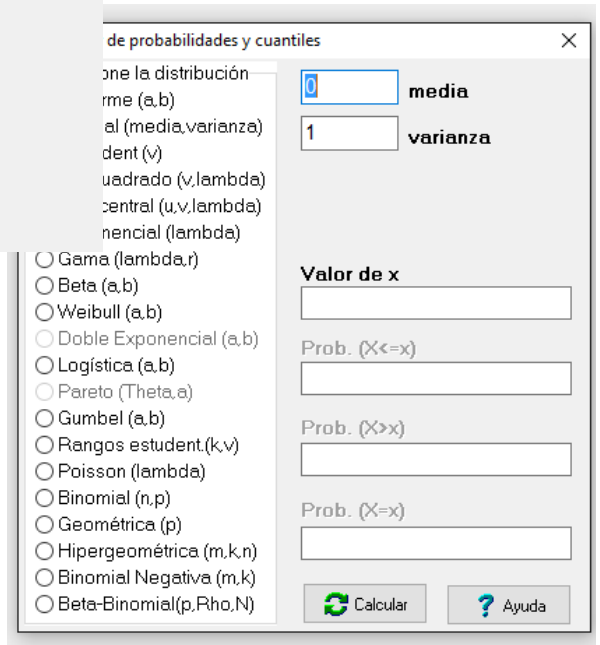
De la misma manera, para la resta o diferencia de dos variables $Y = y_A - y_B$, resulta que la media de la diferencia es:

Cálculo de probabilidades

El porcentaje promedio de humedad de frutos frescos sigue una distribución normal, con un promedio de 20% con una varianza de 5%. ¿Cuál es la probabilidad de que la humedad mida más de 21.2%? ¿y menos de 21.2%? ¿y exactamente 21.2%?



Distribución



Calculador de probabilidades y cuantiles

Seleccione la distribución

- Uniforme (a,b)
- Normal (media,varianza)
- T Student (v)
- Chi Cuadrado (v,lambda)
- F no central (u,v,lambda)
- Exponencial (lambda)
- Gama (lambda,r)
- Beta (a,b)
- Weibull (a,b)
- Doble Exponencial (a,b)
- Logística (a,b)
- Pareto (Theta,a)
- Gumbel (a,b)
- Rangos estudent.(k,v)
- Poisson (lambda)
- Binomial (n,p)
- Geométrica (p)
- Hipergeométrica (m,k,n)
- Binomial Negativa (m,k)
- Beta-Binomial(p,Rho,N)

20 **media**

5 **varianza**

Valor de x

21.2

Prob. ($X \leq x$)


0,7042471589


Prob. ($X > x$)

0,2957528411

Prob. ($X = x$)

0

 **Calcular**

 **Ayuda**

Valor de media

Varianza

Valor de X

Calcular

PRUEBA DE HIPÓTESIS.

7.0. PRUEBA DE HIPÓTESIS.

En el desarrollo ulterior de los métodos estadísticos y en el quehacer diario de la Ciencia y la Tecnología, se presenta con mucha frecuencia la necesidad de tomar una decisión con relación al valor de un parámetro de una población determinada.

Los problemas que involucran una toma de decisión pueden, por regla general, reducirse a un procedimiento que implique el rechazo o la aceptación de una hipótesis o suposición sobre el valor de la media o la proporción poblacional de una distribución.

El procedimiento estadístico que da solución a este tipo de problemática se conoce como **PRUEBA DE HIPÓTESIS**, el cual pasamos a describir a continuación.

Con el objetivo de ilustrar el problema, iniciaremos la descripción del procedimiento mediante una prueba de hipótesis sencilla. Supongamos que se desea tomar una decisión sobre el valor de una media poblacional de la cual conocemos que solo puede tomar los valores 14 o 19. Una formulación de hipótesis para este caso puede ser expresada de la siguiente manera :

$$H_0: \mu = 14 \text{ y } H_1: \mu = 19$$

En términos de una Prueba de Hipótesis, a H_0 se le conoce como la **hipótesis nula** y a H_1 como la **hipótesis alternativa**.

Como vimos anteriormente, decidir si la media poblacional es igual a 14 o igual a 19, se reduce a tomar una de las dos siguientes decisiones :

- ◆ Rechazar H_0 , lo cual implica aceptar H_1 y en consecuencia decidir que $\mu = 19$
- ◆ Aceptar H_0 , lo cual determina que $\mu = 14$

Pero la estadística no es una ciencia exacta, y en consecuencia, cualquiera sea la decisión que tomemos podemos cometer dos tipos de errores :

- ◆ Si rechazamos H_0 y realmente esta hipótesis es verdadera, cometemos el error conocido como **Error de Tipo I**
- ◆ Si aceptamos H_0 y en realidad esta hipótesis es falsa, cometemos el error conocido como **Error de Tipo II**

En términos probabilísticos, a la probabilidad de cometer el Error de Tipo I, se le denomina **nivel de significación** y como señalamos en un epígrafe anterior, se le representa con la letra griega α (alfa). A la probabilidad de cometer el Error de Tipo II se le designa con la letra griega β (Beta).

Con el objetivo de optimizar el método, la formulación de las hipótesis nula y alternativa se realiza de forma tal que el Error de Tipo I sea el de consecuencias mas graves fijando la probabilidad de cometerlo con un valor lo suficientemente pequeño y que sea aceptable para el investigador.

Con relación a las hipótesis anteriormente formuladas, una regla de decisión razonable sería aceptar H_0 si la media de una muestra extraída de la población bajo estudio es menor o igual que un cierto valor crítico z comprendido entre 14 y 19, y aceptar H_1 si esta media es mayor que dicho valor. Supongamos entonces que una muestra de tamaño 16 extraída de la población bajo estudio dio como resultado una media muestral igual a 17, y que conocemos que la desviación típica poblacional tiene un valor igual a 9. Si fijamos el valor de α , entonces la probabilidad de rechazar H_0 siendo verdadera, viene dada por la expresión :

$$P\{ \bar{x} > z \mid \mu = 14 \} = \alpha$$

$$\text{Consideremos } \alpha = 0.05, \text{ entonces, } P\{ \bar{x} > z \mid \mu = 14 \} = 0.05$$

$$P\{ \bar{x} > z \mid \mu = 14 \} = 1 - P\{ \bar{x} \leq z \mid \mu = 14 \} = 0.05$$

$$\text{Tipificando : } 1 - P\left\{ \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq \frac{z - 14}{9/4} \right\} = 0.05$$

De donde :

$$P\left\{ \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq \frac{z - 14}{9/4} \right\} = 1 - 0.05 = 0.95$$

$$N\left(\frac{z-14}{\frac{9}{4}}\right) = 0.95 \quad \frac{z-14}{2.25} = 1.64 \text{ y despejando } z = 17.69$$

Entonces la regla de decisión quedaría como sigue:

Aceptar H_0 y rechazar H_1 si $\bar{x} \leq 17.69$

Rechazar H_0 y aceptar H_1 si $\bar{x} > 17.69$

Y como $\bar{x} = 17$ entonces la decisión final será aceptar H_0 .

Pasemos a continuación a calcular la probabilidad de cometer el error de tipo II o error β para diferentes valores de α :

Para $\alpha = 0.05$ ya vimos que $z = 17.69$

$$\beta = P\{\bar{x} \leq 17.69 \mid \mu = 19\} = N\left(\frac{17.69-19}{\frac{9}{4}}\right) = N(-0.58) = 1 - N(0.58) = 1 - 0.7190 = 0.2810$$

Para $\alpha = 0.01$

$$N\left(\frac{z-14}{\frac{9}{4}}\right) = 0.99 \quad \frac{z-14}{2.25} = 2.33 \text{ y despejando } z = 19.24$$

$$\beta = P\{\bar{x} \leq 19.24 \mid \mu = 19\} = N\left(\frac{19.24-19}{\frac{9}{4}}\right) = N(0.11) = 0.5438$$

Para $\alpha = 0.001$

$$N\left(\frac{z-14}{\frac{9}{4}}\right) = 0.999 \quad \frac{z-14}{2.25} = 3.08 \text{ y despejando } z = 20.93$$

$$\beta = P\{\bar{x} \leq 20.93 \mid \mu = 19\} = N\left(\frac{20.93 - 19}{\frac{9}{4}}\right) = N(0.86) = 0.8051$$

Resumiendo :

α	β
0.05	0.2810
0.01	0.5438
0.001	0.8051

El resumen anterior evidencia que cuando se disminuye la probabilidad de cometer el Error de Tipo I entonces aumenta la probabilidad de cometer el Error de Tipo II y viceversa.

7.1.1. Potencia de la prueba de hipótesis.

Para cualquier investigador, una “buena” prueba de hipótesis es, sin lugar a dudas, aquella en que la probabilidad de rechazar H_0 cuando ésta es falsa, es un valor cercano a 1, o dicho de otra manera, cuando :

$$P[\text{Rechazar } H_0, \text{ cuando } H_0 \text{ es falsa}] =$$

$$1 - P[\text{Aceptar } H_0, \text{ cuando } H_0 \text{ es falsa}] = 1 - \beta$$

es un valor cercano a 1. Al valor $1 - \beta$ se le conoce como la **potencia de la prueba de hipótesis**. Ampliando el resumen del epígrafe anterior :

n	α	β	$1 - \beta$
16	0.05	0.2810	0.7190
16	0.01	0.5438	0.4562
16	0.001	0.8051	0.1949

se puede apreciar que para un tamaño de muestra fijo ($n=16$), cuando se disminuye la probabilidad de rechazar H_0 cuando ésta es verdadera (α) entonces se disminuye también la potencia de la prueba ($1 - \beta$), lo cual es por supuesto algo no deseable. Veamos a continuación de que

forma podemos incrementar la potencia de la prueba de una hipótesis con independencia del valor de α que haya sido escogido.

Como se aprecia en el resumen anterior para un tamaño de muestra igual a 16 y un nivel de significación igual a 0.05, la potencia de la prueba es igual a 0.7190. Calculemos la potencia de la prueba si incrementamos el tamaño de la muestra, por ejemplo, a 25. Tal y como hicimos en el epígrafe anterior :

Para $\alpha = 0.05$ y $n = 25$

$$N\left(\frac{z-14}{\frac{9}{5}}\right) = 0.95 \quad \frac{z-14}{1.8} = 1.64 \text{ y despejando } z = 16.95$$

$$\beta = P\{\bar{x} \leq 16.95 \mid \mu = 19\} = N\left(\frac{16.95-19}{\frac{9}{5}}\right) = N(-1.14) = 1 - N(1.14) = 1 -$$

0.8729 = 0.1271 , y por tanto :

$$1-\beta = 1 - 0.1271 = 0.8729$$

valor que comparado con la potencia de la prueba (0.7190) para n igual a 16 y α igual a 0.05, nos indica que **al aumentar el tamaño de la muestra podemos incrementar el valor de la potencia de la prueba de hipótesis**. El lector podrá comprobar sin mucha dificultad los resultados que se reportan en el siguiente resumen y compararlos con los que aparecen al inicio de este epígrafe :

n	α	β	$1 - \beta$
25	0.05	0.1271	0.8729
25	0.01	0.3264	0.6736
25	0.001	0.3821	0.6179

Prueba de hipótesis sobre la media de una población con varianza poblacional conocida.

En toda prueba de hipótesis se hace necesario asignarle un valor hipotético a la media de la población sobre la cual se desarrolla la investigación. Una prueba de hipótesis para la media de una población se hace necesaria cuando ocurre un evento que de alguna manera nos hace suponer un cambio en la media poblacional. El valor de la media poblacional antes de que tal evento ocurra, es el valor hipotético al que nos referíamos con anterioridad, y se le suele designar como μ_0 . Para una mejor comprensión del aspecto que estamos tratando, analicemos la situación que se presenta a continuación :

La producción promedio de leche diaria en una vaquería en la que se utiliza el ordeño manual, es de 10.1 litros. Se cambia el sistema de ordeño a mecanizado, y en una muestra aleatoria de 16 días después del cambio, se obtuvo una producción media diaria de 10.8 litros. Adicionalmente "**se conoce**" que la varianza poblacional de la producción de leche es de 1.21 litros.

Observe que antes de que se produjera el evento relacionado con el cambio en el sistema de ordeño de la vaquería, existían evidencias que la producción diaria de leche era de 10.1 litros. Este valor representa la producción de leche que de manera tradicional se producía antes de que el sistema de ordeño fuera variado, es decir, antes que el evento hubiese ocurrido. Por tanto, y sin lugar a dudas, podemos concluir que el valor hipotético de la media de la población es de 10.1 litros, y en consecuencia :

$$\mu_0 = 10.1$$

La formulación de la hipótesis nula sería entonces :

$$H_0 : \mu = 10.1$$

Tres hipótesis alternativas podrían entonces ser formuladas en correspondencia con el interés de la investigación. Estas diferentes hipótesis alternativas son :

- ◆ $H_1 : \mu \neq 10.1$ si nos interesa conocer si el cambio en el sistema de ordeño hizo que la producción de leche cambiara.
- ◆ $H_1 : \mu > 10.1$ si esperamos que el nuevo sistema de ordeño haya incrementado la producción de leche.

- ◆ $H_1 : \mu < 10.1$ si por el contrario suponemos que el cambio hizo que disminuyera la producción.

Un resumen de los datos de la prueba de hipótesis se muestra a continuación :

$$\mu_0 = 10.1 \quad n = 16 \quad \bar{x} = 10.8$$

$$\sigma^2 = 1.21 \Rightarrow \sigma = 1.1$$

Desarrollemos a continuación de forma teórica las tres posibles pruebas de hipótesis que pueden ser formuladas, para con posterioridad referirnos específicamente al caso que nos ocupa :

Consideremos en primer lugar las hipótesis:

- ◆ $H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0$

donde μ es la media de la distribución de una población normal con varianza σ^2 y μ_0 es un número real.

Si \bar{x} es la media de una muestra aleatoria de esta población, resulta razonable rechazar H_0 si \bar{x} difiere mucho de μ , o sea, si

$|\bar{x} - \mu| > x$ para x suficientemente grande. Tipificando :

$$\left| \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \right| > \frac{x}{\frac{\sigma}{\sqrt{n}}} = z$$

El problema de la decisión de rechazar H_0 se reduce entonces a determinar el valor de z . Utilicemos el nivel de significación α partiendo de la relación :

$$P\{\text{rechazar } H_0 \mid H_0 \text{ es cierta}\} = \alpha$$

$$P\left[\left| \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| > z \right] = \alpha \quad \text{ya que si } H_0 \text{ es cierta entonces } \mu = \mu_0$$

$$P \left[\left| \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| \leq z \right] = 1 - \alpha \quad P \left[-z \leq \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq z \right] = 1 - \alpha$$

$$N(z) - N(-z) = 1 - \alpha \quad N(z) - 1 + N(z) = 1 - \alpha$$

$$2N(z) = 2 - \alpha \quad N(z) = 1 - \frac{\alpha}{2}$$

y entonces $z = Z_{1 - \frac{\alpha}{2}}$ que es el percentil de orden $1 - \frac{\alpha}{2}$ de la

distribución $N(0,1)$.

