Métodos estadísticos para la Investigación Científica

Dr. Alejandro Néstor Salas Begazo

compás

Métodos estadísticos para la Investigación Científica

Autor:

Dr. Alejandro Néstor Salas Begazo

Métodos estadísticos para la
Investigación Científica

Autor:
Dr. Alejandro Néstor Salas Begazo

Primera edición: octubre 2018

ISBN: 978-9942-33-046-8

Diseño de portada y diagramación:
Equipo Editorial Grupo Compás

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

Cita.
Salas, A. (2018) Métodos estadísticos para la Investigación Científica,Editorial Grupo Compás, Guayaquil Ecuador, 161pag

A la memoria de mis Padres

Alejandro y Benigna

Mientras más inteligente es un hombre, más originalidad encuentra en el hombre. Las personas ordinarias no encuentran diferencias entre los hombres.

Blaise Pascal

INTRODUCCIÓN

Ante la demanda por parte de la sociedad de tener conocimientos para mostrar los acontecimientos que se realizan en nuestra vida cotidiana, así como la interpretación de los mismos, es necesaria la aplicación de métodos estadísticos que permitan resolver estas problemáticas.

En la actualidad, los métodos estadísticos se hacen indispensables como un procedimiento eficaz para la interpretación exacta de información de todas las ciencias y sirve como una herramienta de enlace para cada información.

En este sentido el presente libro Métodos Estadísticos para la Investigación Científica, proporciona una guía amigable para el procesamiento y resolución de problemas con grande data o experimentales con poca data, con problemas resueltos y propuestos que abarcan cada tipo de procesamiento de datos.

La metodología del libro es interactiva y no se focaliza en la demostración de las formulas sino en la aplicación de las mismas para la investigación de procesos físicos e experimentales y así tener una formación integral de nuestros alumnos. Lo que les permitirá una mejor comprensión e interpretación de la información procesada.

El libro esta subdividido en ocho capítulos en los cuales se quiere proporcionar un conocimiento pleno, para llegar a nuestros alumnos con temas como: conceptos generales, interpretación estadística, medidas de posición y tendencia central, medidas de desviación o dispersión, distribuciones de probabilidad, prueba de hipótesis, regresión lineal y análisis factorial.

El autor

.

INDICE

CAPITULO I

CONCEPTOS GENERALES

CAPITULO I

CONCEPTOS GENERALES

BREVE HISTORIA DE LA ESTADISTICA

La colección y registro de datos fue muy común por muchos siglos. En el año 2000 antes de J.C., la población de Judea se estima que fue de 3'800,000. Militares tales como Cesar Augusto, ordenaron censos del mundo, con fines militares y de cobro de impuestos. William el conquistador de Inglaterra ordeno un censo, y los resultados fueron publicados en el Libro "El Juicio Final" en el año 1086. En el desarrollo de la Estadística, el periodo anterior al siglo XVI fue dedicado prioritariamente a la simple colección de datos.

Amplio énfasis en la aplicación y educación estadística comienza a ser común en el siglo XIX y los primeros años del siglo XX. El número de estadísticos aumenta, y Quetelet (1796 - 1874) astrónomo y estadístico Belga promueve el primer congreso de estadística en el año 1853. Algunas de las primeras aplicaciones fueron hechas en astronomía en intentos para fijar las órbitas planetarias. Galton (1822 - 1911) postulo una "ley de regresión universal" en el estudio de la herencia (1888); esta teoría fue sometida a una prueba estadística por Karl Pearson (1857 - 1936) quien ahora es conocido como el "fundador de la ciencia estadística.

La estadística la que nosotros conocemos en estos días, se considera que tiene sus inicios en el trabajo decir Ronald_ a. Fisher (1890 - 1962), un matemático genetista del Rothamsted Experimental Station en Inglaterra..

¿QUÉ ES LA ESTADÍSTICA?

El hombre común y corriente, con frecuencia concibe a la estadística como una masa de números o una colección de datos..

La estadística es la que recolecta, clasifica y presenta los datos, para una toma de decisiones.

La estadística se ocupa de los métodos científicos que se utilizan para recolectar, organizar, resumir, presentar y analizar datos así como para obtener conclusiones válidas y tomar decisiones razonables con base en este análisis. El término estadística también se usa para denotar los datos o los números que se obtienen de esos datos; por ejemplo, los promedios. Así, se habla de estadísticas de empleo, estadísticas de accidentes, etcétera. (Murray R. Spiegel 2009)

DEFINICIONES ESTADISTICAS

Población

Conjunto de elementos, con una o más características bien definida, de los cuales alguno o todos pueden ser observados y representan el universo. Algunos ejemplos:

a) Un conjunto de objetos tales como todos los libros de una biblioteca
b) Un conjunto de valores clasificados, tales como las marcas de cigarrillos

Elementos

Son cada uno de los componentes de la población, en los ejemplos de población los elementos son:

- Los libros
- Las marcas

Características

Representan una propiedad específica cualitativa o cuantitativa de los elementos de una población. Las características tienen un valor específico, una cualidad o una categoría para cada elemento en su conjunto constituye una variable.

Ejemplo si se considera como población los enfermos de un centro de salud se pude considerar varias características, tales como:

- Edad
- Lugar de procedencia
- Peso
- Enfermedad
- Talla
- Ingreso económico mensual

Variable

Dada una característica determinada para una población, la variable es la medida, cualidad o característica concreta de cada uno de los elementos. Así por ejemplo en el caso de los enfermos del Centro de Salud, si se escoge la característica edad, la variable está constituida por el registro de las edades de cada uno de los enfermos. Las variables pueden ser: cualitativa o cuantitativas.

Variables cualitativas, son las características que toman valores que representan cualidades, virtudes, categorías o rangos. Por ejemplo, Estado civil (soltero, casado, viudo, divorciado), Sexo (masculino, femenino), Religión, Estado emocional.

Variable Cuantitativas, son aquellas que únicamente pueden tomar valores numéricos. Así tenemos por ejemplo, las medidas de peso, talla, ingreso económico, edad, etc.

Parámetros

Son valores que representan a una población, con relación a una variable (característica) así tenemos por ejemplo, el promedio, la desviación estándar, la varianza, la proporción, el número total, la suma total.

Estadígrafo

Son valores que representan a una muestra con relación a una variable (característica) así tenemos por ejemplo, el promedio la desviación estándar la varianza, la proporción, el número total, la suma total de la muestra.

Censo

Es un proceso planificado, a través del cual se logra la enumeración de todos los elementos de una población o el estudio de una o más características de los elementos de una población.

Muestra

Es una parte de la población que es obtenida para conocer una o más características de los elementos de una población con la finalidad de realizar inferencias sobre toda la población. Se obtiene una muestra por razones de tiempo y dinero

Encuesta

Es la sistematización en forma de cuestionario del conjunto de características que se requiere estudiar en el censo o muestreo. En la literatura en general se usa en forma indistinta, con el mismo sentido a las palabras muestreo o encuesta; sin embargo es necesario hacer esta precisión.

OBTENCIÓN DE DATOS

En cualquier actividad donde se necesite realizar análisis estadístico, sea esta académica, empresarial o gubernamental; se requiere de datos y Su forma de obtenerla es a través de registros, censos, encuestas y experimentos.

Registros

Las instituciones públicas y privadas por lo general tienen datos que registran sobre su organización, producción, ventas, rendimiento académico, personal, alumnos, trabajadores y otras operaciones de su institución
Por ejemplo en un Centro Educativo tienen datos sobre el número de profesores, estudiantes, administrativos, presupuesto, actividades curriculares, actas académicas entre otras.
Muchas de las veces los directivos de una Institución o Empresa para la toma de decisiones adecuadas, requieren de información externa que se encuentra disponible en boletines informativos y publicaciones de las Instituciones de Gobierno, de Asociaciones Comerciales y otras Organizaciones que recopilan datos.
En el Perú, la institución encargada de centralizar esta información es el Instituto Nacional de Estadística e Informática (INEI), la misma que la sistematiza y oferta al público interesado en forma de compendios anuales

Ejercicio 1

Enumerar todos los registros que se realizan en su Centro Laboral

Censos

Todos los países realizan censos poblacionales con cierta periodicidad, en el caso de nuestro país se ha establecido que por lo menos debe realizarse Wing" sin embargo solo se han

ejecutado los censos de 1940, 1961; 1981, 1993 y 2005; toda esta información es centralizada por el INEI.
Además se programan otros censos más específicos come:

- 111 Censo Económico. 1994
- 111 Censo agropecuario. 1994
- 1 Censo Nacional Universitario 1996

Ejercicio 2

Indicar la población a nivel nacional y de los departamentos de Lima, Apurímac, Arequipa, Cusco, Madre de Dios, Moquegua, Puno y Tacna para los últimos censos; en forma de cuadros y gráficos. Adicionalmente realice un análisis e interpretación de estos datos, precise por lo menos 4 conclusiones.

Muestreo

Muchas veces se requiere información sobre una característica o problema en especial, pero en un determinado tiempo y espacio; puesto que la información censal y la que se tiene en registros no nos da la información requerida; en estos casos se debe recurrir a una encuesta por muestreo a través de un cuestionario que nos permita obtener la información sobre la característica o problema que se quiere conocer.

Experimentos

El término experimento es usado por los Estadísticos pare definir un proceso qua nos permite la recopilación de datos, a través de un diseño que controla uno o más factores y la influencia que ejercen estos sobre una variable de respuesta que es de interés.
Por ejemplo si es un maestro quiere saber cómo influye el use de una nueva metodología de enseñanza, necesita probar que este método es más eficiente que el que está utilizando; por lo tanto deberá realizar un experimento en el cual divida al azar su salón de clases en dos grupos, en uno de ellos seguirá aplicando el método tradicional y en el otro el nuevo método.
Otro ejemplo, sería que el gerente de una empresa desea conocer si un entrenamiento adicional sobre el proceso productivo, produce un aumento en la producción, un experimento para investigar este problema podría consistir en seleccionar al azar a la mitad de los empleados y hacerlos participar en un programa de entrenamiento, dejando que la otra mitad continúe sin entrenamiento adicional y comparando después la productividad de los dos grupos.

PRESENTACION DE DATOS

Los datos obtenidos de las diferentes maneras que se han señalado son mostrados básicamente en forma de cuadros y gráficos

Cuadros

Cuadros de referencia

También se le conoce como Cuadro Fuente, porque están diseñados pare servir como depósito de información estadística. Los Cuadros de referencia contienen datos que permiten el análisis de relaciones entre variables; estos cuadros se encuentran en los aprendices de los estudios analíticos.

En el Cuadro N° 01 se aprecian datos que están contenidos en un cuadro de referencia, donde únicamente se tienen datos, sin realizar ningún análisis ni mucho menos obtener conclusiones.