La regla de decisión quedaría entonces como sigue :

$$\text{Rechazar } H_0 \text{ si } \left| \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| > Z_{1 - \frac{\alpha}{2}}$$

$$\text{Aceptar } H_0 \text{ si } \left| \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| \leq Z_{1 - \frac{\alpha}{2}}$$

Apliquemos la regla de decisión en el caso que estamos estudiando utilizando un nivel de significación del 5 %.

$$\blacklozenge H_0: \mu = 10.1 \quad H_1: \mu \neq 10.1$$

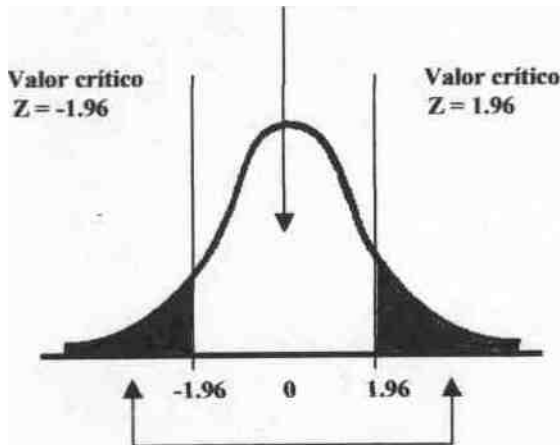
$$\left| \frac{10.8 - 10.1}{\frac{1.1}{\sqrt{16}}} \right| = \left| \frac{0.7}{0.28} \right| = 2.5$$

$$1 - \frac{\alpha}{2} = 1 - \frac{0.05}{2} = 1 - 0.025 = 0.975$$

$$Z_{0.975} = 1.96$$

y como $2.5 > 1.96$, rechazamos H_0 con un nivel de significación del 5%, es decir, el sistema de ordeño mecanizado provocó un cambio en la producción de leche, y la probabilidad de que este resultado no sea cierto es del 5%. Las zonas de rechazo y de aceptación de esta prueba de hipótesis

Zona de aceptación



Zona de rechazo

Cuando se observa el gráfico anterior se puede apreciar que la zona de rechazo está conformada por dos “**colas**” que se extienden a la derecha e izquierda de la curva de la distribución normal. Por esta razón se dice que la prueba de hipótesis que hemos desarrollado es una “**prueba de dos colas**”.

Consideremos las hipótesis:

$$\diamond H_0: \mu = \mu_0 \quad H_1: \mu > \mu_0$$

Si \bar{x} es la media de una muestra aleatoria de esta población, resulta razonable rechazar H_0 si \bar{x} es mucho mayor que μ , o sea, si $\bar{x} - \mu > x$ para x suficientemente grande. Tipificando:

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{x}{\frac{\sigma}{\sqrt{n}}} = z$$

Calculemos el valor de z para un nivel de significación α .

$$P\{\text{rechazar } H_0 \mid H_0 \text{ es cierta}\} = \alpha$$

$$P \left[\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z = \alpha \right] \text{ ya que si } H_0 \text{ es cierta entonces } \mu = \mu_0$$

$$P \left[\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq z = 1 - \alpha \right]$$

$$N(z) = 1 - \alpha$$

y entonces $z = Z_{1-\alpha}$ que es el percentil de orden $1 - \alpha$ de la distribución $N(0,1)$.

La regla de decisión quedaría entonces como sigue :

$$\text{Rechazar } H_0 \text{ si } \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > Z_{1-\alpha}$$

$$\text{Aceptar } H_0 \text{ si } \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq Z_{1-\alpha}$$

En nuestro caso :

$$\blacklozenge H_0: \mu = 10.1 \quad H_1: \mu > 10.1$$

$$\frac{10.8 - 10.1}{\frac{1.1}{\sqrt{16}}} = \frac{0.7}{0.28} = 2.5$$

$$1 - \alpha = 1 - 0.05 = 0.95$$

$$Z_{0.95} = 1.64$$

y como $2.5 > 1.64$, rechazamos H_0 con un nivel de significación del 5%, es decir, el sistema de ordeño mecanizado provocó un aumento en la producción de leche, y la probabilidad de que este resultado no sea cierto es del 5%.

PRUEBAS NO PARAMÉTRICAS

PRUEBAS NO PARAMÉTRICAS

8.1 INTERVALOS DE CONFIANZA

Cuando tenemos un conjunto de datos, necesitamos determinar entre que valores se va a encontrar los intervalos respecto al μ , a un determinado nivel de confianza respecto a la media se requiere del error estándar de media, teniendo como producto un valor límite inferior y límite superior, es decir se infiere respecto a la probabilidad $1-\alpha$ (y en los extremos $\alpha/2$ (colas) respecto a μ se emplea la siguiente fórmula:

$$P (-t_{.05} \times s < \mu < +t_{.05} \times s) = 0.95$$

Ejercicio:

Supongamos que se ha probado siete mermeladas de frutas tropicales para ello se tomó los grados Brix obteniéndose los siguientes resultados:

Mermelada	Brix
1	45
2	60
3	54
4	48
5	70
6	65
7	50
Promedio	56
D. Estan	9.29
CV(%)	16.59
Error. Estandar	3.51

$$= \frac{392}{7} = 56s^2 = \sum \frac{xi^2 - (xi)^2/n}{n-1}$$

$$= \frac{22470 - 153664/7}{7-1} = \frac{22470 - 21952}{6} = \frac{518}{6} = 86.33$$

$$s = \sqrt{s^2} = \sqrt{86.33} = 9.29$$

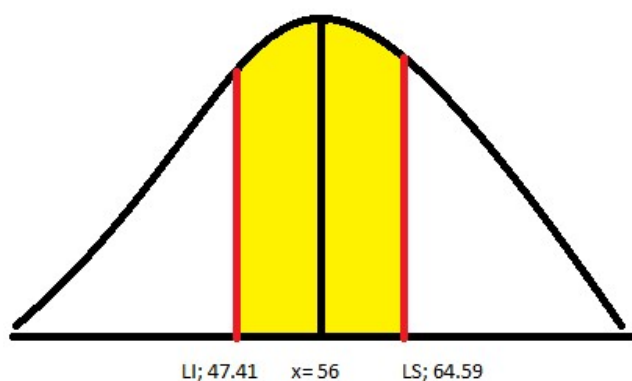
$$s = \frac{9.29}{\sqrt{7}} = \frac{9.29}{2.65} = 3.51$$

$$CV(\%) = \frac{9.29}{56} * 100 = 16.59$$

Intervalo de confianza al 95%,

$$P (-t_{.05} \times s < \mu < +t_{.05} \times s) = 56 - (2.447) (3.51) < \mu < 56 + (2.447) (3.51)$$

$$47.41 < \mu < 64.59$$



8.2 PRUEBA DE HIPOTESIS

Prueba deriva del latín, concretamente de "probus", que puede traducirse como "bueno".

-Hipótesis, por su parte, emana del griego, ya que se encuentra conformada por la suma de dos partes diferenciadas: el prefijo "hypo-", que es sinónimo de "debajo", y el sustantivo "thesis", que es equivalente a "conclusión".

Las pruebas de hipótesis son de gran utilidad en el campo estadístico ya que tiene por objeto comprobar las teorías tentativas, es decir se comprueba dos hipótesis estadísticas ($H_0 =$ nula; $\mu_1 = \mu_2$ y $H_a =$ alternativa; $\mu_1 \neq \mu_2$) que se requiere dará respuesta a una incertidumbre, basado en un parámetro conocido a una probabilidad estimada, a partir de una muestra, la misma que se fundamente de la varianza de una población normalmente distribuida o en distribución normal.

PRUEBA DE HIPÓTESIS PARA UN GRUPO ALEATORIO DE PROMEDIO DE UNA POBLACIÓN NORMAL

El procedimiento depende del tamaño de la muestra por lo general trabajos diarios en laboratorio o campo, es decir cuando las observaciones son de 30 como mínimo, coincidiendo con la distribución de "z" y de "t"

Pasos a seguir

Se conoce que ($\bar{x} - \mu$) se inicia con la obtención del error estándar de media empleando la siguiente fórmula:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Según el nivel de probabilidad usado será el grado determinará la seguridad en la inferencia esto requiere los siguientes pasos

- 1) Se determina mediante el uso de las tablas de t o se z igual o mayor al estimado
- 2) Se compara con el valor tabular con el calculado; si es mayor al tabular se rechaza la hipótesis nula.

Problema:

Con el ejemplo de probar la hipótesis (1.0) en ancho de almendras de cacao de 8 muestras escogidas al azar se registraron los siguientes datos:

MUESTRA	1	2	3	4	5	6	7	8	$\sum x$	
Ancho (mm)	0.6	0.7	0.7	0.8	1.3	1.0	0.6	0.7	6.4	0.8
$(X - \bar{x})^2$	0.04	0.01	0.01	0.00	0.25	0.04	0.04	0.01	$\sum x^2 = 0.40$	

$$s^2 \frac{\sum x^2}{n-1} = \frac{0.40}{7} = 0.057$$

$$s = \sqrt{0.057} = 0.239$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{0.239}{2.83} = 0.084$$

Teniendo las estimaciones necesarias, se procede a calcular el valor de t

$$\frac{0.8 - 1.0}{0.084} = \frac{-0.200}{0.084} = t \text{ calculado} - 2.38 \quad t \text{ tabla } (0.05 \text{ gl } 7) = 2.36$$

Dado que la distribución de t y z es simétrica y el valor de la del t 2.38 calculado es mayor que el t tabular al 5% de probabilidad escogido se rechaza la hipótesis nula y se acepta la alternativa, todo cabe indicar que el ancho de almendras es mayor al promedio.

Ejercicio en clase:

Con objeto de evaluar si la producción anual de resina por árbol de pinus es igual o menor a 2kg, se registraron los datos de 10 árboles al azar obteniéndose los siguientes datos:

ÁRBOL	1	2	3	4	5	6	7	8	9	10	TOTAL
Cantidad de resina	2.8	3.5	4.2	1.6	3.3	4.0	2.8	3.1	3.5	1.8	30.60
x²	7.84	12.25	17.64	2.56	10.89	16.00	7.84	9.61	12.25	3.24	100.12

Promedio= 3.06

$$s^2 = \frac{\sum x^2}{n-1} = \frac{100.12 - (30.60)^2/10}{9} = 0.7211$$

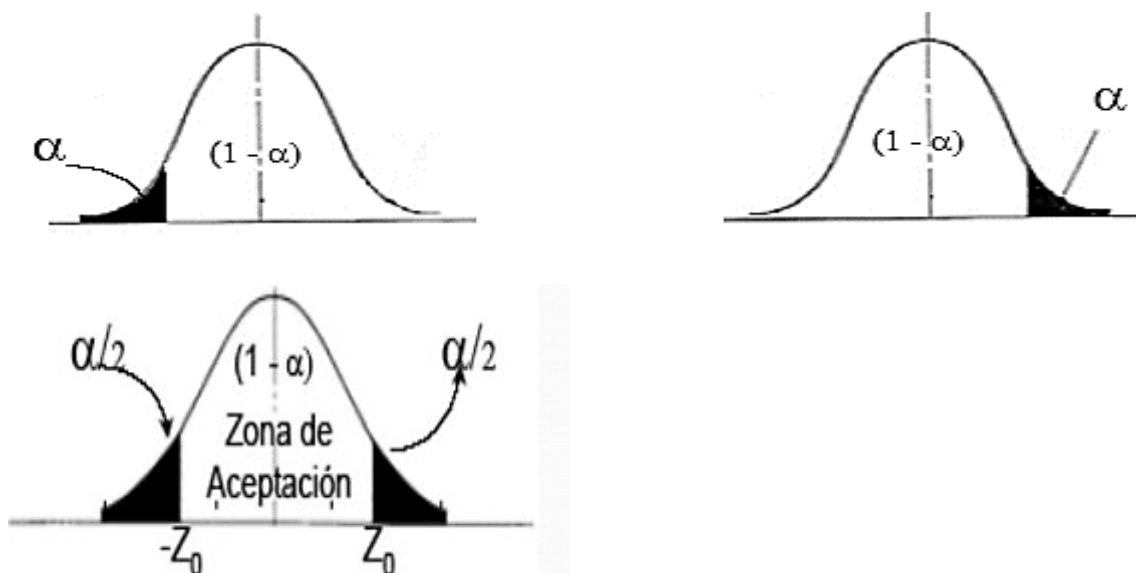
$$s = \sqrt{0.7211} = 0.8492$$

$$t = \frac{3.06 - 2.00}{0.2685} = 3.9478 **$$

(t .05 tabla 2.26 y t.01 3.25)

En este caso la hipótesis nula señala que la producción es de 2 kg, la hipótesis indica que la producción de resina supera al valor establecido, por lo tanto estadísticamente se dice que es altamente significativo por lo cual las hipótesis pueden ser de una cola o unilateral:

$$H_0 = \mu \leq 2 \quad \text{vs} \quad H_a = \mu \geq 2$$



En el problema establecido del rendimiento de 2Kg lo valores caen dentro de la zona de rechazo por lo cual se debe rechazar la hipótesis nula propuesta, se acepta la alternativa.

8.3. PRUEBA DE CHI CUADRADO

En ocasiones un valor estimado puede generar dudas generado por varianzas, es decir se requiere definir la variación existente (suponiendo rendimiento) ante lo cual se podría emplear el test de Chi cuadrado o Jí-Cuadrada, aprovechando la relación entre varianzas, es decir la suma se sus valores elevado al cuadrado de las variables de carácter independiente, dentro de una distribución normal para ello se emplea la siguiente fórmula:

$$x^2 = \sum_z 2 = \sum \left(\frac{Xi - u}{\delta}\right)^2$$

El estadístico de prueba es

$$\chi_{prueba}^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \frac{(o_3 - e_3)^2}{e_3} + \dots + \frac{(o_n - e_n)^2}{e_n}$$

$$\chi_{prueba}^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

Donde:

χ es la letra griega ji

χ^2 se lee ji cuadrado

Ejemplo:

Asumamos que tenemos los siguientes valores al lanzar dos dado 6 veces:

Sucesos	1	2	3	4	5	6
valores	6	8	9	15	14	8

R= 2 ; K= 5

Hipotesis:

$H_0 = \mu_1 = \mu_2$

$H_a = \mu_1 \neq \mu_2$

Grados de libertad = (2-1) (6-1) = 5

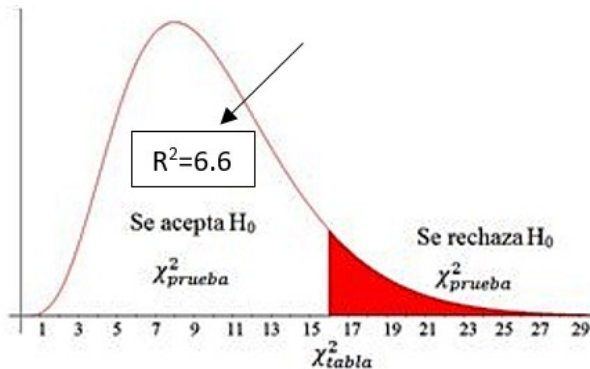
Solución:

$$x^2 = \frac{(6 - 10)^2}{10} + \frac{(8 - 10)^2}{10} + \frac{(9 - 10)^2}{10} + \frac{(15 - 10)^2}{10} + \frac{(14 - 10)^2}{10} + \frac{(8 - 10)^2}{10}$$

$$x^2 = 1.6 + 0.4 + 0.1 + 2.5 + 1.6 + 0.4$$

$$x^2 = 6.6$$

Se acepta la hipótesis H_0 ; ya que el valor de chi cuadrado es de 6.6 y al compararlo con el valor tabular al 0.05 es de 11.071 con 5 grados de libertad, por lo cual los valores caen dentro de la zona aceptación.