Cuadro N° 01
Peso en Kilogramos de 100 niños de 14 años de la Gran Unidad Escolar Mariano Melgar Arequipa – Año 2016

67.9	58.6	65.6	57.7	64.6	67.0	58.2	60.2	60.4	61.51
60.7	61.9	63.3	63.6	60.4	61.0	65.7	64.8	52.9	63.2
64.4	60.6	59.6	65.8	60.5	60.7	66.5	65.7	63.3	63.6'
56.4	65.6	63.0	62.8	58.61	62.3	60.9	65.8	63.7	61.3
60.4	61.7	63.6	62.6	63.61	60.3	65.81	60.2	66.6	63.6
64.8	65.6	65.9	63.0	60.4	64.8	62.6	60.9	57.2	61.2
61.6	58.8	67.2	63.71	62.6	63.6	65.8	62.9	60.8	63.2
62.7	60.9	62.6	56.8	60.7	65.4	64.8	61.4	65.2	61.0
65.9	65.6	65.9	63.7	61.2	61.9	65.61	67.0	63.0	62.9
63.5	60.9	63.8	63.6	62.9	62.7	65.11	58.7	58.0	61.8!

Fuente datos obtenidos a través de un muestreo del autor

Cuadros Analíticos

Elaboran Cuadros Analíticos cuando deben estudiarse las relaciones mutual entre factores o variables. La diferencia entre Cuadros de referencia y Cuadros analíticos esta fundamentalmente en el uso que se le da y no en su construcción.

CUADRO N° 2

VENTAS DE AUTOMOVILES EN LA CIUDAD DE AREQUIPA				
MES	2016	2017	INCREMENTO	INDICE VENTAS%
ENERO	500	700	200	140%
FEBRERO	450	400	-50	89%
MARZO	600	500	-100	83%
ABRIL	200	400	200	200%
MAYO	350	395	45	113%
JUNIO	500	450	-50	90%
JULIO	700	600	-100	86%
AGOSTO	550	500	-50	91%
SETIEMBRE	280	350	70	125%
OCTUBRE	350	450	100	129%
NOVIEMBRE	290	500	210	172%
DICIEMBRE	500	489	-11	98%

Partes Estructurales de un Cuadro

Cada autor tiene su propia forma de presentar cuadros, sin embargo la metodología estadística nos ofrece un formato básico de las partes estructurales principales que nos permite estandarizar la presentación de cuadros, sin embargo cada uno de los interesados puede realizar variantes de acuerdo a sus necesidades y criterios de estética de presentación

Número del Cuadro

Para facilitar la ubicación de un cuadro en un reporte, artículo, libro o cualquier otro documento es necesario numerarlo de acuerdo al orden de aparición. Cuando aparece un solo cuadro en el reporte o presentación, es posible obviar el número.

Título

El título de un cuadro estadístico debe expresar con claridad el contenido respondiendo fundamentalmente a las siguiente preguntas ¿Qué contiene?, ¿Dónde se obtuvieron los datos? Y ¿cuándo se obtuvieron los datos?, también se suele considerar la pregunta ¿Cómo están clasificados los datos?
Al redactar el título se debe evitar los extremos de: una redacción demasiada elaborada y extensa o la de una redacción breve que no proporciona la información suficiente sobre el cuadro. El título debe permitirnos saber a primera vista el contenido del cuadro

Nota en el encabezado del Cuadro

Las notas en el encabezado del cuadro generalmente se ponen debajo del título y entre paréntesis; proporcionan información complementaria al cuadro o a las unidades de medida de los datos

Filas

Las filas están ubicadas en la primera columna del cuadro y contiene el encabezado de la Clasificación A y los rótulos de cada una de las filas. El encabezado de las filas debe usarse para denominar la característica del factor A y cada fila debe tener un rotulo que sea suficientemente claro. Por lo general en las filas esta la variable cualitativa que tiene varias categorías.

Encabezado Principal

El cajón de encabezados contiene el título de la segunda variable o Clasificador B y se supone a todas las demás columnas del cuadro.

Encabezados

Todas las columnas a partir de la 2da tienen un encabezado que es el que corresponde al encabezado principal de las columnas, que de ser necesario puede omitirse, este encabezado corresponde al factor B.
A su vez, cada una de las columnas tiene un rotulo, donde debe estar el total de la clasificación B, el rótulo de la clase B_1, el rótulo de la clase B_2, y así sucesivamente hasta completar todos los niveles del Factor B

Cuerpo

En el cuerpo o también denominado campo del cuadro están las celdas, donde se registran los datos estadísticos. Una celda es la intersección de una fila y una columna, el dato corresponde a los factores considerados en la intersección de las clasificaciones A y B.
Es importante que cada celda se registre un dato, si el dato es cero, este debe ser consignado y si no hay información debe indicarse ND (no disponible).
Cuando una celda se deja en blanco, el lector no sabe si este dato ha sido omitido o no está disponible.

Fuente

Debe consignarse al pie del cuadro la fuente o origen de la información, esto con la finalidad de que si se quiere recurrir a los datos originales, estos puedan ser ubicados con facilidad: y por otro lado para conocer de la confiabilidad de los datos.

Cuadro de Distribución de Frecuencias

Un cuadro donde se especifican las clases y las frecuencias ya sea en valores absolutos (cantidades) o relativos (porcentajes) se le conoce como Cuadro de Distribución de Frecuencias.

Cuadro N° 03
Número de Ovinos de las Familias de una Comunidad del Distrito de Santa Rosa Caylloma – Año 2014

Número de Ovinos (Clases) X_1	Número de Familias (Frecuencia) n_1
0 – 1	22
2 – 3	225
4 – 5	432
6 – 7	231
8 – 9	60
TOTAL	1000

Fuente: Ministerio de Agricultura

Los datos consistente en un gran número de observaciones, provenientes de una muestra o una población; pueden ser resumidos en categorías o clases. Por ejemplo el peso de los estudiantes u otras medidas, pueden ser asignados a clases con diferentes intervalos.

Los datos resumidos en un arreglo tabular que muestren el número de individuos en cada clase, es denominado cuadro de distribución de frecuencias o tablas de frecuencias, siendo estas estructuras la base para una presentación gráfica o indicadores de tendencias.

CAPÍTULO II
INTERPRETACIÓN ESTADÍSTICA
TABLAS DE FRECUENCIA

CAPÍTULO II

INTERPRETACIÓN ESTADÍSTICA

TABLAS DE FRECUENCIA

Son una ordenación de los datos estadísticos, asignando a cada dato su frecuencia correspondiente.

Pueden ser para datos cualitativos o cuantitativos

Ejemplos:

a) Un docente tiene las notas del curso de diseño de reactores de 30 alumnos.
b) Marcas de gaseosa que consume la población arequipeña.

CONSTRUCCIÓN DE LA TABLA DE FRECUENCIAS

1) <u>Rango datos</u>

$$\mathbf{R = X_{máximo} - X_{mínimo.}}$$

$X_{MAX} = 67.9$
$X_{MIN} = 56.4$

$$R = 67.9 - 56.4$$

$$R = 11.5$$

Nro de filas del Cuadro de distribución.

El número de categorías, puede también determinarse, aplicando la regla de STURGES pero ésta no es de obligatoria utilización. La fórmula de Sturges es la siguiente:

$$\mathbf{K = 1+3.3\log(n)}$$

Donde:

K = Número de clases
n = Tamaño de muestra = 100

$$K = 1 + 3.3 \log(100)$$

$$K = 1 + 3.3 * 2$$

$$K \cong 7.6 \cong 8 \text{ clases}$$

Calculamos el tamaño de Intervalo

$$I = \frac{R}{k}$$

R = 11.5
K = 8

$$I = \frac{11.5}{8}$$

$$I = 1.4375$$

$$I \cong 1,5$$

Registramos las categorías

Calculamos primero la amplitud de la tabla

4.1 Amplitud de la Tabla (A)

Es el producto entre número de clases y el intervalo de clase.

$$A = K \times I$$

Donde:
K = 8
I = 1.5

$$A = 8 \times 1.5$$

$$A = 12$$

Determinación del límite inferior de la primera clase (L)

$$Li = X_{MIN} - \frac{A - R}{2}$$

Donde:
$L_{INFERIOR}$ = 56.4
R = 11.5
A = 12

$$Li = 56.4 - \frac{12 - 11.5}{2}$$

$$Li = 56.15 \approx 56.2$$

Cuadro N° 04

	a<X≤ b	
Clase 1	56.2	57.7
Clase 2	57.7	59.2
Clase 3	59.2	60.7
Clase 4	60.7	62.2
Clase 5	62.2	63.7
Clase 6	63.7	65.2
Clase 7	65.2	66.7
Clase 8	66.7	68.2

Computar los datos de cada clase

Con la computación de los datos de cada clase tenemos la frecuencia de los mismos y lo llamaremos "f",

Cumplidos los anteriores pasos, la tabla de frecuencia de nuestro ejemplo.

Cuadro N° 05

	a<X≤ b		f_i	X	F	h %	H %
Clase 1	56.2	57.7	4	56.95	4	4	4
Clase 2	57.7	59.2	6	58.45	10	6	10
Clase 3	59.2	60.7	13	59.95	23	13	23
Clase 4	60.7	62.2	17	61.45	40	17	40
Clase 5	62.2	63.7	30	62.95	70	30	70
Clase 6	63.7	65.2	9	64.45	79	9	79
Clase 7	65.2	66.7	17	65.95	96	17	96
Clase 8	66.7	68.2	4	67.45	100	4	100
	TOTAL		**100**				

Dónde:

1. Marca de clase **"X",** se calcula con la siguiente formula:

$$X = \frac{L_S + L_I}{2}$$

Dónde:

L_S = Límite superior del intervalo
L_I = Límite inferior del intervalo

Ejemplo

Para la Clase 1

$$X = \frac{57.7 + 56.2}{2} = 56.95$$

2. La **"F"** frecuencia absoluta acumulada, la cual se obtiene sumando el primer valor de la frecuencia con el segundo y así sucesivamente al final se debe obtener el número total de datos (100).

3. La "h %", es la frecuencia relativa

Se calcula con la siguiente fórmula:

$$h_i = \frac{f_i}{N} x100$$

Dónde:

f_i = es la frecuencia
N = es el total de datos

4. La "H %", es la frecuencia relativa acumulada, la cual se obtiene sumando el primer valor de la frecuencia con el segundo y así sucesivamente al final se debe obtener el 100%.

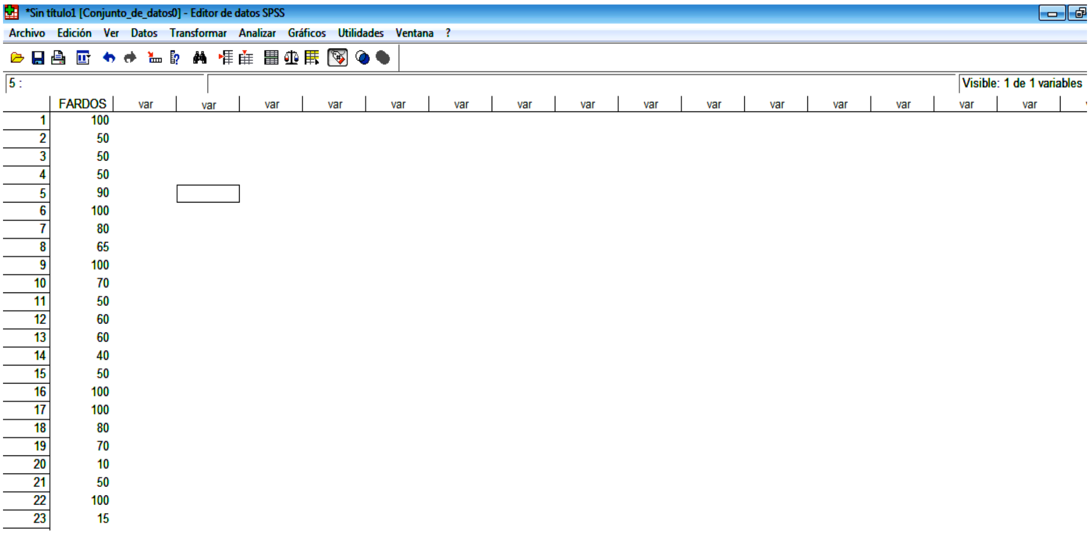
APLICACIÓN DEL SOFTWARE ESTADÍSTICO SPSS

Ejemplo

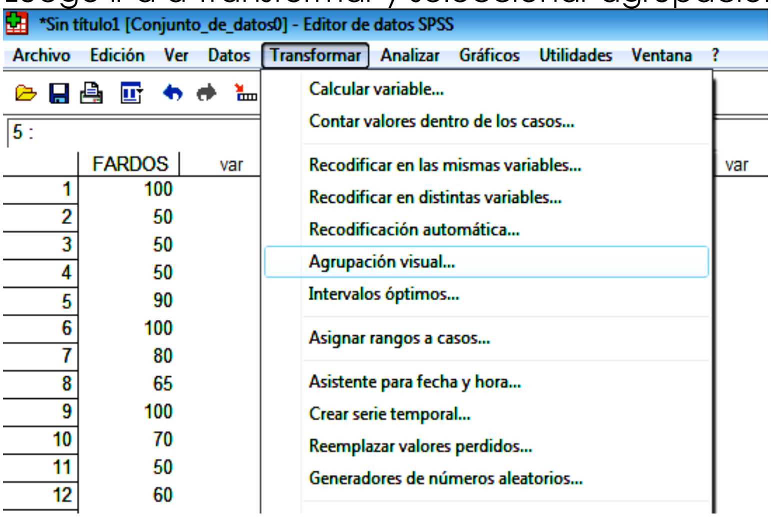
Dados los siguientes datos, que corresponde al peso de fardos de lana de la Empresa Michell & Cia., construir un Cuadro de Distribución de Frecuencias.

100	60	15	30	80	90	110	40	80	70	100
50	60	100	70	80	60	110	15	100	65	80
50	40	40	120	100	80	80	80	30	100	50
50	50	100	120	80	40	60	65	100	100	60
90	100	80	60	100	60	40	100	100	100	100
100	100	80	120	80	100	80	100	45	100	40
80	80	100	100	100	70	80	80	40	100	55
65	70	100	60	50	120	70	50	80	80	70
100	10	100	50	35	100	35	80	75	80	100
70	50	70	110	80	50	100	30	40	100	60
50	100	70	110	110	50	100	100	30	100	100

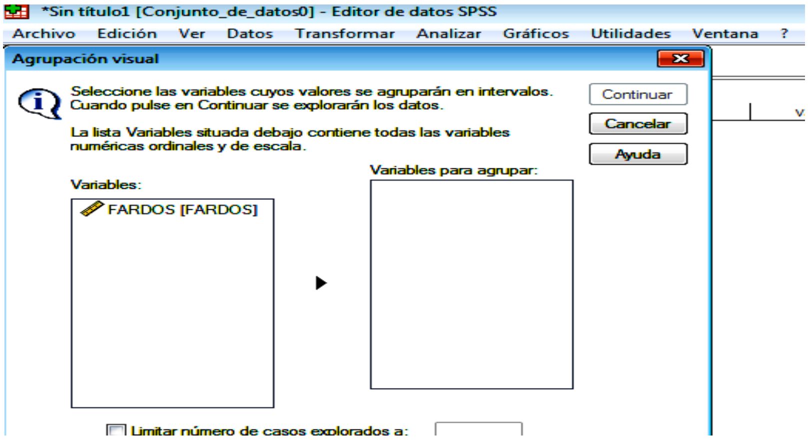
Abrir programa SPSS y colocar los datos



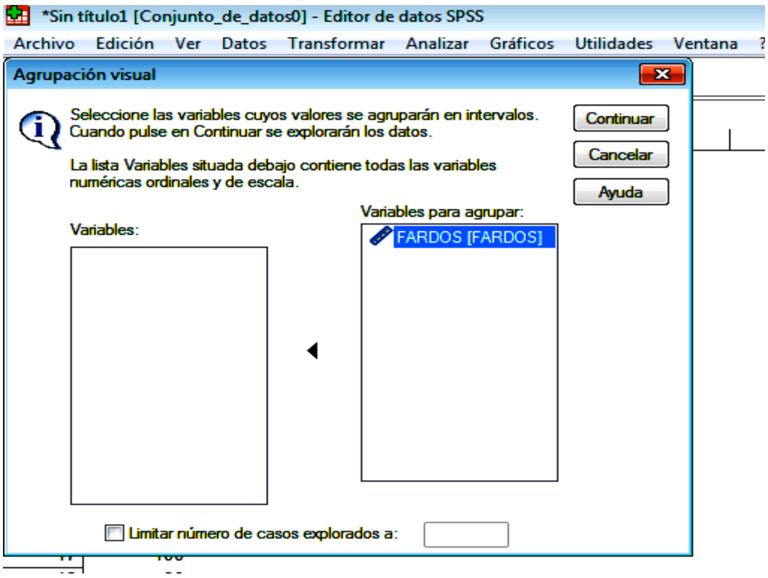
Luego ir a a Transformar y seleccionar agrupacion visual