8.4. PRUEBA DE "t" STUDENT

Las técnicas empleadas en la comparación de dos grupos aleatorios son muy útiles para el investigador, dado que de dos valores poblacionales se la realiza para ver el comportamiento en la distribución de t cuando la muestra es de 30 observaciones, en las cuales se puede inferir, a la vez verificar las hipótesis planteadas, por lo cual existen cinco casos los cuales se describen:

- 1.- Cuando las varianzas son iguales: $\sigma_1^2 = \sigma_2^2$; $n_1=n_2$
- 2.- Cuando las varianzas son iguales: $\sigma_1^2 = \sigma_2^2$; $n_1 \neq n_2$
- 3.- Cuando las varianzas son desiguales: $\sigma_1^2 \neq \sigma_2^2$; $n_1=n_2$
- 4.- Cuando las varianzas son desiguales: $\sigma_1^2 \neq \sigma_2^2$; $n_1 \neq n_2$
- 5.- Observaciones Apareadas.

En los anteriores casos se pondrá a prueba las hipótesis estadísticas las cuales indican que:

Hipotesis:

$$H_0 = \mu_1 = \mu_2$$

$$H_a = \mu_1 \neq \mu_2$$

Además de hipótesis alterantes: $\mu_1 \leq \mu_2$ y $\mu_1 \geq \mu_2$.

8.4.1. PRIMER CASO: VARIANZAS IGUALES DE MUESTRAS IDENTICAS.

Para la aplicación de este caso, es preciso que ambas muestras estén con el mismo número de observaciones.

El procedimiento consiste en calcular la diferencia entre medias ($\mu_1 - \mu_2$), expresado en la siguiente fórmula:

$$t = (\mu_1 - \mu_2) / s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad \text{o} \quad t = \frac{\mu_1 - \mu_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Ejemplo:

Se desea determinar el crecimiento diamétrico de un lote de nueve variedades T cacao diferente de otro T. bicolor con el mismo número de árboles en el mismo sitio se registraron los siguientes datos.

Variedad	Valores del DAP (cm)									$\sum x_i$	\bar{x}
T. bicolor	7.0	5.0	6.5	9.5	8.0	8.0	10.0	10.0	6.0	70.0	7.78
T. cacao	5.0	4.5	4.2	4.0	7.0	6.8	5.5	4.2	3.0	44.2	4.91

$$s^2 = \frac{\sum x_i^2 - (\sum x_i)^2 / n}{n-1}$$

$$s_1^2 = \frac{570.50 - 70.00^2 / 9}{8} = \frac{26.055}{8} = 3.257$$

$$s_2^2 = \frac{231.02 - 44.2^2 / 9}{8} = \frac{13.949}{8} = 1.744$$

$$s^2 = \frac{3.257 + 1.744}{2} = 2.500$$

$$s = \frac{2.500}{9} = 0.278 \quad \text{y} \quad 2(0.278) = 0.556$$

$$s = \sqrt{0.556} = 0.746$$

$$\mu_1 - \mu_2 = 7.78 - 4.91 = 2.870$$

$$t = \frac{2.870}{0.746} = 3.85 \quad ** \quad t_{.01} (gl 16) = 2.921$$

El valor calculado de t es de 3.85 al comparar con el valor tabular al 0.01% es igual a 2.921 lo que indica los diámetros en las dos variedades son diferentes es decir tienen efecto demostrando que no hay igualdad es decir se acepta la hipótesis alternativa los valores caen dentro de la zona de rechazo.

8.4.2. SEGUNDO CASO: HOMOGENEIDAD DE LAS VARIANZAS DIFERENTES MUESTRAS

Al igual que el primer caso no hay variación en el cálculo, sin embargo, la variante será en el cálculo del error estándar de la media. Debido a que las observaciones no son iguales por lo cual se usa la siguiente fórmula:

Ejemplo:

$$s_{1-s_2} = \sqrt{\frac{\sum x^2}{n(n-1)}} = \sqrt{\frac{26.06+13.95}{9(8)}} = 0.746$$

$$s_{1-s_2} = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} = s^2 = \left(\frac{1}{n_1} + \frac{1}{n_2}\right) = s \sqrt{\left(\frac{n_1+n_2}{n_1 n_2}\right)}$$

Ejemplo

El número de cocones de Neodiprion, obteniéndose en 12 cuadros de muestreos uno, y 9 cuadros en un segundo muestreo por lo cual se registraron los siguientes datos:

Exposición	Valores												n	
Muestra 1	6	9	12	15	7	10	13	12	10	8	16	4	12	10.167
Muestra 2	15	10	9	12	11	16	16	11	7	-	-	-	9	11.889

$$\text{Grados de libertad} = (n_1-1) + (n_2-1) = n_1+n_2 - 2$$

$$s^2_{gl\ total} = \frac{\sum x_1^2 + \sum x_2^2}{19} = \frac{143.667+80.889}{19} = 11.819$$

Aplicando la fórmula:

$$s_{1-s_2} = \sqrt{11.819 \left(\frac{21}{108}\right)} = 1.516$$

$$1-2 = 11.889-10.167 = 1.722$$

$$t = \frac{1.722}{1.516} = 1.13 \text{ ns} ; t_{.05 \text{ gl } 16} = 2.093$$

El valor calculado de t es de 1.13 al comparar con el valor tabular al 0.05% es igual a 2.093 lo que indica que las muestras son iguales decir no tienen efecto demostrando que hay igualdad en la distribución es decir se acepta la hipótesis nula ya que los valores caen dentro de la zona aceptación.

8.4.3. TERCER Y CUARTO CASO: VARIANZAS DE DOS POBLACIONES CON DIFERENTES MUESTRAS

En el primer y segundo caso implican heterogeneidad de las varianzas

Ejemplo

Se requiere estimar si las medias de dos muestras a continuación son diferentes:

Muestra	Valores											$\sum X_i$	X_i^2		S^2
A	12	18	14	16	17	12	12	14	13	12	11	151	54.18	13.73	5.42
B	6	10	24	30	8	6	20	12	4	26	4	150	918.55	13.64	91.85

Lo primero que hay que hacer es una prueba de homogeneidad de las varianzas para determinar si el problema aplica al primer o segundo caso.

Haciendo la prueba de F:

$F = \frac{91.85}{5.42} = 16.95 **$ $F_{.01} \text{ gl } 10-10 = 4.85$ por lo cual indica que las varianzas no son homogéneas, por lo que se emplea la siguiente fórmula:

$$s^2 = \frac{SA^2 + SB^2}{2} = 48.62$$

$$s_{1-s} = \sqrt{2 \left(\frac{s^2}{n} \right)} = \sqrt{\frac{2 * 48.62}{11}} = 2.97$$

$$t = \frac{0.09}{2.97} = 0.03 \text{ ns} ; t_{.05 \text{ gl } 10} = 2.228$$

No existe diferencias entre los grupos en estudio

8.4.4. Cuarto caso:

Ejemplo:

Suponga ahora que, se registró valores numéricos de dos muestras A y B con dos datos faltantes:

Muestra	Valores										
A	12	18	14	16	17	12	12	14	13	12	11
B	24	30	8	6	20	12	4	26	4	-	-

$$\sum Xa = 151$$

$$\sum Xb = 134$$

$$a = 13.73$$

$$b = 14.89$$

$$\sum Xa^2 = 54.18$$

$$\sum Xb^2 = 832.89$$

$$SA^2 = 5.42$$

$$SB^2 = 104.11$$

Entonces aplicamos la fórmula:

$$s^2 = \frac{Sc \text{ Total}}{gl} = \frac{887.07}{18} = 49.28$$

$$s_{x1 - x2} = \sqrt{s^2 \left(\frac{n1+n2}{n1n2} \right)} = \sqrt{49.28 \left(\frac{20}{99} \right)} = \sqrt{9.95} = 3.16$$

$$t = \frac{1.16}{3.16} = 0.367 \text{ ns} ; t_{.05} = 2.30$$

No existe diferencias entre los grupos en estudio

8.5. KRUSKAL WALLIS PARA K MUESTRAS

Esta es una prueba para comparar k muestras en experimentos de un factor reemplazando al ANOVA-1 (Diseño Completamente Aleatorio) cuando este no es posible aplicar, basado en el uso de rangos para K=2 es equivalente a una prueba de U Mann-Whitney en las pruebas de rangos.

$$H = \frac{12}{N(N+1)} \left(\frac{1}{N} \sum R_i^2 - \left(\sum R_i \right)^2 / NT \right)$$

Donde:

n_i : Tamaño de la i -ésima muestra, $i = 1, 2, 3, \dots, k$ (Tratamientos)

N : $\sum n_i$ (Unidad Experimental)

R_i : Suma de rangos para i -ésima muestra. Para obtener estas sumas todos los datos de todas las muestras son ordenadas y asignadas un rango de menor a mayor.

H- test: Como se suele llamar a esta prueba, se aproxima a la distribución de X^2 con $k-1$ grados de libertad cuando $n_i \geq 5$.

$$Divisor = 1 - \frac{\sum(t^3 - t)}{N^3 - n}$$

Ejercicio:

La Altura de plantas de cuatro zonas diferentes, en cada zona se tomó 5 muestras al azar (cuadrantes) de plantas por flor, se registraron los siguientes datos:

Zonas	REPETICIONES				
	I	II	III	IV	V
A	120	112	115	119	117
B	117	114	112	115	115
C	123	116	118	114	118
D	115	117	120	112	116

Hipótesis propuestas:

$H_0 = A=B=C=D$

$H_a = A \neq B \neq C \neq D$

ASIGNACIÓN DE RANGOS:

IMC-67 $(1+2+3/3=2.0)$

$$(1+2+3)/3=2.0 \quad (9/2=4.5)$$

Nº	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Índice	112	112	112	114	114	115	115	115	115	115	116	117	117	117	118	118	119	120	120	123
Rango	2.0	2.0	2.0	4.5	4.5	7.5	7.5	7.5	7.5	10.5	10.5	13.0	13.0	13.0	15.5	15.5	17.0	18.5	18.5	20.0

Zonas	REPETICIONES					$\sum Ri$	$\sum Ri^2$
	I	II	III	IV	V		
A	18.5	2.0	7.5	17.0	13.0	58.0	3364.0
B	13.0	4.5	2.0	7.5	7.5	34.5	1190.25
C	20.0	10.5	15.5	4.5	15.5	66.0	4356.00
D	7.5	13.0	18.5	2.0	10.5	51.5	2652.25
\sum						210.0	11562.50

N= número de unidades experimentales

$$NT= 4 \times 5 = 20$$

$$H = \frac{12}{20(21)} \left(\frac{11562.50}{5} - \frac{210.0^2}{20} \right) = 3.07$$

Para la corrección de H calculamos el divisor, calcular de coincidencias (ti).

Empates	112	114	115	116	117	118	120	$\sum(t^3 - t)$
ti	3	2	4	2	3	2	2	132
$t^3 - t$	24	6	60	6	24	6	6	

$$C = Divisor = 1 - \frac{132}{20^3 - 20} = 0.9835 \quad \frac{3.07}{0.9835} = 3.12$$

$$VALOR \text{ tabular de } X^2_{.05} = 7.81$$

Entonces. $3.12 < 7.81$ ns (no significativo) se acepta la hipótesis $H_0 = A=B=C=D$

Se concluye que el valor de H corregido respecto al valor teórico de X^2 , no existe cambios en la altura de plantas en diferentes zonas donde se hizo el muestreo.

REGRESIÓN Y CORRELACIÓN

9.0 REGRESIÓN Y CORRELACIÓN

9.1 Introducción

Es posible establecer una relación funcional entre la velocidad de un móvil y el tiempo que demora en recorrer una determinada distancia con relación al espacio recorrido, es decir, $e = v t$.

De esta manera, si conocemos la velocidad con que un móvil ha recorrido una determinada distancia y el tiempo que ha empleado en ello, basta con sustituir estos valores en la expresión anterior para obtener el espacio recorrido. En esta relación, a las variables v y t se les llama variables independientes mientras que e se conoce como variable dependiente.

Pero la Física es una ciencia exacta, y en muchos otros casos, no es posible encontrar una relación funcional, fórmula o modelo matemático que exprese de forma exacta la relación que existe entre diferentes variables independientes (o series de datos) y una variable dependiente (o serie de datos).

Resulta evidente, por ejemplo, que existe una relación entre la temperatura (variable independiente) en una ciudad y el número de colas vendidas (variable dependiente) en un supermercado de este lugar. Pero el número de colas vendidas no solo depende de la temperatura, sino mas bien, de un grupo de factores que intervienen en ello, y por tanto, resulta difícil, sino imposible, obtener una relación funcional o fórmula entre ambas variables que permita calcular de forma exacta el número de colas vendidas conocido el valor de la temperatura por lo cual abordaremos la teoría de la regresión y la correlación.

La regresión de una variable dependiente (Y) sobre una o más variables independientes (X_1, X_2, \dots) expresa la variación que sufre la primera como consecuencia de la variación de las segundas. La teoría de la regresión permite establecer una expresión matemática (no funcional) entre estas variables de forma que evaluando la expresión en las variables independientes se pueda hacer una estimación mas o menos acertada de la variable dependiente.

Por su parte la teoría de la correlación tiene por objeto determinar el grado o intensidad de la interdependencia entre dos o más variables, permitiendo definir con claridad si esta interdependencia es o no cierta.