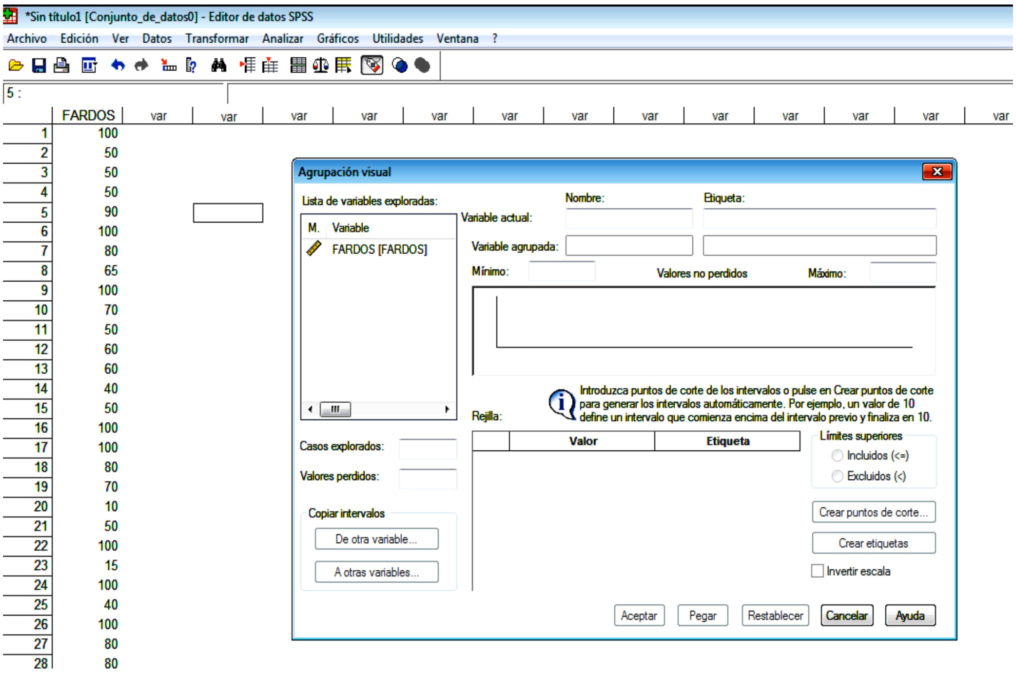


Aparecera

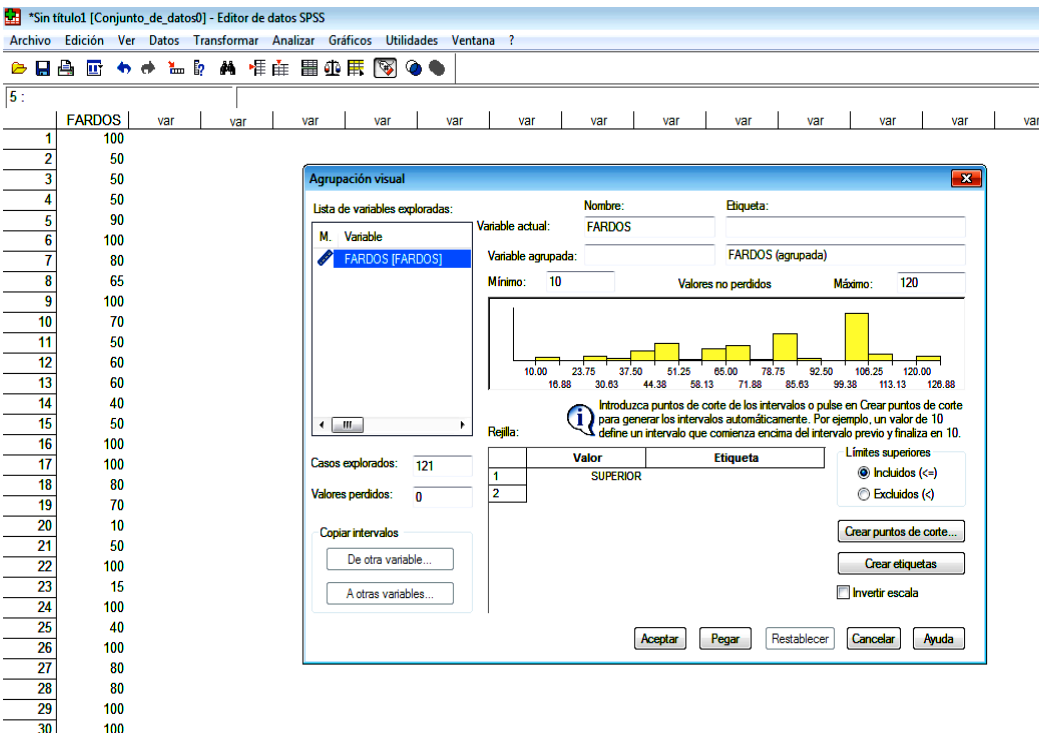


Seleccionar la variable fardos y pasar a variables para agrupar y continuar

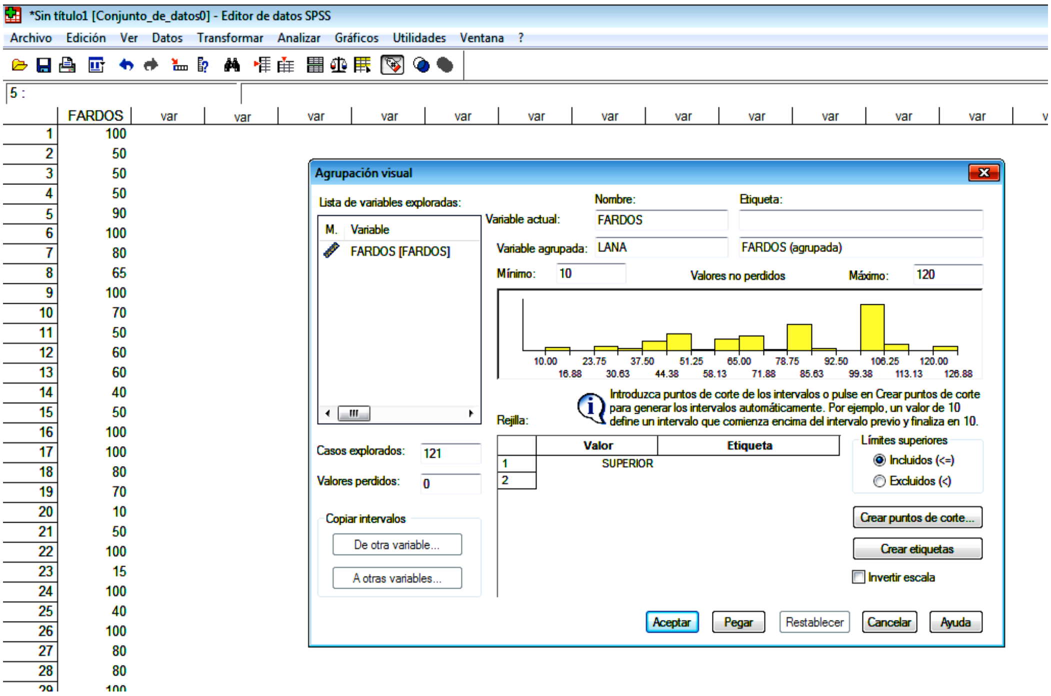




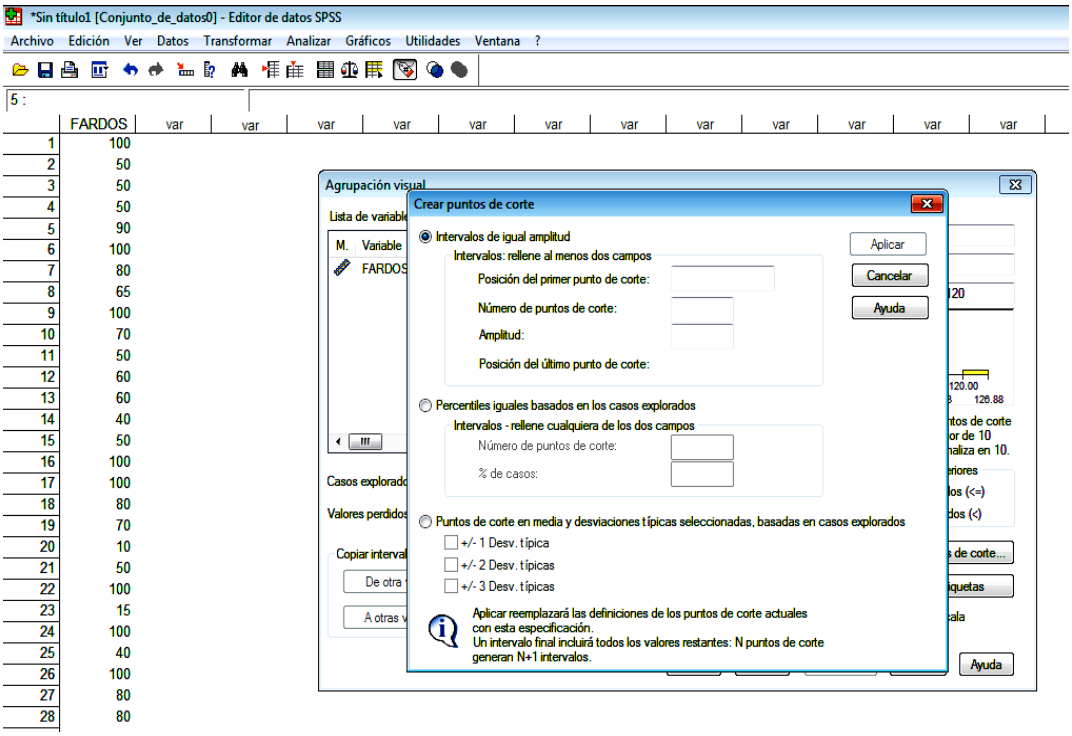
Dar doble clic en la variable fardos



Poner nombre en la variable agrupada, le pondremos LANA

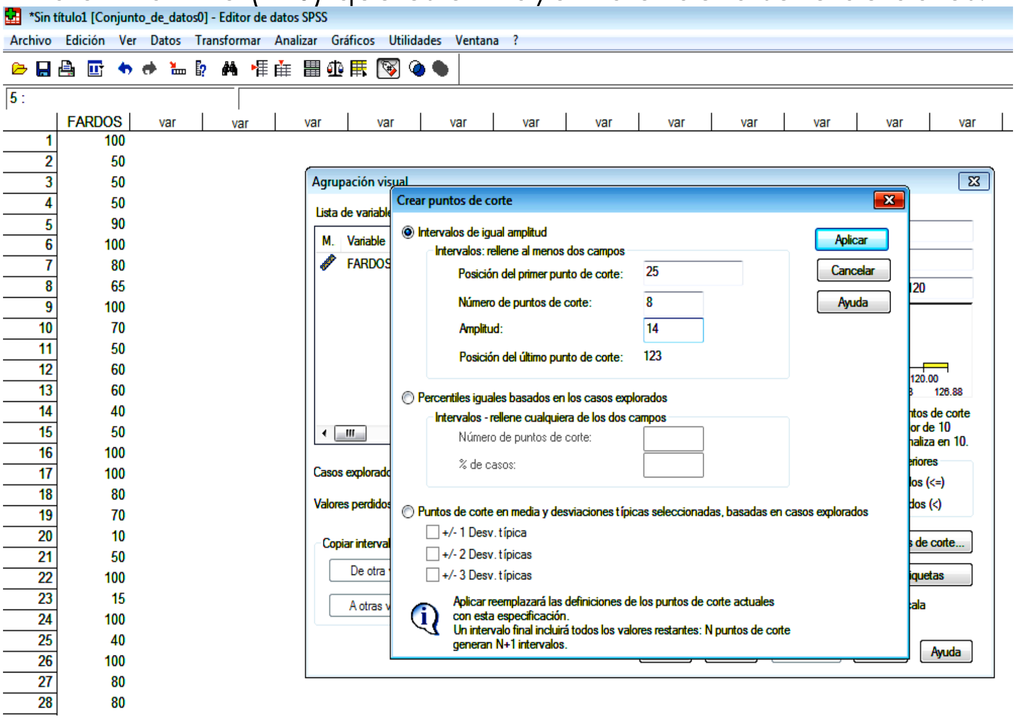


Dar clic en crear puntos de corte

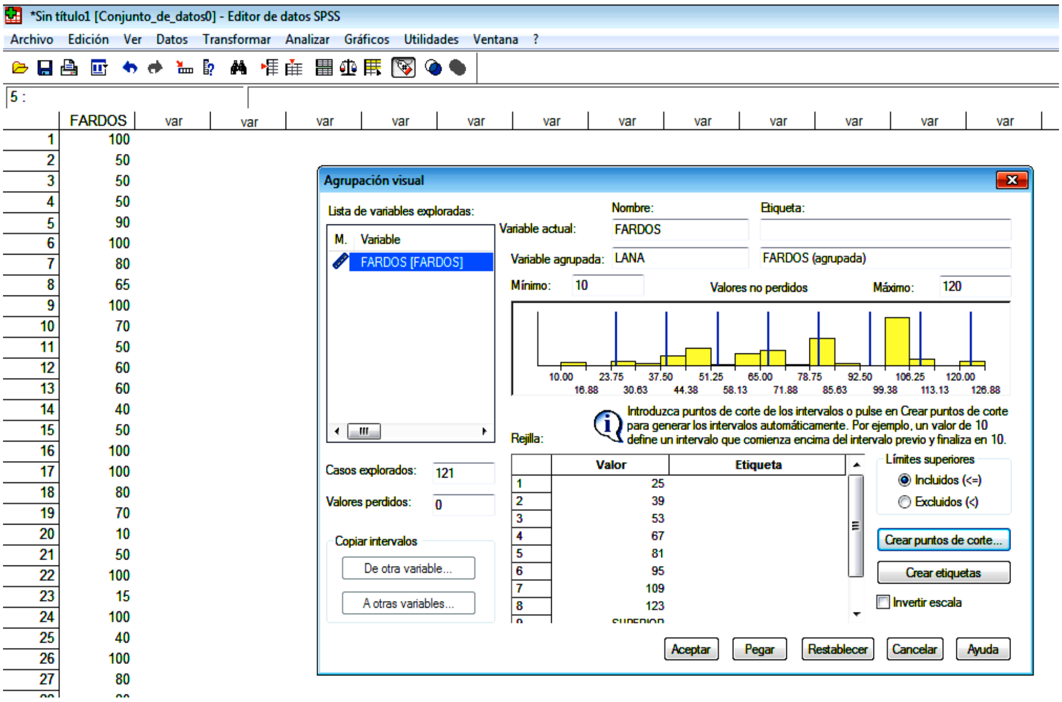


Llenar :

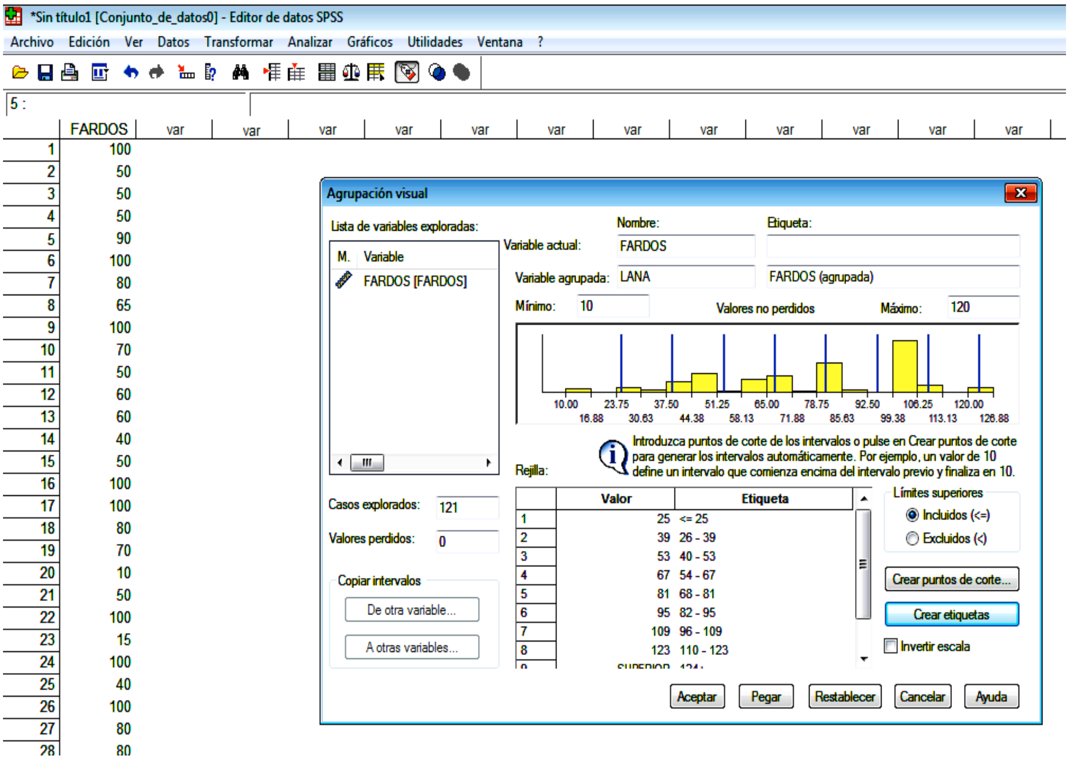
1. Posición del primer punto de corte: valor mínimo 25
2. Número de puntos de corte: 8 (Apoyados por la fórmula de Sturges)
3. Anchura de intervalo: 14
4. Se debe de jugar con estas variables hasta obtener o pasar el valor máximo (120) que es el mayor valor de la serie de datos.