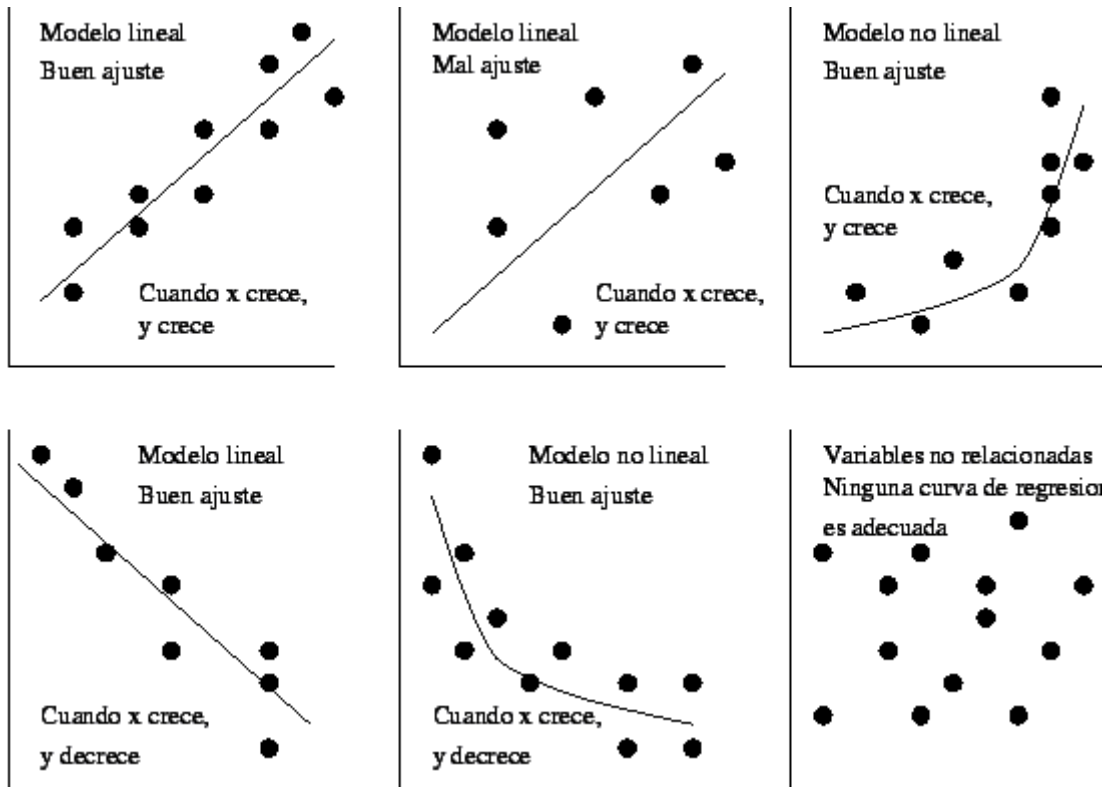
9.2 REGRESIÓN LINEAL SIMPLE

Para iniciar este estudio, consideremos dos variables o series de datos entre las cuales se considera pueda existir una relación de tipo lineal, es decir, que la representación algebraica de esta relación es una línea recta según el caso.

Al valor **b** se le conoce como **coeficiente de regresión**, y su expresión numérica tiene la siguiente interpretación:

- a) Cuando $b > 0$, a incrementos de la variable independiente X corresponden incrementos de la variable dependiente Y , es decir, la línea recta es creciente. En este caso, es posible obtener una estimación del valor de Y sustituyendo en la ecuación un valor determinado de X .
- b) Cuando $b < 0$, la línea recta es decreciente, es decir, a incrementos de la variable independiente X corresponden decrecimientos de la variable dependiente Y . También en este caso es posible predecir el valor de Y para un determinado valor de X .
- c) Cuando $b = 0$, la ecuación de la línea recta asume la forma $Y = a$ y por tanto es paralela al eje de las X . En esta situación resulta imposible predecir un valor de Y .

TIPOS REGRESIONES Y SUS RESPECTIVOS MODELOS:



9.3. COEFICIENTE DE CORRELACIÓN LINEAL SIMPLE

En el epígrafe anterior, concretamos en forma matemática la posible relación existente entre dos variables o series de valores entre las cuales la dependencia podía ser expresada mediante una línea recta.

A menudo resulta muy favorable poder expresar en términos numéricos el grado de dependencia que eventualmente pueda existir entre dos variables o entre las series de datos que la representan. Este valor matemáticamente calculado recibe el nombre de **coeficiente de correlación**.

Para obtener la expresión matemática de este coeficiente partamos de la igualdad que se muestra a continuación.

$$\begin{aligned} \sum X_i Y_i &= \sum [\bar{X} + (X_i - \bar{X})][\bar{Y} + (Y_i - \bar{Y})] = \\ &= \sum \bar{X} \bar{Y} + \sum \bar{X} (Y_i - \bar{Y}) + \sum \bar{Y} (X_i - \bar{X}) + \sum (X_i - \bar{X})(Y_i - \bar{Y}) = \\ &= n \bar{X} \bar{Y} + \bar{X} \sum (Y_i - \bar{Y}) + \bar{Y} \sum (X_i - \bar{X}) + \sum (X_i - \bar{X})(Y_i - \bar{Y}) \end{aligned}$$

pero,

$$\sum (X_i - \bar{X}) = 0 \text{ y } \sum (Y_i - \bar{Y}) = 0, \text{ por tanto dividiendo para } n :$$

$$\frac{\sum X_i Y_i}{n} = \bar{X} \bar{Y} + \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

De la teoría de probabilidades sabemos que si X e Y son independientes, entonces se cumple que:

$$\frac{\sum X_i Y_i}{n} = \frac{\sum X_i}{n} + \frac{\sum Y_i}{n} = \bar{X} + \bar{Y} \text{ y por tanto, en ese caso:}$$

$$\frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n} = 0 \quad (1)$$

Si por el contrario, X e Y no son independientes, entonces:

$$\frac{\sum X_i Y_i}{n} \neq \frac{\sum X_i}{n} + \frac{\sum Y_i}{n} \quad \text{y :}$$

$$\frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n} \neq 0$$

Podemos concluir entonces, que la expresión (1) puede ser utilizada para calcular de forma numérica la intensidad o el grado de dependencia entre dos variables o series de valores, en concordancia con el hecho de que dicha expresión se acerque o esté alejada significativamente de 0.

Si en (1), expresamos los valores de las desviaciones con relación a su media de cada una de las variables en unidades de desviación típica, tenemos

$$\frac{\sum \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\sigma_X \sigma_Y}}{n} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n \sigma_X \sigma_Y} \quad (2)$$

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}} \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{n}}} \text{ de donde,}$$

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} \quad (3)$$

Expresión que recibe el nombre de **coeficiente de correlación** y se denota con la letra **R²**.

Si escribimos la expresión (2) en términos de desviaciones, obtenemos:

la cual expresada en términos de sumas de cuadrados queda:

$$r = \frac{SPC_{XY}}{\sqrt{SCC_X \cdot SCC_Y}}$$

donde:

$$SPC_{XY} = \sum_i X_i Y_i - \frac{\sum_i X_i \sum_i Y_i}{n}$$

$$SCC_X = \sum_i X_i^2 - \frac{\left(\sum_i X_i\right)^2}{n}$$

$$SCC_Y = \sum_i Y_i^2 - \frac{\left(\sum_i Y_i\right)^2}{n}$$

n = número de pares de observaciones

Puede demostrarse matemáticamente que el coeficiente de correlación siempre es una cantidad que oscila entre 1 y -1, es decir,

$$-1 \leq r \leq 1.$$

Cuando $r > 0$, la correlación es directa, y a valores crecientes de una de las variables corresponden valores que tienden a ser crecientes de la otra.

Cuando $r < 0$, la correlación es inversa, y a valores crecientes de una de las variables corresponden valores que tienden a ser decrecientes de la otra.

Cuando $r = 1$ o $r = -1$, la correlación es perfecta. Cuanto más se acerque a la unidad el valor del coeficiente de correlación más alta será ésta. En cambio, valores próximos a cero indican, en general, ausencia de correlación entre los valores simultáneos de ambas variables.

Calculemos el coeficiente de correlación entre las dos series de datos que hemos estado trabajando.

$$r = \frac{SPC_{XY}}{\sqrt{SCC_X \cdot SCC_Y}} = \frac{169 - \frac{(20)(38)}{5}}{\sqrt{\left(90 - \frac{(20)^2}{5}\right)\left(322 - \frac{(38)^2}{5}\right)}} =$$

$$17/18.22 = 0.933$$

El valor de r obtenido es positivo y cercano a 1, lo cual indica que existe una alta correlación directa entre ambas variables.

A menudo se suele reportar el valor del cuadrado del coeficiente de correlación en lugar de éste. Este valor recibe el nombre de **coeficiente de determinación** y se denota como **R²**. En nuestro caso, **R² = 0.87**.

Si como hemos dicho, el valor de r oscila entre -1 y 1, entonces el valor de **R²** es siempre menor o igual a 1 y mayor o igual a 0.

Un método alternativo para calcular el coeficiente de determinación, y en consecuencia, el coeficiente de correlación, es mediante las sumas de cuadrados del análisis de regresión. El coeficiente de determinación se define en estos términos como **el porcentaje de la variación de la variable dependiente Y explicado por la variable independiente X**, o lo

$$R^2 = \frac{SCC_{REG}}{SCC_T}$$

que es lo mismo, mediante la siguiente expresión:

De la tabla del análisis de regresión tenemos:

$$R^2 = \frac{28.9}{33.2} = 0.87$$

Valor que coincide exactamente con el calculado anteriormente.

Podemos decir entonces que el 87% de la variación de la variable dependiente Y está ocasionada por su relación lineal con la variable independiente X.

EL término regresión lineal implica, cuantificar de forma gráfica y cuantificable mediante una gráfica o diagrama de dispersión, cuyos puntos dispersos corresponden a dos variables independiente (X) y dependiente (Y), matemáticamente "Y es en función de X" en el ejemplo desarrollado la presión está en función de la edad.

Ejemplo

En un muestreo de personas de diferentes edades se tomaron datos de la edad (x) y años promedio (Y) de la presión sanguínea (en mm de mercurio). Los valores obtenidos son:

Pares	Edad (X)	Presión (Y)	productos XY	Cuadrados	
				X ²	Y ²
1	19	122	2318	361	14884
2	25	125	3125	625	15625
3	30	126	3780	900	15876
4	42	129	5418	1764	16641
5	46	130	5980	2116	16900
6	52	135	7020	2704	18225
7	57	138	7866	3249	19044
8	62	142	8804	3844	20164
9	70	145	10150	4900	21025
10	73	146	10658	5329	21316
Σ	476	1 338	65 119	25 792	179 700

= 47.6;

$$\bar{y} = 133.8$$

$$\sum x^2 = 19^2 + 25^2 + 30^2 \dots \dots + 73^2 - \frac{476^2}{10} = 25792 - 22657.6 = 3134.4$$

$$\sum y^2 = 122^2 + 125^2 + 126^2 \dots \dots + 146^2 - \frac{1338^2}{10} = 179700 - 179024.4 = 675.6$$

$$\sum xy = 19x122 + 25x125 + 30x126 \dots \dots + 73x146 - \frac{476x 1338}{10} = 1430.2$$

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} = \frac{1430.2}{\sqrt{3134.4 x 675.6}} = \frac{1430.2}{1455.19} = 0.984$$

Por lo tanto, las hipótesis planteadas son:

Ho= p=0; no hay correlación

Ha= p≠0; hay correlación, por tanto; 0.983 > valor tabular de la correlación 0.01 (8) = 0.765**

Se puede concluir que existe una correlación altamente significativa con tendencia positiva es decir hay asociatividad fuerte entre la edad y la presión ($p < 0.01$), por tanto, la correlación es positiva.

La regresión se usa en la mayoría de ciencias e Ingenierías como; biología, química, medicina, agronomía etc. Por tanto, la regresión se expresa como ($y = a + bx$) se mide por medio de coeficiente de regresión. En una población se simboliza con β en una muestra con b . Su valor indica el incremento promedio de Y al aumentar X en la unidad. Se calcula por medio de la expresión.

$$by/x = b = \frac{\sum xy}{\sum x^2}$$

El término by/x se lee "regresión de Y sobre X "

Ejemplo:

Cálculo del coeficiente de regresión de la presión de la presión sanguínea (Y) sobre la edad (X) del ejemplo numérico descrito al calcular la correlación:

$$b = \frac{\sum xy}{\sum x^2} = \frac{1430.2}{3134.4} = 0.456$$

El valor de 0.456 es el incremento promedio de la presión sanguínea en mm Hg por cada año de edad,

$$bx = \frac{\sum xy}{\sum y^2} = \frac{1430.2}{675.6} = 2.12$$

El valor de 2.12 indica el incremento promedio por cada mm de Hg de presión sanguínea es de 2.12 años por tanto la ecuación lineal será:

$$\hat{X} = \bar{X} + (Y - \bar{Y})$$

$$\hat{X} = 47.6 + 2.12 (Y - 133.8)$$

$$\hat{X} = 2.12Y - 236.06$$

Se observa que en la edad de 70 años la presión sanguínea fue de 145, con esto notamos que hay una estrecha concordancia entre lo observado y lo teórico, pudiendo ser reflejado en la regresión, en muchos fenómenos biológicos es posible considerar indistintamente a Y o X como variables dependientes e independientes.

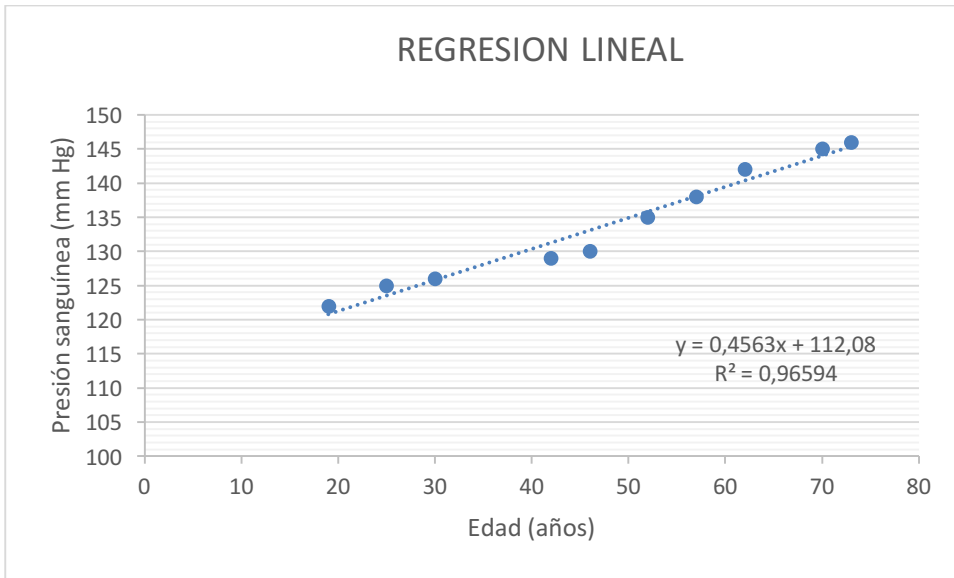


Figura 1. Diagrama de dispersión y líneas de regresión entre la presión versus la edad en humanos.

$$r = \sqrt{b \cdot bx/r} = \sqrt{0.4563 \times 2.12} = 0.984$$

Coeficiente de determinación (CD)

El coeficiente de determinación es un valor que se calcula usando en cuadrado del valor del coeficiente de correlación y se expresa en porcentaje.