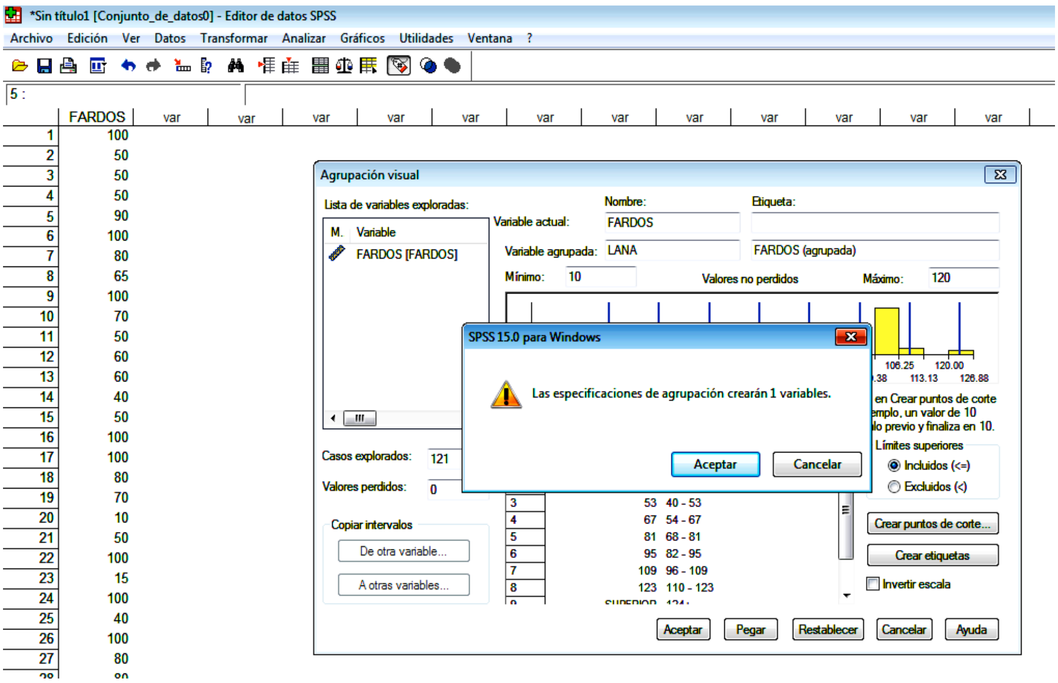
Dar clic en aplicar



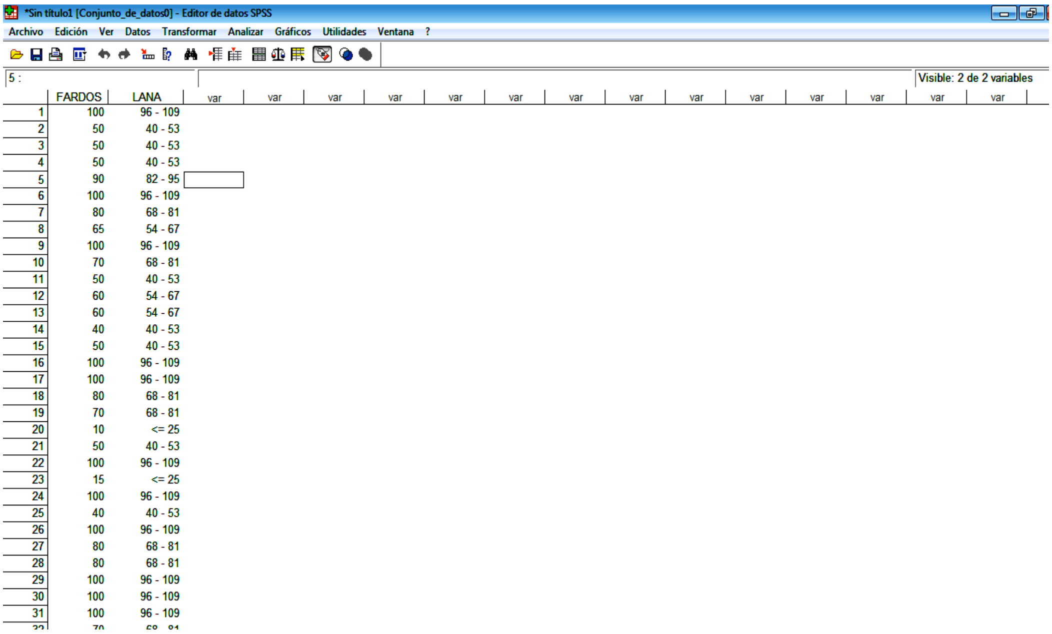
Luego clic en crear etiqueta



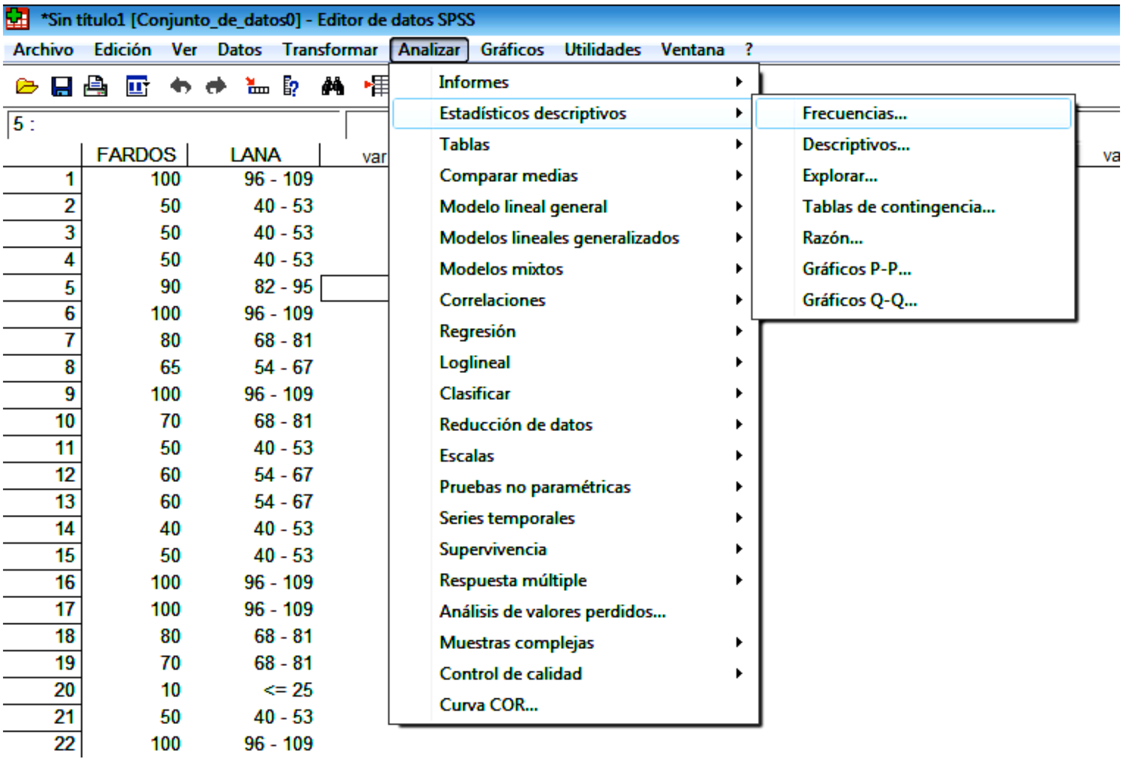
Clic en aceptar

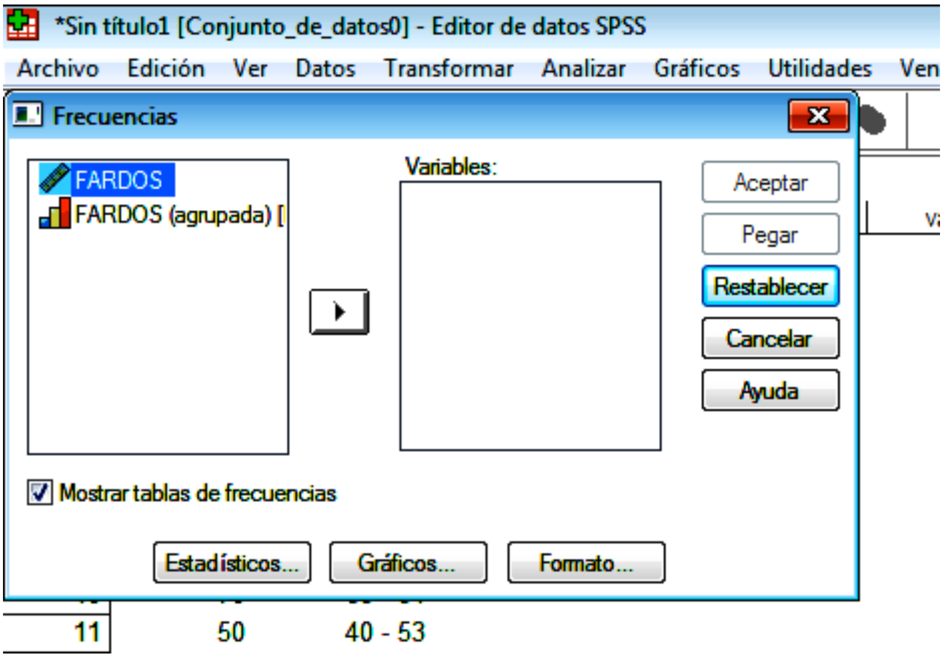


Luego aceptar

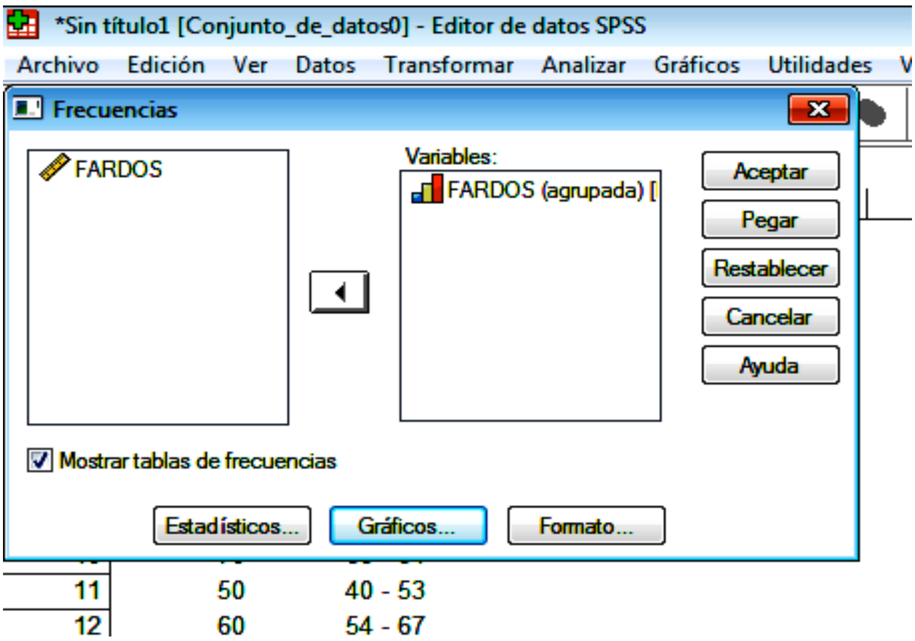


Luego ir a analizar – Estadisticos descriptivos – Frecuencias





Pasar la variable FARDOS (Agrupada) ir a graficos y escoger histograma con curva normal luego continuar.



Luego aceptar

Estadísticos

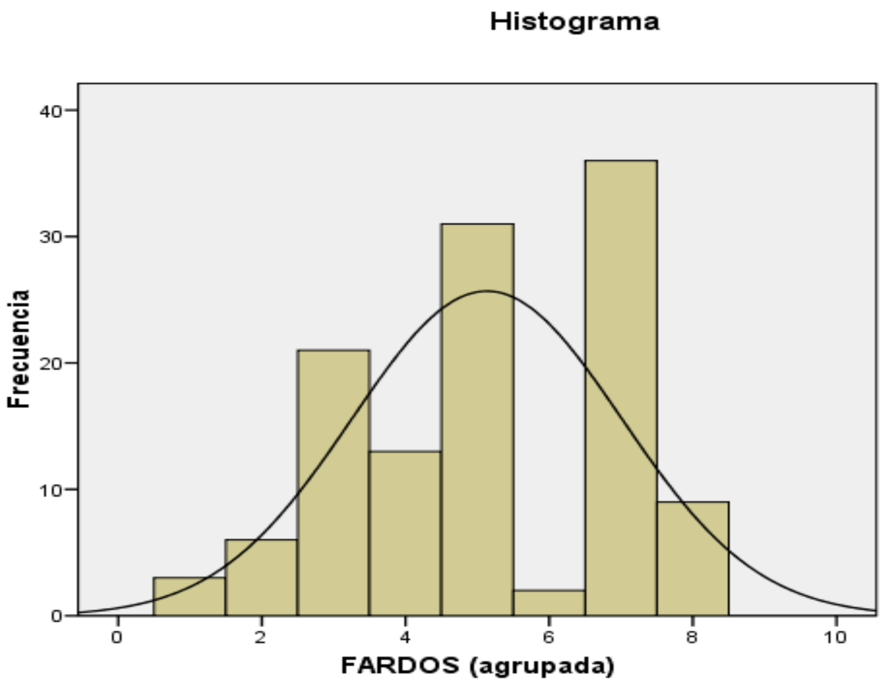
FARDOS (agrupada)

N	Válidos	121
	Perdidos	0

FARDOS (agrupada)

-

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	<= 25	3	2.5	2.5	2.5
	26 - 39	6	5.0	5.0	7.4
	40 - 53	21	17.4	17.4	24.8
	54 - 67	13	10.7	10.7	35.5
	68 - 81	31	25.6	25.6	61.2
	82 - 95	2	1.7	1.7	62.8
	96 - 109	36	29.8	29.8	92.6
	110 - 123	9	7.4	7.4	100.0
	Total	121	100.0	100.0	

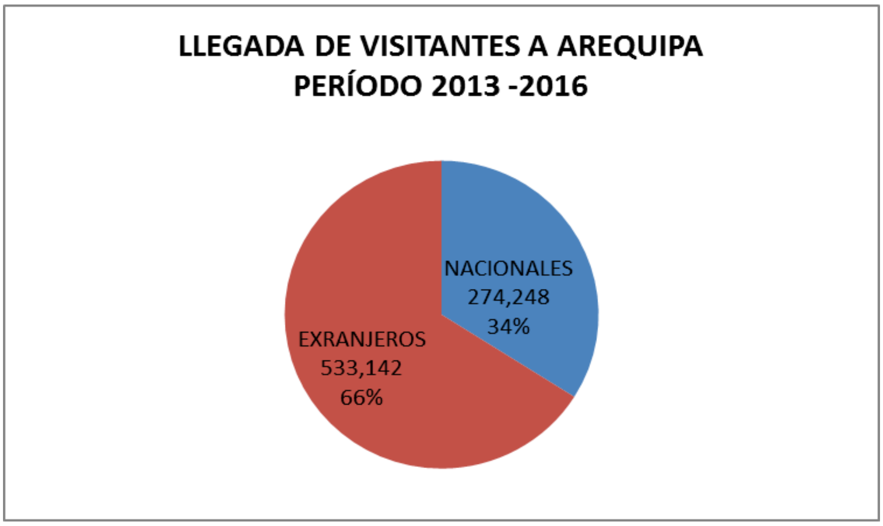


GRÁFICOS

GRÁFICOS CIRCULARES

Estos gráficos sirven para comparar las partes con respecto al Todo a resaltar las diferencias entre las partes. Esta representación es a través de una circunferencia o un pastel que representa el todo y que se la divide en sectores circulares o fracciones del pastel que están en proporción a la frecuencia de la partes.

Ejemplo



PICTOGRAMAS

Sirven para realizar representaciones de características específicas a través de figuras que se aproximan a la descripción de una característica. Por ejemplo si se quiere especificar el sexo para una población la representación grafica sería de la manera siguiente:

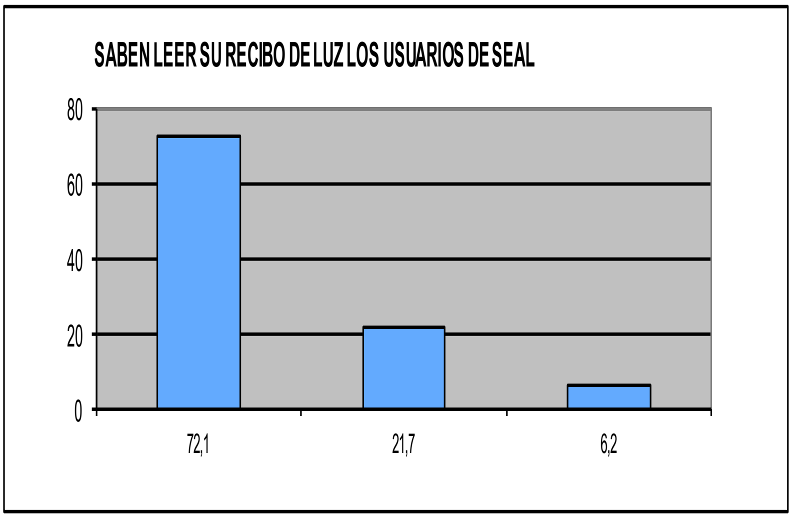
POBLACIÓN DEL DEPARTAMENTO DE AREQUIPA

Número
635,830
Porcentaje
54%

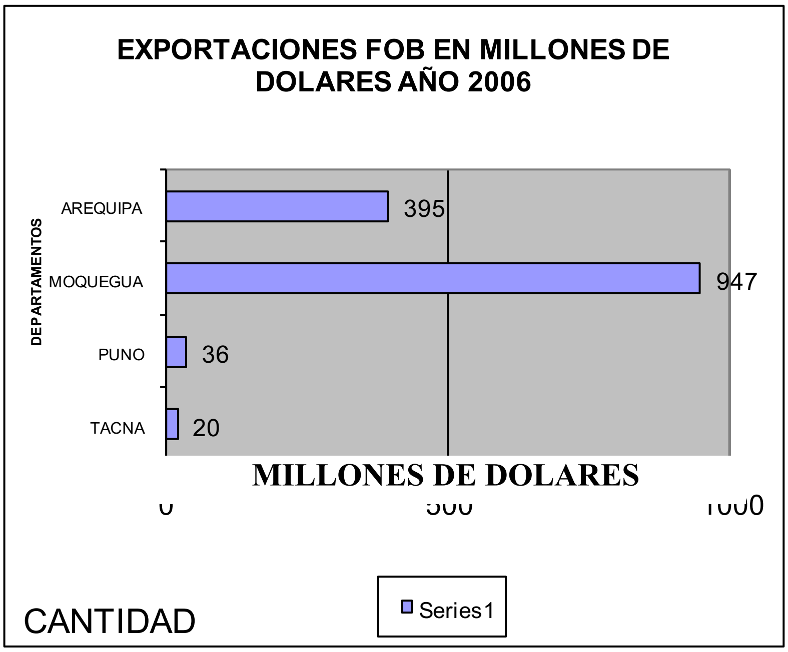
Número
544,019
Porcentaje
46%

GRÁFICAS DE BARRAS

Esta es una de las gráficas más usadas, son graficas que se dibujan sobre plano cartesiano las barras que pueden estar en el eje de las abscisas o en el de las ordenadas se las usa normalmente para graficar cuadros de distribución de frecuencias.