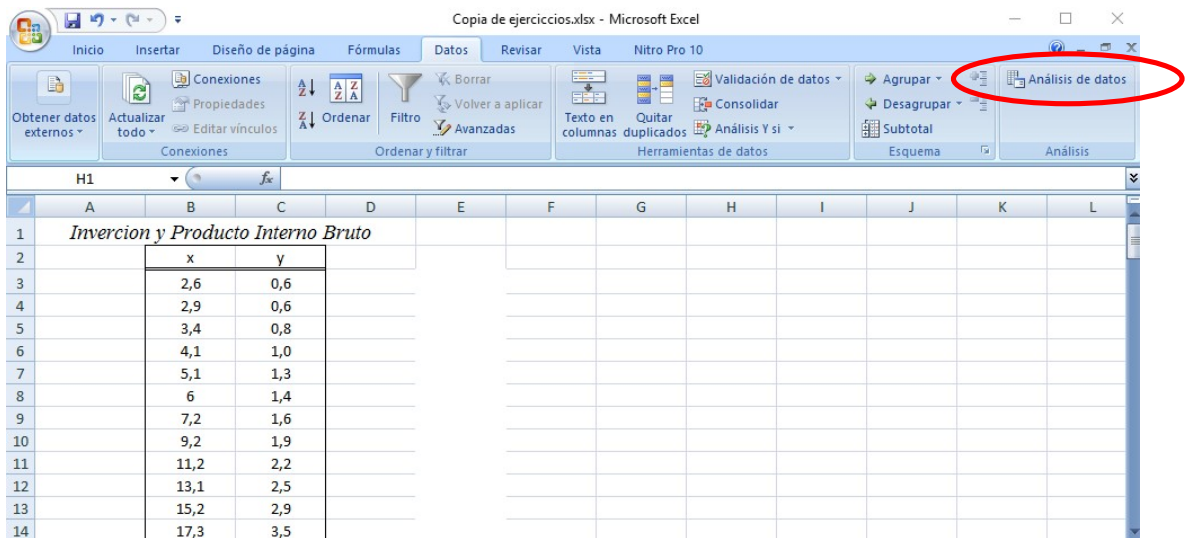
$$CD = 0.984^2 \times 100 = 96.63\% \text{ entonces el efecto medio} = 100 - 96.63 = 3.37\%$$

9.4. USANDO EL EXCEL E INFOSTAT

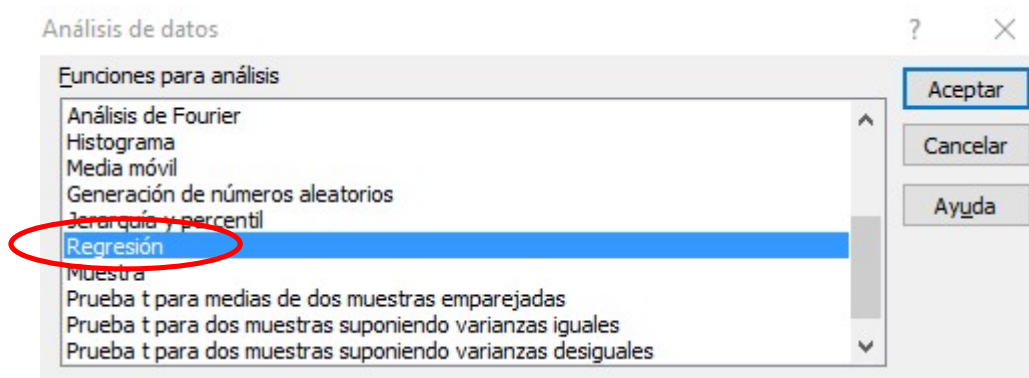
Se tabulan las variables X y Y en el Excel

A	B	C	D	E
<i>Invercion y Producto Interno Bruto</i>				
	x	y		
	2,6	0,6		
	2,9	0,6		
	3,4	0,8		
	4,1	1,0		
	5,1	1,3		
	6	1,4		
	7,2	1,6		
	9,2	1,9		
	11,2	2,2		
	13,1	2,5		
	15,2	2,9		
	17,3	3,5		
	19,9	3,9		

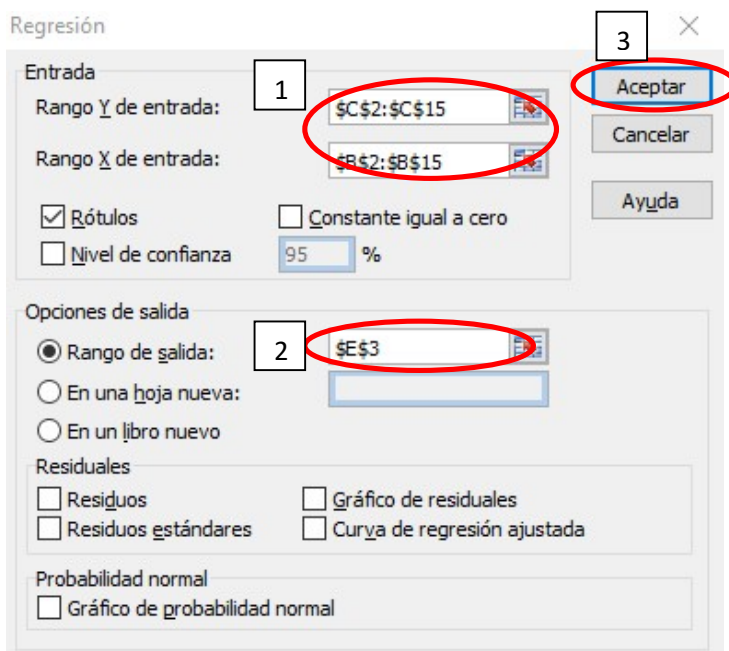
Escogemos la opción datos y luego elegimos Análisis de datos



Se elige la opción regresión lineal



- 1) Se seleccionan las variables para X y Y
- 2) se elige el rango de salida
- 3) Aceptar

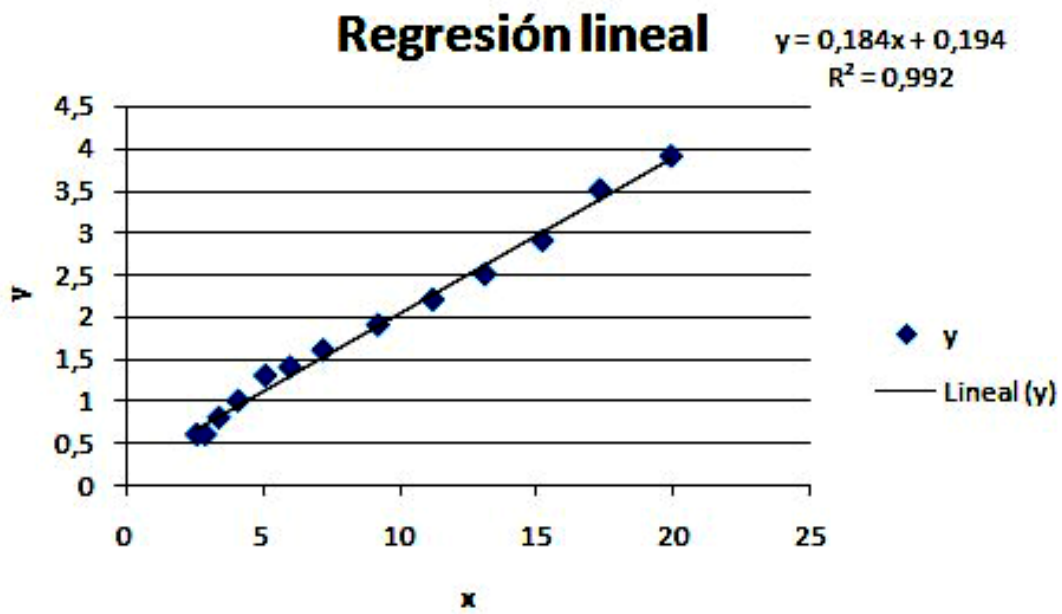
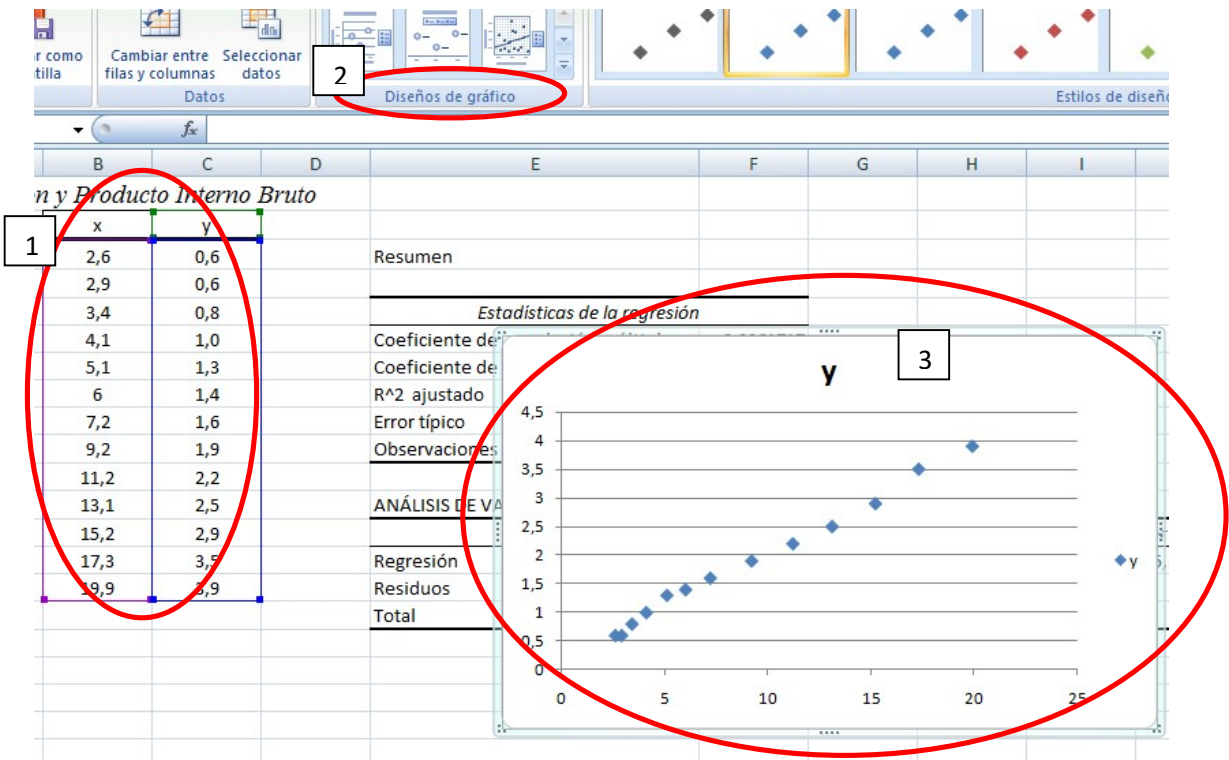


Finalmente se visualiza en ANDEVA de la REGRESION y al Correlación

Resumen						
<i>Estadísticas de la regresión</i>						
Coeficiente de correlación múltiple	0,9961717					
Coeficiente de determinación R ²	0,99235805					
R ² ajustado	0,99166333					
Error típico	0,09894026					
Observaciones	13					
ANÁLISIS DE VARIANZA						
		<i>Grados de libertad</i>	<i>de cuadrado de los cua</i>	<i>F</i>	<i>valor crítico de</i>	
Regresión	1	13,9830883	13,9830883	1428,42356	5,3755E-13	
Residuos	11	0,10768093	0,00978918			
Total	12	14,0907692				

Podemos observar un valor de F 1428.42356 versus el valor crítico pudiendo definir una alta significancia estadística y un coeficiente de $R^2 = 0.99$ lo cual indica una tendencia lineal positiva con una alto grado de asociación entre variables.

Si queremos visualizar la regresión hacemos el siguiente paso: 1) Sombreamos los datos 2) Ir a la opción insertar y elegimos gráficos de dispersión 3) Ir a Diseño de grafico opción $f(x)$ y visualizamos la regresión lineal con buen ajuste.



EN INFOSTAT

Pegamos los datos incluyendo nombre de columna

Caso	x	y
1	2,60	0,60
2	2,90	0,60
3	3,40	0,80
4	4,10	1,00
5	5,10	1,30
6	6,00	1,40
7	7,20	1,60
8	9,20	1,90
9	11,20	2,20
10	13,10	2,50
11	15,20	2,90
12	17,30	3,50
13	19,90	3,90

Opción estadísticas regresión lineal y separamos las variables en X y Y luego aceptamos

Análisis de regresión lineal

Caso

Variables Particiones ...

Variable dependiente

-> y

<-

Regresoras

-> x

<-

Pesos (solo una)

->

<-

1(0)

Seleccionar si contiene..

(...) [...] {...} {g}

Cancelar Limpiar

Aceptar

Análisis de regresión lineal

General Diagnóstico Polinomios Hipótesis Selección de modelos

Mostrar...

Coeficientes de regresión y estadísticos asociados

Tabla de análisis de la varianza SC Tipo I

Tabla resumen criterios de diagnóstico

Matriz de covarianzas de los coef. de regresión

Solo mostrar los coeficientes de regresión

Opciones

Ordenada al origen

Regresoras centradas

Prueba de Atkinson

Error puro

Todas las reg. simples

Método de Estimación

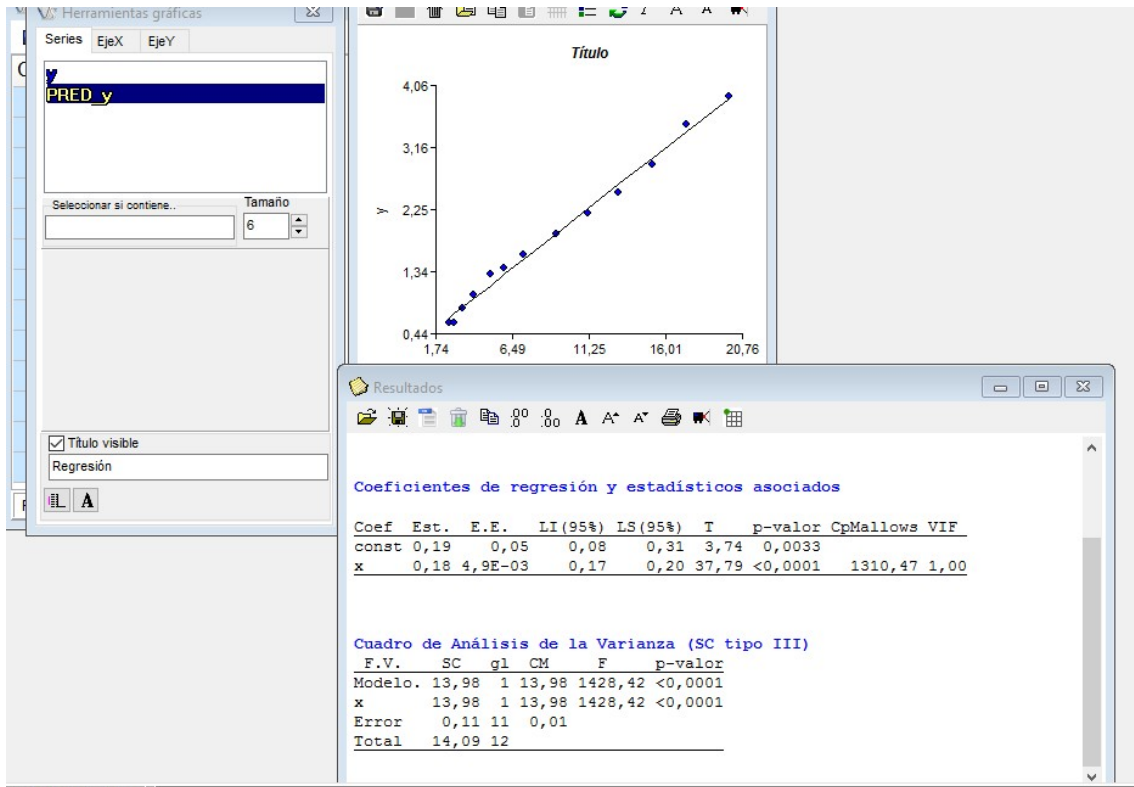
Mínimos cuadrados

Aceptar

Cancelar

Ayuda

Y visualizamos el análisis y el análisis de regresión.



BIBLIOGRAFÍA

Bibliografía

- CORRALES, L. 2013. *Curso de Estadística y técnicas experimentales para las ciencias biológicas*. Ecuador.
- SALTOS, A. 1993. *Diseño Experimental.- Aplicación de Procesos Tecnológicos*. Ambato Ecuador.
- SILVA E. 2012. *Curso de Diseños experimentales aplicados a las ciencias biológicas*. Ecuador.
- ALARCÓN, A. *Análisis sensorial y especial*
- COCHRAN, W. Y COX, G. M. 1987. *Diseños Experimentales*. Editorial Trillas, México. 661 p
- ALVARADO, J. 1996 *PRINCIPIOS DE INGENIERIA APLICADA A ALIMENTOS*. Secretaría General de la OEA. Programa de Desarrollo Científico y Tecnológico Multinacional de Biotecnología y Tecnología de Alimentos. Impreso por: Radio Comunicaciones.- Quito -Ecuador.
- REYES C., P. 2003. *Diseños de Experimentos Aplicados*. Trillas, México. 348 p.
- GONZÁLEZ B., G. 1974. *Métodos Estadísticos y Principios de Diseño Experimental*. Universidad Central del Ecuador, Quito.

TABLAS ESTADISTICAS

TABLA 1:	Distribución Normal tipificada
TABLA 2:	Distribución t de Student
TABLA 3:	Distribución F de Fisher
TABLA 4:	Prueba de rango múltiple de Duncan
TABLA 5:	Distribución Chi-Cuadrado
TABLA 6:	Arcoseno de la raíz cuadrada de datos expresados en porcentaje
TABLA 7:	Coefficientes de Correlación
TABLA 8	Coefficientes de polinomios ortogonales

TABLA 1 Distribución Normal tipificada

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7549
0.7	0.7580	0.7612	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916

2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9986	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	1	1	1

TABLA 2 Distribución t de Student

Prueba de 2 Colas	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.001
Prueba del Cola	0.25	0.2	0.15	0.10	0.05	0.025	0.01	0.005	0.0005
G.L.									
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.941
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.405
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.371	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.833
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792

23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	0.842	1.936	1.282	1.645	1.960	2.326	2.576	3.291

TABLA 3 Distribución F de Fisher

G.L.		G.L. DEL NUMERADOR								
ERROR	α	1	2	3	4	5	6	7	8	9
1	0.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	0.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
	0.025	647.79	799.48	864.15	899.60	921.83	937.11	948.20	956.64	963.28
	0.010	4052.18	4999.34	5403.53	5624.26	5763.96	5858.95	5928.33	5980.95	6022.40
	0.005	16212.46	19997.36	21614.13	22500.75	23055.82	23439.53	23715.20	23923.81	24091.45
	0.001	405311.58	499725.34	540256.50	562667.85	576496.12	586032.87	593185.42	597953.80	602245.33
2	0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
	0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
	0.010	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39
	0.005	198.50	199.01	199.16	199.24	199.30	199.33	199.36	199.38	199.39
	0.001	998.38	998.84	999.31	999.31	999.31	999.31	999.31	999.31	999.31
3	0.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34
	0.005	55.55	49.80	47.47	46.20	45.39	44.84	44.43	44.13	43.88
	0.001	167.06	148.49	141.10	137.08	134.58	132.83	131.61	130.62	129.86
4	0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	0.005	31.33	26.28	24.26	23.15	22.46	21.98	21.62	21.35	21.14

	0.001	74.13	61.25	56.17	53.43	51.72	50.52	49.65	49.00	48.47
5	0.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
	0.005	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77
	0.001	47.18	37.12	33.20	31.08	29.75	28.83	28.17	27.65	27.24
6	0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
	0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
	0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
	0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
	0.005	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39
	0.001	35.51	27.00	23.71	21.92	20.80	20.03	19.46	19.03	18.69
7	0.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
	0.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
	0.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
	0.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
	0.005	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51
	0.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33

TABLA 3 Distribución F de Fisher (continuación 1)

G.L.		G.L. DEL NUMERADOR								
ERROR	α	1	2	3	4	5	6	7	8	9
8	0.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
	0.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
	0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
	0.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
	0.005	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34
	0.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.77
9	0.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
	0.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
	0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
	0.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
	0.005	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54
	0.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11
10	0.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
	0.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
	0.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
	0.005	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
	0.001	21.04	14.90	12.55	11.28	10.48	9.93	9.52	9.20	8.96
11	0.100	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
	0.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
	0.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
	0.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63

	0.005	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54
	0.001	19.69	13.81	11.56	10.35	9.58	9.05	8.65	8.35	8.12
12	0.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
	0.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
	0.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
	0.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
	0.005	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20
	0.001	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48
13	0.100	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
	0.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
	0.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
	0.010	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
	0.005	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94
	0.001	17.82	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98
14	0.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
	0.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
	0.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
	0.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
	0.005	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72
	0.001	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.58

TABLA 3 Distribución F de Fisher (Continuación 2)

G.L. ERROR	α	G.L. DEL NUMERADOR								
		1	2	3	4	5	6	7	8	9
15	0.100	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
	0.050	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
	0.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12

	0.010	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
	0.005	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54
	0.001	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26
16	0.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
	0.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
	0.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
	0.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
	0.005	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38
	0.001	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.20	5.98
17	0.100	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
	0.050	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
	0.025	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
	0.010	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68
	0.005	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25
	0.001	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75
18	0.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
	0.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
	0.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
	0.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
	0.005	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14
	0.001	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56
19	0.100	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
	0.050	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
	0.025	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
	0.010	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
	0.005	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04
	0.001	15.08	10.16	8.28	7.27	6.62	6.18	5.85	5.59	5.39

20	0.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
	0.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
	0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
	0.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
	0.005	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96
	0.001	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24
21	0.100	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
	0.050	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
	0.025	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
	0.010	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
	0.005	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88
	0.001	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11

TABLA 3 Distribución F de Fisher (Continuación 3)

G.L. ERROR	α	G.L. DEL NUMERADOR								
		1	2	3	4	5	6	7	8	9
22	0.100	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
	0.050	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
	0.025	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
	0.010	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
	0.005	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81
	0.001	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99
23	0.100	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
	0.050	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
	0.025	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
	0.010	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30

	0.005	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75
	0.001	14.20	9.47	7.67	6.70	6.08	5.65	5.33	5.09	4.89
24	0.100	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
	0.050	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
	0.025	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
	0.010	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
	0.005	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69
	0.001	14.03	9.34	7.55	6.59	5.98	5.55	5.24	4.99	4.80
25	0.100	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
	0.050	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
	0.025	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
	0.010	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
	0.005	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64
	0.001	13.88	9.22	7.45	6.49	5.89	5.46	5.15	4.91	4.71
26	0.100	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
	0.050	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
	0.025	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
	0.010	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
	0.005	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60
	0.001	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64
27	0.100	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
	0.050	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
	0.025	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
	0.010	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
	0.005	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56
	0.001	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57
	0.100	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87

	0.050	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
28	0.025	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
	0.010	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
	0.005	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52
	0.001	13.50	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50

TABLA 3 Distribución F de Fisher (Continuación 4)

G.L.										
ERROR	α	1	2	3	4	5	6	7	8	9
29	0.100	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
	0.050	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
	0.025	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
	0.010	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
	0.005	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48
	0.001	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45
30	0.100	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
	0.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
	0.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
	0.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
	0.005	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45
	0.001	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39
40	0.100	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
	0.050	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
	0.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
	0.010	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
	0.005	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22
	0.001	12.61	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02
60	0.100	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
	0.050	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
	0.025	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
	0.010	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
	0.005	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01

	0.001	11.97	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69
120	0.100	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
	0.050	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
	0.025	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
	0.010	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
	0.005	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81
	0.001	11.38	7.32	5.78	4.95	4.42	4.04	3.77	3.55	3.38
∞	0.100	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63
	0.050	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88
	0.025	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11
	0.010	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41
	0.005	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62
	0.001	10.83	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10

TABLA 3 Distribución F de Fisher (Continuación 5)

G.L.		G.L. DEL NUMERADOR									
ERROR	α	10	12	15	20	24	30	40	60	120	∞
1	0.100	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
	0.050	241.88	243.90	245.95	248.02	249.05	250.10	251.14	252.20	253.25	254.32
	0.025	968.63	976.72	984.87	993.08	997.27	1001.40	1005.60	1009.79	1014.04	1018.26
	0.010	6055.93	6106.68	6156.97	6208.66	6234.27	6260.35	6286.43	6312.97	6339.51	6365.59
	0.005	24221.84	24426.73	24631.62	24836.51	24937.09	25041.40	25145.71	25253.74	25358.05	25466.08
	0.001	605583.19	610351.56	616073.61	620841.98	623703.00	626087.19	628471.37	631332.40	634193.42	636577.61
2	0.100	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
	0.050	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
	0.025	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
	0.010	99.40	99.42	99.43	99.45	99.46	99.47	99.48	99.48	99.49	99.50
	0.005	199.39	199.42	199.43	199.45	199.45	199.48	199.48	199.48	199.49	199.51
	0.001	999.31	999.31	999.31	999.31	999.31	999.31	999.31	999.31	999.31	999.31
3	0.100	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
	0.050	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
	0.025	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
	0.010	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
	0.005	43.68	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	41.83
	0.001	129.22	128.32	127.36	126.43	125.93	125.44	124.97	124.45	123.98	123.46
4	0.100	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
	0.050	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
	0.025	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
	0.010	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
	0.005	20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.32

	0.001	48.05	47.41	46.76	46.10	45.77	45.43	45.08	44.75	44.40	44.05
5	0.100	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.11
	0.050	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
	0.025	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
	0.010	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
	0.005	13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	12.14
	0.001	26.91	26.42	25.91	25.39	25.13	24.87	24.60	24.33	24.06	23.79
6	0.100	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
	0.050	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
	0.025	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
	0.010	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
	0.005	10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88
	0.001	18.41	17.99	17.56	17.12	16.90	16.67	16.44	16.21	15.98	15.75
7	0.100	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
	0.050	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
	0.025	4.76	4.67	4.57	4.47	4.41	4.36	4.31	4.25	4.20	4.14
	0.010	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
	0.005	8.38	8.18	7.97	7.75	7.64	7.53	7.42	7.31	7.19	7.08
	0.001	14.08	13.71	13.32	12.93	12.73	12.53	12.33	12.12	11.91	11.70

TABLA 3 Distribución F de Fisher (Continuación 6)

G.L. ERROR	α	G.L. DEL NUMERADOR									
		10	12	15	20	24	30	40	60	120	∞
8	0.100	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
	0.050	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
	0.025	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
	0.010	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
	0.005	7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95
	0.001	11.54	11.19	10.84	10.48	10.30	10.11	9.92	9.73	9.53	9.33
9	0.100	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
	0.050	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
	0.025	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
	0.010	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
	0.005	6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19
	0.001	9.89	9.57	9.24	8.90	8.72	8.55	8.37	8.19	8.00	7.81
10	0.100	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
	0.050	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
	0.025	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
	0.010	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
	0.005	5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64
	0.001	8.75	8.45	8.13	7.80	7.64	7.47	7.30	7.12	6.94	6.76
11	0.100	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
	0.050	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
	0.025	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
	0.010	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
	0.005	5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.45	4.34	4.23

	0.001	7.92	7.63	7.32	7.01	6.85	6.68	6.52	6.35	6.18	6.00
12	0.100	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
	0.050	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
	0.025	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
	0.010	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
	0.005	5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90
	0.001	7.29	7.00	6.71	6.40	6.25	6.09	5.93	5.76	5.59	5.42
13	0.100	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
	0.050	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
	0.025	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
	0.010	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
	0.005	4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65
	0.001	6.80	6.52	6.23	5.93	5.78	5.63	5.47	5.30	5.14	4.97
14	0.100	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
	0.050	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
	0.025	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
	0.010	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
	0.005	4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44
	0.001	6.40	6.13	5.85	5.56	5.41	5.25	5.10	4.94	4.77	4.60

TABLA 3 Distribución F de Fisher (Continuación 7)

G.L.		G.L. DEL NUMERADOR									
ERROR	α	10	12	15	20	24	30	40	60	120	∞
15	0.100	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
	0.050	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
	0.025	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
	0.010	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
	0.005	4.42	4.25	4.07	3.88	3.79	3.69	3.59	3.48	3.37	3.26
	0.001	6.08	5.81	5.54	5.25	5.10	4.95	4.80	4.64	4.48	4.31
16	0.100	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
	0.050	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
	0.025	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
	0.010	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
	0.005	4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11
	0.001	5.81	5.55	5.27	4.99	4.85	4.70	4.54	4.39	4.23	4.06
17	0.100	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
	0.050	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
	0.025	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
	0.010	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
	0.005	4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98
	0.001	5.58	5.32	5.05	4.78	4.63	4.48	4.33	4.18	4.02	3.85
18	0.100	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
	0.050	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
	0.025	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
	0.010	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
	0.005	4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87

	0.001	5.39	5.13	4.87	4.59	4.45	4.30	4.15	4.00	3.84	3.67
19	0.100	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
	0.050	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
	0.025	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
	0.010	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
	0.005	3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78
	0.001	5.22	4.97	4.70	4.43	4.29	4.14	3.99	3.84	3.68	3.51
20	0.100	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
	0.050	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
	0.025	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
	0.010	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
	0.005	3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69
	0.001	5.08	4.82	4.56	4.29	4.15	4.00	3.86	3.70	3.54	3.38
21	0.100	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
	0.050	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
	0.025	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
	0.010	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
	0.005	3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61
	0.001	4.95	4.70	4.44	4.17	4.03	3.88	3.74	3.58	3.42	3.26

TABLA 3 Distribución F de Fisher (Continuación 8)

G.L.		G.L. DEL NUMERADOR									
ERROR	α	10	12	15	20	24	30	40	60	120	∞
22	0.100	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
	0.050	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
	0.025	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
	0.010	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
	0.005	3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55
	0.001	4.83	4.58	4.33	4.06	3.92	3.78	3.63	3.48	3.32	3.15
23	0.100	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
	0.050	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
	0.025	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
	0.010	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
	0.005	3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48
	0.001	4.73	4.48	4.23	3.96	3.82	3.68	3.53	3.38	3.22	3.05
24	0.100	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
	0.050	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
	0.025	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
	0.010	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
	0.005	3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43
	0.001	4.64	4.39	4.14	3.87	3.74	3.59	3.45	3.29	3.14	2.97
25	0.100	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
	0.050	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
	0.025	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
	0.010	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
	0.005	3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38

	0.001	4.56	4.31	4.06	3.79	3.66	3.52	3.37	3.22	3.06	2.89
26	0.100	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
	0.050	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
	0.025	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
	0.010	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
	0.005	3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33
	0.001	4.48	4.24	3.99	3.72	3.59	3.44	3.30	3.15	2.99	2.82
27	0.100	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
	0.050	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
	0.025	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
	0.010	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
	0.005	3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.29
	0.001	4.41	4.17	3.92	3.66	3.52	3.38	3.23	3.08	2.92	2.75
28	0.100	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
	0.050	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
	0.025	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
	0.010	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
	0.005	3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.25
	0.001	4.35	4.11	3.86	3.60	3.46	3.32	3.18	3.02	2.86	2.69

TABLA 3 Distribución F de Fisher (Continuación 9)

G.L.		G.L. DEL NUMERADOR									
ERROR	α	10	12	15	20	24	30	40	60	120	∞
29	0.100	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
	0.050	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
	0.025	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
	0.010	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
	0.005	3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.21
	0.001	4.29	4.05	3.80	3.54	3.41	3.27	3.12	2.97	2.81	2.64
30	0.100	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
	0.050	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
	0.025	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
	0.010	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
	0.005	3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18
	0.001	4.24	4.00	3.75	3.49	3.36	3.22	3.07	2.92	2.76	2.59
40	0.100	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
	0.050	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
	0.025	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
	0.010	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
	0.005	3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93
	0.001	3.87	3.64	3.40	3.15	3.01	2.87	2.73	2.57	2.41	2.23
60	0.100	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
	0.050	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
	0.025	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
	0.010	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
	0.005	2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	1.69

	0.001	3.54	3.32	3.08	2.83	2.69	2.55	2.41	2.25	2.08	1.89
120	0.100	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
	0.050	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
	0.025	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
	0.010	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
	0.005	2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	1.43
	0.001	3.24	3.02	2.78	2.53	2.40	2.26	2.11	1.95	1.77	1.54
∞	0.100	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00
	0.050	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00
	0.025	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00
	0.010	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00
	0.005	2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00
	0.001	2.96	2.74	2.51	2.27	2.13	1.99	1.84	1.66	1.45	1.00

TABLA 4 Prueba de rango múltiple de Duncan

G.L. ERROR	α	NUMERO DE MEDIAS CONSIDERADAS													
		2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	0.05	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
	0.01	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	0.05	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
	0.01	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	0.05	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
	0.01	8.26	8.50	8.60	8.70	8.80	8.90	8.90	9.00	9.00	9.00	9.10	9.20	9.30	9.30
4	0.05	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
	0.01	6.51	6.80	6.90	7.00	7.10	7.1	7.20	7.20	7.30	7.30	7.40	7.40	7.50	7.50
5	0.05	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
	0.01	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.50	6.60	6.60	6.70	6.70	6.80
6	0.05	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
	0.01	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.00	6.10	6.20	6.20	6.30	6.30
7	0.05	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61
	0.01	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.80	5.80	5.90	5.90	6.00	6.00
8	0.05	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56
	0.01	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.50	5.60	5.70	5.70	5.80	5.80
9	0.05	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52
	0.01	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.40	5.50	5.50	5.60	5.70	5.70
10	0.05	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.48
	0.01	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36	5.42	5.48	5.54	5.55
11	0.05	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.46	3.46	3.46	3.47	3.48
	0.01	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24	5.28	5.34	5.38	5.39

12	0.05	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46	3.46	3.46	3.47	3.48
	0.01	4.32	4.55	4.68	4.76	4.84	4.92	4.96	5.02	5.07	5.13	5.17	5.22	5.24	5.26
13	0.05	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.45	3.46	3.46	3.47	3.47
	0.01	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04	5.08	5.13	5.14	5.15
14	0.05	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47
	0.01	4.21	4.42	4.55	4.63	4.70	4.78	4.83	4.87	4.91	4.96	5.00	5.04	5.06	5.07
15	0.05	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47
	0.01	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90	4.94	4.97	4.99	5.00
16	0.05	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47
	0.01	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.84	4.88	4.91	4.93	4.94
17	0.05	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47
	0.01	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80	4.83	4.86	4.88	4.89
18	0.05	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47
	0.01	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76	4.79	4.82	4.84	4.85
19	0.05	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47
	0.01	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67	4.72	4.76	4.79	4.81	4.82
20	0.05	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	3.46	3.47
	0.01	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69	4.73	4.76	4.78	4.79

TABLA 4 Prueba de rango múltiple de Duncan (Continuación 2)

G.L. ERROR	α	NUMERO DE MEDIAS CONSIDERADAS													
		2	3	4	5	6	7	8	9	10	12	14	16	18	20
22	0.05	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.46	3.47
	0.01	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65	4.64	4.71	4.74	4.75
24	0.05	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.46	3.47
	0.01	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62	4.64	4.67	4.70	4.72
26	0.05	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47
	0.01	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58	4.62	4.65	4.67	4.69
28	0.05	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47
	0.01	3.91	4.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56	4.60	4.62	4.65	4.67
30	0.05	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47
	0.01	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54	4.58	4.61	4.63	4.65
40	0.05	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47
	0.01	3.82	3.99	4.10	4.17	4.24	4.30	4.34	4.37	4.41	4.46	4.51	4.54	4.54	4.59
60	0.05	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47
	0.01	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39	4.44	4.47	4.50	4.53
100	0.05	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47
	0.01	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.35	4.38	4.42	4.45	4.48
∞	0.05	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47
	0.01	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26	4.31	4.34	4.38	4.41

TABLA 5 Distribución Chi-Cuadrado

G.L.	0.995	0.990	0.975	0.950	0.900	0.750	0.500	0.250	0.100	0.050	0.025	0.010	0.005	0.001
1	0.001	0.001	0.001	0.004	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88	10.83
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60	13.82
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86	18.47
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75	20.51
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55	22.46
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28	24.32
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.95	26.12
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.73	26.76	31.26
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	7.434	8.260	9.591	10.851	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00	45.31
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40	46.80
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80	48.27
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18	49.73
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56	51.18
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93	52.62

26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29	54.05
27	11.81	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.19	46.96	49.65	55.48
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99	56.89
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34	58.30
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.81	55.76	59.34	63.69	66.77	73.40
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49	86.66
60	35.53	37.48	40.48	43.19	46.46	52.29	59.33	66.98	74.40	79.08	83.30	88.38	91.95	99.61
70	43.28	45.44	48.76	51.74	55.33	61.70	69.33	77.58	85.53	90.53	95.02	100.43	104.21	112.32
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33	88.13	96.58	101.88	106.63	112.33	116.32	124.84
90	59.20	61.75	65.65	69.13	73.29	80.62	89.33	98.65	107.57	113.15	118.14	124.12	128.30	137.21
100	67.33	70.06	74.22	77.93	82.36	90.13	99.33	109.14	118.50	124.34	129.56	135.81	140.17	149.45
120	83.85	86.92	91.57	95.70	100.62	109.22	119.33	130.05	140.23	146.57	152.21	158.95	163.65	173.62

TABLA 6 Arcoseno de la raíz cuadrada de datos expresados en porcentaje.

%	0	1	2	3	4	5	6	7	8	9
0	0.00	1.81	2.56	3.14	3.63	4.05	4.44	4.80	5.13	5.44
1	5.74	6.02	6.29	6.55	6.80	7.03	7.27	7.49	7.71	7.92
2	8.13	8.33	8.53	8.72	8.91	9.10	9.28	9.46	9.63	9.80
3	9.97	10.14	10.30	10.47	10.63	10.78	10.94	11.09	11.24	11.39
4	11.54	11.68	11.83	11.97	12.11	12.25	12.38	12.52	12.66	12.79
5	12.92	13.05	13.18	13.31	13.44	13.56	13.69	13.81	13.94	14.06
6	14.18	14.30	14.42	14.54	14.65	14.77	14.89	15.00	15.12	15.23
7	15.34	15.45	15.56	15.68	15.79	15.89	16.00	16.11	16.22	16.32
8	16.43	16.54	16.64	16.74	16.85	16.95	17.05	17.15	17.26	17.36
9	17.46	17.56	17.66	17.76	17.85	17.95	18.05	18.15	18.24	18.34
10	18.43	18.53	18.63	18.72	18.81	18.91	19.00	19.09	19.19	19.28
11	19.37	19.46	19.55	19.64	19.73	19.82	19.91	20.00	20.09	20.18
12	20.27	20.36	20.44	20.53	20.62	20.70	20.79	20.88	20.96	21.05
13	21.13	21.22	21.30	21.39	21.47	21.56	21.64	21.72	21.81	21.89
14	21.97	22.06	22.14	22.22	22.30	22.38	22.46	22.54	22.63	22.71
15	22.79	22.87	22.95	23.03	23.11	23.18	23.26	23.34	23.42	23.50
16	23.58	23.66	23.73	23.81	23.89	23.97	24.04	24.12	24.20	24.27
17	24.35	24.43	24.50	24.58	24.65	24.73	24.80	24.88	24.95	25.03
18	25.10	25.18	25.25	25.33	25.40	25.47	25.55	25.62	25.70	25.77
19	25.84	25.91	25.99	26.06	26.13	26.21	26.28	26.35	26.42	26.49
20	26.56	26.64	26.71	26.78	26.85	26.92	26.99	27.06	27.13	27.20
21	27.27	27.34	27.42	27.49	27.56	27.62	27.69	27.76	27.83	27.90
22	27.97	28.04	28.11	28.18	28.25	28.32	28.38	28.45	28.52	28.59
23	28.66	28.73	28.79	28.86	28.93	29.00	29.06	29.13	29.20	29.27

24	29.33	29.40	29.47	29.53	29.60	29.67	29.73	29.80	29.87	29.93
25	30.00	30.07	30.13	30.20	30.26	30.33	30.40	30.46	30.53	30.59
26	30.66	30.72	30.79	30.85	30.92	30.98	31.05	31.11	31.18	31.24
27	31.31	31.37	31.44	31.50	31.56	31.63	31.69	31.76	31.82	31.88
28	31.95	32.01	32.08	32.14	32.20	32.27	32.33	32.39	32.46	32.52
29	32.58	32.65	32.71	32.77	32.83	32.90	32.96	33.02	33.09	33.15
30	33.21	33.27	33.34	33.40	33.46	33.52	33.58	33.65	33.71	33.77
31	33.83	33.89	33.96	34.02	34.08	34.14	34.20	34.27	34.33	34.39
32	34.45	34.51	34.57	34.63	34.70	34.76	34.82	34.88	34.94	35.00
33	35.06	35.12	35.18	35.24	35.30	35.37	35.43	35.49	35.55	35.61
34	35.67	35.73	35.79	35.85	35.91	35.97	36.03	36.09	36.15	36.21
35	36.27	36.33	36.39	36.45	36.51	36.57	36.63	36.69	36.75	36.81
36	36.87	36.93	36.99	37.05	37.11	37.17	37.23	37.29	37.35	37.41
37	37.46	37.52	37.58	37.64	37.70	37.76	37.82	37.88	37.94	38.00
38	38.06	38.12	38.17	38.23	38.29	38.35	38.41	38.47	38.53	38.59
39	38.65	38.70	38.76	38.82	38.88	38.94	39.00	39.06	39.11	39.17
40	39.23	39.29	39.35	39.41	39.47	39.52	39.58	39.64	39.70	39.76
41	39.82	39.87	39.93	39.99	40.05	40.11	40.16	40.22	40.28	40.34
42	40.40	40.45	40.51	40.57	40.63	40.69	40.74	40.80	40.86	40.92
43	40.98	41.03	41.09	41.15	41.21	41.27	41.32	41.38	41.44	41.50
44	41.55	41.61	41.67	41.73	41.78	41.84	41.90	41.96	42.02	42.07
45	42.13	42.19	42.25	42.30	42.36	42.42	42.48	42.53	42.59	42.65

TABLA 6 Arcoseno de la raíz cuadrada de datos expresados en porcentaje (Continuación 1)

%	0	1	2	3	4	5	6	7	8	9
46	42.71	42.76	42.82	42.88	42.94	42.99	43.05	43.11	43.17	43.22
47	43.28	43.34	43.39	43.45	43.51	43.57	43.62	43.68	43.74	43.80
48	43.85	43.91	43.97	44.03	44.08	44.14	44.20	44.25	44.31	44.37
49	44.43	44.48	44.54	44.60	44.66	44.71	44.77	44.83	44.89	44.94
50	45.00	45.06	45.11	45.17	45.23	45.29	45.34	45.40	45.46	45.52
51	45.57	45.63	45.69	45.74	45.80	45.86	45.92	45.97	46.03	46.09
52	46.15	46.20	46.26	46.32	46.38	46.43	46.49	46.55	46.61	46.66
53	46.72	46.78	46.83	46.89	46.95	47.01	47.06	47.12	47.18	47.24
54	47.29	47.35	47.41	47.47	47.52	47.58	47.64	47.70	47.75	47.81
55	47.87	47.93	47.98	48.04	48.10	48.16	48.22	48.27	48.33	48.39
56	48.45	48.50	48.56	48.62	48.68	48.73	48.79	48.85	48.91	48.97
57	49.02	49.08	49.14	49.20	49.26	49.31	49.37	49.43	49.49	49.55
58	49.60	49.66	49.72	49.78	49.84	49.89	49.95	50.01	50.07	50.13
59	50.18	50.24	50.30	50.36	50.42	50.48	50.53	50.59	50.65	50.71
60	50.77	50.83	50.89	50.94	51.00	51.06	51.12	51.18	51.24	51.30
61	51.35	51.41	51.47	51.53	51.59	51.65	51.71	51.77	51.83	51.88
62	51.94	52.00	52.06	52.12	52.18	52.24	52.30	52.36	52.42	52.48
63	52.53	52.59	52.65	52.71	52.77	52.83	52.89	52.95	53.01	53.07
64	53.13	53.19	53.25	53.31	53.37	53.43	53.49	53.55	53.61	53.67
65	53.73	53.79	53.85	53.91	53.97	54.03	54.09	54.15	54.21	54.27
66	54.33	54.39	54.45	54.51	54.57	54.63	54.69	54.76	54.82	54.88
67	54.94	55.00	55.06	55.12	55.18	55.24	55.30	55.37	55.43	55.49
68	55.55	55.61	55.67	55.73	55.80	55.86	55.92	55.98	56.04	56.10
69	56.17	56.23	56.29	56.35	56.41	56.48	56.54	56.60	56.66	56.73
70	56.79	56.85	56.91	56.98	57.04	57.10	57.17	57.23	57.29	57.35

71	57.42	57.48	57.54	57.61	57.67	57.73	57.80	57.86	57.92	57.99
72	58.05	58.12	58.18	58.24	58.31	58.37	58.44	58.50	58.56	58.63
73	58.69	58.76	58.82	58.89	58.95	59.02	59.08	59.15	59.21	59.28
74	59.34	59.41	59.47	59.54	59.60	59.67	59.74	59.80	59.87	59.93
75	60.00	60.07	60.13	60.20	60.27	60.33	60.40	60.47	60.53	60.60
76	60.67	60.73	60.80	60.87	60.94	61.00	61.07	61.14	61.21	61.27
77	61.34	61.41	61.48	61.55	61.61	61.68	61.75	61.82	61.89	61.96
78	62.03	62.10	62.17	62.24	62.31	62.37	62.44	62.51	62.58	62.65
79	62.73	62.80	62.87	62.94	63.01	63.08	63.15	63.22	63.29	63.36
80	63.43	63.51	63.58	63.65	63.72	63.79	63.87	63.94	64.01	64.08
81	64.16	64.23	64.30	64.38	64.45	64.52	64.60	64.67	64.75	64.82
82	64.90	64.97	65.05	65.12	65.20	65.27	65.35	65.42	65.50	65.57
83	65.65	65.73	65.80	65.88	65.96	66.03	66.11	66.19	66.27	66.34
84	66.42	66.50	66.58	66.66	66.74	66.81	66.89	66.97	67.05	67.13
85	67.21	67.29	67.37	67.46	67.54	67.62	67.70	67.78	67.86	67.94
86	68.03	68.11	68.19	68.28	68.36	68.44	68.53	68.61	68.70	68.78
87	68.87	68.95	69.04	69.12	69.21	69.30	69.38	69.47	69.56	69.64
88	69.73	69.82	69.91	70.00	70.09	70.18	70.27	70.36	70.45	70.54
89	70.63	70.72	70.81	70.91	71.00	71.09	71.19	71.28	71.37	71.47
90	71.56	71.66	71.76	71.85	71.95	72.05	72.15	72.24	72.34	72.44
91	72.54	72.64	72.74	72.84	72.95	73.05	73.15	73.26	73.36	73.46

TABLA 6 Arcoseno de la raíz cuadrada de datos expresados en porcentaje (Continuación 2)

%	0	1	2	3	4	5	6	7	8	9
92	73.57	73.68	73.78	73.89	74.00	74.11	74.21	74.32	74.44	74.55

93	74.66	74.77	74.88	75.00	75.11	75.23	75.35	75.46	75.58	75.70
94	75.82	75.94	76.06	76.19	76.31	76.44	76.56	76.69	76.82	76.95
95	77.08	77.21	77.34	77.48	77.62	77.75	77.89	78.03	78.17	78.32
96	78.46	78.61	78.76	78.91	79.06	79.22	79.37	79.53	79.69	79.86
97	80.03	80.19	80.37	80.54	80.72	80.90	81.09	81.28	81.47	81.67
98	81.87	82.08	82.29	82.51	82.73	82.96	83.20	83.45	83.71	83.98
99	84.26	84.56	84.87	85.20	85.56	85.94	86.37	86.86	87.44	88.19
100	90.00									

TABLA 7 Coeficientes de Correlación.

G.L. del Error	α	Variables independientes								
		1	2	3	4		1	2	3	4
1	0.05	0.997	0.999	0.999	0.999	10	0.576	0.671	0.726	0.763
	0.01	1.000	1.000	1.000	1.000		0.708	0.776	0.814	0.840
2	0.05	0.950	0.975	0.983	0.987	11	0.553	0.648	0.703	0.741
	0.01	0.990	0.995	0.997	0.998		0.684	0.753	0.793	0.821
3	0.05	0.878	0.930	0.950	0.961	12	0.532	0.627	0.683	0.722
	0.01	0.959	0.976	0.983	0.987		0.661	0.732	0.773	0.802
4	0.05	0.811	0.881	0.912	0.930	13	0.514	0.608	0.664	0.703
	0.01	0.917	0.949	0.962	0.970		0.641	0.712	0.755	0.785
5	0.05	0.754	0.836	0.874	0.898	14	0.497	0.590	0.646	0.686
	0.01	0.874	0.917	0.937	0.949		0.623	0.694	0.737	0.768
6	0.05	0.707	0.795	0.839	0.867	15	0.482	0.574	0.630	0.670
	0.01	0.834	0.886	0.911	0.927		0.606	0.677	0.721	0.752
7	0.05	0.666	0.758	0.807	0.838	16	0.468	0.559	0.615	0.655
	0.01	0.798	0.855	0.885	0.904		0.590	0.662	0.706	0.738
8	0.05	0.632	0.726	0.777	0.811	17	0.456	0.545	0.601	0.641
	0.01	0.765	0.827	0.860	0.882		0.575	0.647	0.691	0.724
9	0.05	0.602	0.697	0.750	0.786	18	0.444	0.532	0.587	0.628
	0.01	0.735	0.800	0.836	0.861		0.561	0.633	0.678	0.710
19	0.05	0.433	0.520	0.575	0.615	28	0.361	0.439	0.490	0.529
	0.01	0.549	0.620	0.665	0.698		0.463	0.530	0.573	0.605
20	0.05	0.423	0.509	0.563	0.604	29	0.355	0.432	0.482	0.521
	0.01	0.537	0.608	0.652	0.685		0.456	0.522	0.565	0.598
21	0.05	0.413	0.498	0.522	0.592	30	0.349	0.426	0.476	0.514

	0.01	0.526	0.596	0.641	0.674		0.449	0.514	0.558	0.591
22	0.05	0.404	0.488	0.542	0.582	35	0.325	0.397	0.445	0.482
	0.01	0.515	0.585	0.630	0.663		0.418	0.481	0.523	0.556
23	0.05	0.396	0.479	0.532	0.572	40	0.304	0.373	0.419	0.455
	0.01	0.505	0.574	0.619	0.652		0.393	0.454	0.494	0.526
24	0.05	0.388	0.470	0.523	0.562	45	0.288	0.353	0.397	0.432
	0.01	0.496	0.565	0.609	0.642		0.372	0.430	0.470	0.501
25	0.05	0.381	0.462	0.514	0.553	50	0.273	0.336	0.379	0.412
	0.01	0.487	0.555	0.600	0.633		0.354	0.410	0.449	0.479
26	0.05	0.374	0.454	0.506	0.545	60	0.250	0.308	0.348	0.380
	0.01	0.478	0.546	0.590	0.624		0.325	0.377	0.414	0.442
27	0.05	0.367	0.446	0.498	0.536	70	0.232	0.286	0.324	0.354
	0.01	0.470	0.538	0.582	0.615		0.302	0.351	0.386	0.413

TABLA 7 Coeficientes de Correlación (Continuación 2)

G.L. del		Variables independientes			
Error	α	1	2	3	4
80	0.05	0.217	0.269	0.304	0.332
	0.01	0.283	0.330	0.362	0.389
90	0.05	0.205	0.254	0.288	0.315
	0.01	0.267	0.312	0.343	0.368
100	0.05	0.195	0.241	0.274	0.300
	0.01	0.254	0.297	0.327	0.351
150	0.05	0.159	0.198	0.225	0.247
	0.01	0.208	0.244	0.270	0.290
200	0.05	0.138	0.172	0.196	0.215
	0.01	0.181	0.212	0.234	0.253
300	0.05	0.113	0.141	0.160	0.176
	0.01	0.148	0.174	0.192	0.208
400	0.05	0.098	0.122	0.139	0.153
	0.01	0.128	0.151	0.167	0.180
500	0.05	0.088	0.109	0.124	0.137
	0.01	0.115	0.135	0.150	0.162
1000	0.05	0.062	0.077	0.088	0.097
	0.01	0.081	0.096	0.106	0.115

TABLA 8 Coeficientes de polinomios ortogonales

Coeficientes

	Tratamiento	Grado	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉	a ₁₀	Coef2	
<u>Lineal</u>	2	1	-1	+1									2	
	3	1	-1	0	1								2	
	4	1	-3	-1	1	3							20	
	5	1	-2	-1	0	1	2						10	
	6	1	-5	-3	-1	1	3	5					70	
	7	1	-3	-2	-1	0	1	2	3				28	
	8	1	-7	-5	-3	-1	1	3	5	7			168	
	9	1	-4	-3	-2	-1	0	1	2	3	4		60	
	10	1	-9	-7	-5	-3	-1	1	3	5	7	9	330	
	<u>Cuadrática</u>	3	2	1	-2	1								6
4		2	1	-1	-1	1							4	
5		2	2	-1	-2	-1	2						14	
6		2	5	-1	-4	-4	-1	5					84	
7		2	5	0	-3	-4	-3	0	5				84	
8		2	7	1	-3	-5	-5	-3	1	7			168	
9		2	28	7	-8	-17	-20	-17	-8	7	28		2.772	
10		2	6	2	-1	-3	-4	-4	-3	-1	2	6	132	
<u>Cúbica</u>		4	3	-1	3	-3	1							20
		5	3	-1	2	0	-2	1						10
	6	3	-5	7	4	-4	-7	5					180	
	7	3	-1	1	1	0	-1	-1	1				6	
	8	3	-7	5	7	3	-3	-7	-5	7			264	
	9	3	-14	7	13	9	0	-9	-13	-7	14		990	
	10	3	-42	14	35	31	12	-12	-31	-35	-14	42	8.580	
<u>Cuártica</u>	5	4	1	-4	6	-4	1						70	

Quíntica

6	4	1	-3	2	2	-3	1					28
7	4	3	-7	1	6	1	-7	3				154
8	4	7	-13	-3	9	9	-3	-13	7			616
9	4	14	-21	-11	9	18	9	-11	-21	14		2.002
10	4	18	-22	-17	3	18	18	3	-17	-22	18	2.860
6	5	-1	5	-10	10	-5	1					252
7	5	-1	4	-5	0	5	-4	1				84
8	5	-7	23	-17	-15	15	17	-23	7			2.184
9	5	-4	11	-4	-9	0	9	4	-11	4		468
10	5	-6	14	-1	-11	-6	6	11	1	14	6	780

Nota: Las tablas presentadas en este documento han sido adaptadas de la literatura citada

Tabla de números aleatorios

10097	85017	84532	13618	23157	86952	02438	76520	91499	38631	79430	62421	97959	67422	69992	68479
37542	16719	82789	69041	05545	44109	05403	64894	80336	49172	16332	44670	35089	17691	89246	26940
08422	65842	27672	82186	14871	22115	86529	19645	44104	89232	57327	34679	62235	79655	81336	85157
99019	76875	20684	39187	38976	94324	43204	09376	12550	02844	15026	32439	58537	48274	81330	11100
12807	93640	39160	41453	97312	41548	93137	80157	63606	40387	65406	37920	08709	60623	02237	16505
66065	99478	70086	71265	11742	18226	29004	34072	61196	80240	44177	51171	08723	39323	05798	26457
31060	65119	26486	47353	43361	99436	42753	45571	15474	44910	99321	72173	56239	04595	10836	95270
85269	70322	21592	48233	93806	32584	21828	02051	94557	33663	86347	00926	44915	34823	51770	67897
63573	58133	41278	11697	49540	61777	67954	05325	42481	86430	19102	37420	41976	76559	24358	97344
73796	44655	81255	31133	36768	60452	38537	03529	23529	31379	68588	81675	15694	43438	36879	73208
98520	02295	13487	98662	07092	44673	61303	14905	04493	98086	32533	17767	14523	52494	24826	75246
11805	85035	54881	35587	43310	48897	48493	39808	00549	33185	04805	05431	94598	97654	16232	64051
83452	01197	86935	28021	61570	23350	65710	06288	35963	80951	68953	99624	81949	15307	00406	26898
88685	97907	19078	40646	31352	48625	44369	86507	59808	79752	02529	40200	73742	08391	49140	45427
99594	63268	96905	28797	57048	46359	74294	87517	46058	18633	99970	67348	49329	95236	32537	01390
65481	52841	59684	67411	09243	56092	84369	17468	32179	74029	74717	17674	90446	00597	45240	87379
80124	55722	71399	10916	07959	21225	13018	17727	69234	54178	10805	35635	45266	61406	41941	20117
74350	11434	51908	62171	93732	26958	02400	77402	19565	11664	77602	99817	28573	41430	96382	01758
69916	62375	99292	21177	72721	66995	07289	66252	45155	48324	32135	26803	16213	14938	71961	19476
09893	28337	20923	87929	61020	62841	31374	14225	94864	69074	45753	20505	78317	31994	98145	36168

Ing. Jaime Fabián Vera Chang

ORCID: <http://orcid.org/0000-0001-6105-3366>

Ingeniero Agropecuario graduado en la Universidad Técnica Estatal de Quevedo, Magíster en Procesamiento de Alimentos en la Universidad Agraria del Ecuador. Presidente del comité del Comité de Investigación de la Facultad de Ciencias Pecuarias periodo (2016- hasta la actualidad). Ex miembro del Comité de Investigación de la Facultad de Ciencias Pecuarias (2011-2013) y el periodo (2015-2016). Ex investigador de la Unidad de Investigación Científica y tecnológica de la UTEQ (2008-2013). Ex asistente investigador en el Laboratorio de Biotecnología en enero (2007-2008). Investigador del área agrícola, ciclo perenne cacao desde enero 2009 hasta 2013. Docente Facultad de Ciencias Pecuarias, Carreras de Ingeniería en Alimentos, e Ingeniería Agropecuaria, hasta la actualidad. Ha realizado y participado aproximadamente en 90 seminarios, conferencias y cursos nacionales e internacionales. Director de dos proyectos de investigación FOCICYT-UTEQ en el Área de calidad y post-cosecha del cacao y uso de residuos para uso alimentario. Arbitraje Internacional de artículos científicos del Publindex de Universidad Francisco de Paula Santander- Colciencias-Colombia. Y Revista de Biología Tropical/ International Journal of Tropical Biology and Conservation Universidad de Costa. Par externo de la Revista el Misionero de la Universidad Agraria del Ecuador, Revista Ciencia y Tecnología de la Universidad Técnica Estatal de Quevedo. Reconocimiento en calidad de tutor en el área temática Agricultura, Silvicultura y Veterinaria el tercer lugar en el concurso Galardones Nacionales del Senescyt 2017. Guayaquil – Ecuador, por haber participado en calidad de grupo académico para la aprobación del rediseño de la carrera de Ingeniería en Alimentos Resolución RCP-SO-25 N° 426-2016, por 10 años de servicio en la Universidad Técnica Estatal de Quevedo. Profesor no titular de las materias: Diseño Experimental, Metodología de la Investigación Científica, Métodos Estadísticos, Paquetes Estadísticos, Probabilidad y Estadística, Estadística descriptiva e Inferencial. Calidad e Inocuidad de los Alimentos, Tecnología de Frutas y Hortalizas, Cadenas Agroalimentarias y Elaboración de productos agropecuarios.

ISBN: 978-9942-33-085-7