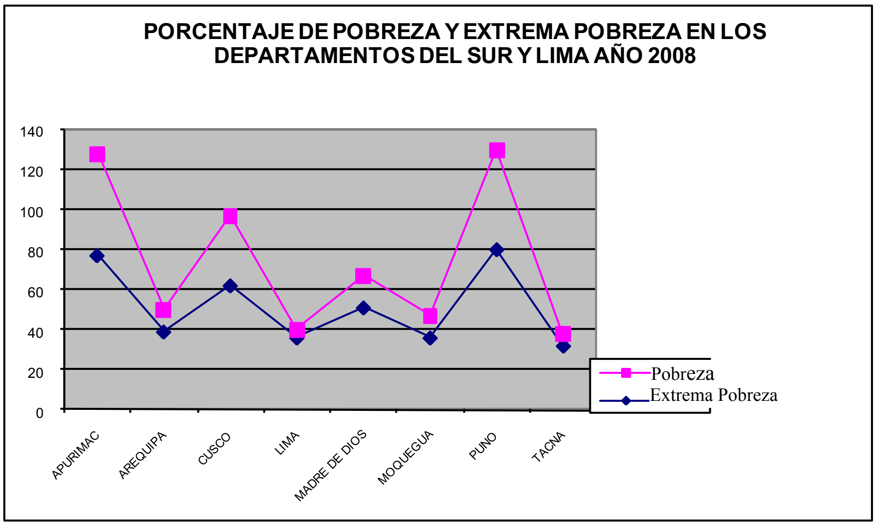


SI NO NO SABE/ NO RESPONDE



GRÁFICAS LINEALES

Es otra de las gráficas que se usa en plano cartesiano donde por lo general la clase de categoría se ubica en el eje de las abscisas y la frecuencia en el eje de las ordenadas. Observar los siguientes ejemplos.



EJERCICIOS

1. Dibuje dos gráficas circulares dos pictogramas, dos gráficas de barras y dos gráficas lineales con datos de su Centro de Trabajo.

2. Dados los siguientes datos, que corresponde a la nota final del Curso de Métodos Estadísticos de los estudiantes de la Maestría de Educación de la UANCV - Arequipa, construir un Cuadro de Distribución de Frecuencias.

20	14	09	14	16	11	18	08	10	10
17	18	12	16	20	12	19	08	11	13
13	19	12	14	20	13	20	12	13	13
17	16	13	15	17	14	15	13	13	14
19	19	13	16	17	14	13	14	15	15

3. Construir una gráfica lineal de la temperatura con respecto al tiempo con los siguientes datos

Tiempo	5	10	15	20	25	30	35	40	45	50	55	60	65
Temp.°C	18	23	27	28	30	32	33	34	35.5	37	38	49	40

4. Los tipos de sangre en un distrito de la región Arequipa están dados por:

T. Sangre	O^+	A	B	AB
Habitantes	50,000	23,000	15,000	12,000

   Construir un Gráfico Circular

5. Construir un diagrama de barras con la siguiente información:

EDAD	ADMINISTRACIÓN V SEMESTRE UNIVERSIDAD "A"	ADMINISTRACIÓN V SEMESTRE UNIVERSIDAD "B"
18	12	17
19	16	11
20	14	9
21	8	13

CAPITULO III
MEDIDAS DE POSICIÓN

Y TENDENCIA CENTRAL

MEDIDAS DE POSICIÓN

CAPITULO III

MEDIDAS DE POSICIÓN Y TENDENCIA CENTRAL

MEDIDAS DE POSICIÓN

LOS CUANTILES

Para calcularlos se procede a ordenar los datos de menos a mayor. Son de tres clases: Cuartiles (Q) , Deciles (D) y Percentil (P). Se calculan con la siguiente formula:

$$Posición = \frac{\%}{100} n$$

Dónde:

N = Número de datos
% = equivalente a la posición a ubicar

a) CUARTILES

Divide el conjunto de datos en cuatro grupos iguales.

$$\begin{array}{cccc} \mathbf{Q_1} & \mathbf{Q_2} & \mathbf{Q_3} & \mathbf{Q_4} \\ \downarrow & \downarrow & \downarrow & \downarrow \\ \mathbf{25\%} & \mathbf{50\%} & \mathbf{75\%} & \mathbf{100\%} \end{array}$$

b) DECILES

Divide el conjunto de datos en diez grupos iguales.

$$\begin{array}{ccccc} \mathbf{D_1} & \mathbf{D_2} & \mathbf{D_3} & \ldots\ldots\ldots\ldots & \mathbf{D_{10}} \\ \downarrow & \downarrow & \downarrow & & \downarrow \\ \mathbf{10\%} & \mathbf{20\%} & \mathbf{30\%} & & \mathbf{100\%} \end{array}$$

c) PERCENTILES

Divide el conjunto de datos en cien grupos iguales.

$$\begin{array}{ccccc} \mathbf{P_1} & \mathbf{P_2} & \mathbf{P_3} & \ldots\ldots\ldots\ldots & \mathbf{P_{100}} \\ \downarrow & \downarrow & \downarrow & & \downarrow \\ \mathbf{10\%} & \mathbf{20\%} & \mathbf{30\%} & & \mathbf{100\%} \end{array}$$

CALCULO DE LOS CUANTILES EN DATOS NO AGRUPADOS

Ejemplo

Con los siguientes datos: 22, 49, 4, 16, 49, 90, 91, 15, 17, 9; calcular el C_1, D_3 y P_{20}

Primero ordenamos de menor a mayor

4	9	15	16	17	22	49	49	90	91
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
1	2	3	4	5	6	7	8	9	10

Luego aplicamos la formula

a) Q_1 $$Posición = \frac{30}{100}(10) = 15$$

Q_1 = 15

b) D_3 $$Posición = \frac{25}{100}(10) = 2.5$$

D_1 = 15

c) P_{20} $$Posición = \frac{20}{100}(10) = 2$$

P_{20} = 9

CALCULO DE LOS CUANTILES EN DATOS AGRUPADOS

Ejemplo

De la siguiente tabla de frecuencias calcular Q_1, D_6 y P_{40}

Salario Mensual	Nro Trabajadores	Frecuencia Acumulada	UBICACIÓN DEMEDIDA
[500 - 1000[	5	5	
1000 - 1500	8	13	
1500 - 2000	10	23	(Q_1)
2000 - 2500	17	40	
2500 - 3000	13	53	(D_6)
3000 - 3500	11	64	
3500 - 4000	5	69	(P_{90})
4000 - 4500	3	72	
TOTAL	72		

Pasos:

1. Efectuar la Frecuencia Acumulada
2. Para saber en qué fila está el Cuartil, Decil y Percentil se hace uso de la Frecuencia Acumulada y de la siguiente formula:

$$\frac{\beta}{\lambda}(N)$$

Dónde:

β = Que número de Cuartil, Decil o Percentil se quiere obtener.
N = Cantidad de Datos
λ = Tomara el valor de:
4 si es CUARTIL
10 si es DECIL
100 si es PERCENTIL

a) Entonces para el CUARTIL 1 (Q_1). Sería:

$$\frac{1}{4}(72) = 18$$

Este valor estaría en el intervalo de clase de 1500 - 2000 y FA= 23 que contiene al valor de 18

Y para ubicar el valor del Cuartil 1 (**Q1**), se hará uso de la siguiente Formula:

$$Q_1 = L_i + \left[\frac{\frac{\beta(N)}{4} - FA}{f} \right] i$$

En que:

Li = dato menor del Intervalo del Cuartil
= 1500

$$\frac{\beta}{\lambda}(N) = \frac{1(72)}{4} = 18$$

FA = Frecuencia Acumulada que antecede al Cuartil = 13
f = frecuencia que contiene al Cuartil = 10
i = Valor del Intervalo de clase que contiene el Cuartil
= 2000 – 1500 = 500

Reemplazando en la formula, tenemos

$$Q_1 = 1500 + \left[\frac{18 - 13}{10} \right] 500 = 1{,}750$$

CONCLUSIÓN

- El 25 % de trabajadores gana menos de S/. 1,750
- El 75 % de trabajadores gana más de S/. 1,750

b) Entonces para el DECIL 6 (D6). Sería:

$$\frac{6}{10}(72) = 43.2$$

Este valor estaría en el intervalo de clase de 2500 - 3000 y FA= 53 que contiene al valor de 43.2
Y para ubicar el valor del Decil 6 (**D6**), se hará uso de la siguiente Formula:

$$D_1 = L_i + \left[\frac{\frac{\beta(N)}{10} - FA}{f} \right] i$$

Dónde:

L_i = Límite Inferior del Intervalo de clase que contiene al Decil
= 2500

$$\frac{\beta}{\lambda}(N) = \frac{6(72)}{10} = 43.2$$

FA = Frecuencia Acumulada que antecede al Decil = 40
f = frecuencia que contiene al Cuartil = 13
i = Valor del Intervalo de clase que contiene el Cuartil
= 3000 – 2500 = 500

Reemplazando en la formula, tenemos

$$D_6 = 2500 + \left[\frac{43.2 - 40}{13}\right]500 = 2,623.01$$

CONCLUSIÓN

- El 60 % de trabajadores gana menos de S/. 2,623.01
- El 40 % de trabajadores gana más de S/. 2,623.01

c) Entonces para el PERCENTIL 90 (P_{90}), Sería:

$$\frac{90}{100}(72) = 64.8$$

Este valor estaría en el intervalo de clase de 3500 - 4000 y FA= 69 que contiene al valor de 64.8
Y para ubicar el valor del Percentil 90 (**P_{90}**), se hará uso de la siguiente Formula:

$$P_{90} = L_i + \left[\frac{\frac{\beta(N)}{100} - FA}{f}\right]i$$

Dónde:

L_i = Límite Inferior del Intervalo de clase que contiene al Cuartil = 3500

$$\frac{\beta}{\lambda}(N) = \frac{90(72)}{100} = 64.8$$

FA = Frecuencia Acumulada que antecede al Percentil = 64
f = frecuencia que contiene al Cuartil = 5
i = Valor del Intervalo de clase que contiene el Cuartil
= 4000 – 3500 = 500
Reemplazando en la formula, tenemos

$$P_{90} = 3500 + \left[\frac{64.8 - 64}{5}\right]500 = 3{,}580$$

CONCLUSIÓN

- El 90 % de trabajadores gana menos de S/. 3,580
- El 10 % de trabajadores gana más de S/. 3,580

MEDIDAS DE TENDENCIA CENTRAL

Ente las Medidas de tendencia central tenemos: la Media, la Mediana y la Moda

3.2.1 MEDIA

Es la más importante y útil de un mayor manejo

3.1.1.1 Notación

μ = Promedio de población
$\overline{X}$ = promedio de muestra

3.1.1.2 DEFINICIÓN

Es la cantidad total de la variable distribuida a partes iguales entre cada observación.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{N} = \frac{Sumatoria\ \ de\ \ valores\ \ de\ \ observaciones}{Sumatoria\ \ de\ \ observaciones\ \ de\ \ la\ \ muestra}$$

Ejemplo

Determinar el promedio de 4, 5 8, 10, 6

$$\overline{X} = \frac{\sum_i^n X_i}{N} = \frac{4+5+8+10+6}{5} \qquad \overline{X} = 6.6$$

PROPIEDADES

a) La suma de las desviaciones de cada una de las observaciones con respecto al promedio es igual a cero

$$\sum_i^n \left(X_i \quad - \overline{X} \right) = 0$$

b) El promedio depende de cada una de las observaciones en la proporción de X_i / n

c) Si a cada una de las observaciones se le suma una constante, el nuevo promedio es igual a la promedio de las observaciones más la constante.

$$\overline{X_c} = \frac{\sum_i^n \ (X_i + c)}{N} = \overline{X} + C$$

d) Si cada una de las observaciones se le multiplica una constante el nuevo promedio es igual al promedio de las observaciones multiplicando por la constante.

$$X_c = \frac{\sum_i^n \ cx}{n} = cX$$

Ejemplo

En una oficina trabajan cinco mecanógrafos que tiene una velocidad de mecanografiado de 40, 65, 53, 48 y 56 palabras por minuto. Calcular el promedio de velocidad de mecanografiado de las cinco secretarias.

$$\overline{X} = \frac{\sum_i^n X_i}{n} = \frac{40+65+53+48+56}{5} \qquad \overline{X} = \frac{262}{5} \qquad \overline{X} = 52.4$$

Ejemplo

En una muestra se toma el número de partos diarios que se realiza en un centro de salud entre los años 2013 y 2015. Calcular el promedio aritmético simple.

5	4	6	7	8	7	5	7	8	7
4	4	5	6	9	7	7	6	9	7
7	7	4	5	7	7	8	7	6	6
3	5	9	4	9	8	7	8	7	5
4	6	10	3	5	6	7	9	8	5
9	11	5	6	6	7	6	6	9	4
5	6	7	7	7	7	8	7	10	3

$$X = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{5+4+...+3}{70}$$

$$X = \frac{458}{70}$$

$$X = 6.5$$

MEDIA ARITMÉTICA EN DATOS AGRUPADOS

Definición

Es el valor obtenido del cociente entre la suma de los productos de las marcas de clase por su frecuencia, dividido entre el número total de observaciones.

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i x_i}{n}$$

Ejemplo 1

Determinar el promedio aritmético simple para el siguiente cuadro de distribución de frecuencias sobre el número de ovinos que tiene las familias de una comunidad campesina de Pampacolca.

$A< X_i \leq b$	X_i	n_i	n_iX_i
2 – 6	4	21	84
6 – 10	8	25	200
10 – 14	12	30	360
14 – 18	16	40	640
18 - 22	20	20	400
TOTAL		136	1684

$$\overline{X} = \frac{\sum_i^n n_i x_i}{n} \qquad \overline{X} = \frac{(4x21)+(8x25)+.....+(20x20)}{136} \qquad \overline{X} = \frac{1684}{136} = 12.38$$

Ejemplo 2

En una empresa de producción de chompas de fibra de alpaca se toma una muestra de obreros para determinar el rendimiento promedio por día, con los siguientes resultados. Calcular el promedio

Nº Chompas / Obrero / Día										
4.2	4.3	8.0	6.9	4.2	7.3	6.4	6.4	8.0	4.6	4.6
3.9	6.5	5.7	5.4	3.9	6.9	6.5	5.1	5.7	3.6	7.0
6.1	5.2	6.4	6.5	4.3	5.7	4.3	6.9	6.4	8.2	5.3
5.6	3.8	7.3	7.3	5.4	5.9	5.4	4.6	7.3	3.5	5.7
5.9	7.1	5.1	8.4	6.9	6.1	6.5	3.7	5.1	5.7	7.7
6.1	4.5	4.5	6.9	5.7	5.2	4.3	3.8	5.4	6.4	7.3
5.2	5.6	5.6	5.7	5.1	8.0	5.4	7.1	6.5	7.3	5.1

1. Determinar el Numero de clases

   n = 77
   K =1 + 3.3 log (77)
   K = 1 + 3.3 * 2
   K $\cong$ 7.6 $\cong$ 8 clases

2. Determinación del rango (R)

   R = $X_{MAX} - X_{MIN}$
   X_{MAX} = 8.4

$X_{MIN} = 3.5$
$R = 4.9$

3. Determinación del Intervalo de clase (I)

$$I = \frac{R}{k}$$

$$I = \frac{4.9}{8}$$

$I = 0.6125$
$I \cong 0.7$

4. Determinación de la amplitud de la tabla (A)

$A = k \times I$
$A = 8 \times 0.7$
$A = 12.0$

5. Determinación del límite inferior de la primera clase (L)

$$Li = X_{MIN} - \frac{A - R}{2}$$

$$Li = 3.9 - \frac{5.6 - 4.9}{2}$$

$$Li = 3.15 \approx 3.2$$

6. Determinación de los Límites restantes

Se le agrega el intervalo de clase (0.7) al límite inferior de la primera clase (3.2) y así sucesivamente hasta completar los límites de las 8 clases:

a<X≤ b	
3.2	3.9
3.9	4.6
4.6	5.3
5.3	6.0
6.0	6.7
6.7	7.4
7.4	8.1
8.1	8.8

7. Conteo para determinar la frecuencia (n_1)

   Consiste en ubicar los valores de cada una de las observaciones dentro de las clases del cuadro de distribución de frecuencias y determinar la frecuencia de cada clase.

A < X ≤ b		n	X_i	$n_i X_i$
3.2	3.9	7	3.55	24.85
3.9	4.6	11	4.25	45.75
4.6	5.3	9	4.95	44.55
5.3	6.0	17	5.65	96.05
6.0	6.7	13	6.35	82.55
6.7	7.4	14	7.05	98.70
7.4	8.1	4	7.775	31.00
8.1	8.8	2	8.45	16.90
TOTAL		77		441.35

Calcular el promedio aritmético simple en datos agrupados

$$X = \frac{\sum_{i=1}^{n} n.x_i}{n}$$

$$X = \frac{441.35}{77} = 5.7318$$

$$X = 5.73$$

PROMEDIO ARITMÉTICO PONDERADO

Notación

μ = Promedio de población
$\overline{X}$ = promedio de muestra

Definición

Es el valor obtenido de la suma de cada una de las observaciones multiplicadas por sus ponderados, que luego se divide entre la suma de los ponderados.

$$\overline{X} = \frac{\sum_i^n p_i X_i}{\sum_i^n p_i}$$

Ejemplo

Calcular la esperanza de vida promedio para los departamentos del sur, teniendo en cuenta su población y esperanza de vida departamental.

Población y Esperanza de Vida al Nacer (años)

Departamento	Esperanza de vida en años X_i	Población 2012 P_i	Población 2012 p_iX_i
Apurímac	61.8	485,934	30030721.2
Arequipa	71.9	1139.599	81937168.1
Cusco	60.2	1252,201	75382500.2
Madre de Dios	67.2	107,664	7235020.8
Moquegua	72.5	167,251	12125697.5
Puno	60.6	1313,571	79602402.6
Tacna	72.8	317,619	23122663.2
Total		4783,839	309436173.6

$$\overline{X} = \frac{\sum_i^n p_i X_i}{\sum_i^n p_i} = \frac{(61.8x485934) + (71.9x1139599) + ... + (72.8x317619)}{4783839}$$

$$\overline{X} = \frac{309436173.6}{4783839}$$

$$\overline{X} = 64.7 \quad años$$

La esperanza de vida a nivel del sur del Perú es de 64.7 años

PROMEDIO CUADRÁTICO

Definición

Es el valor obtenido con respecto a datos observados

$$\overline{X} = \sqrt{\frac{\sum_i^n X^2}{N}}$$

Ejemplo 1

Calcular el promedio cuadrático de 7, 9, 8, 5, 6

$$\overline{X} = \sqrt{\frac{7^2 + 9^2 + 8^2 + 5^2 + 6^2}{5}}$$

$$\overline{X} = \sqrt{\frac{255}{5}} = 7.14$$

MEDIANA (Med)

La **mediana** resulta de acomodador los números de menor a mayor y el número de en medio es la mediana y si son dos números los sumas y los divides entre dos

Ejemplo

Hallar la Mediana de: 23, 30, 28, 27, 22, 21, 20

Ordenando: 20, 21, 22, 23, 27, 28, 30

Mediana

Ejemplo

Hallar la Mediana de: 5, 9, 7, 8, 4, 6, 3, 11

Ordenando: 3, 4, 5, 6, 7, 8, 9, 11

La Mediana debe de estar entre 6 y 7

$$Mediana = \frac{6+7}{2} = 6.5$$

LA MODA

Se extrae observando qué número se repite más veces

Ejemplo

El peso los recién nacidos en kilogramos son: 3.500, 4.000, 2.800, 2.950, 3.500, 2,600, 3.500, 4.000

Ordenando 2.600, 2.800, 2.950, 3.500, 3.500, 3.500, 4.000, 4.000

El peso que se repite más es de 3.500 kg.

LA MEDIA GEOMÉTRICA (MG)

Para extraer la MG, se usa:

$$MG = \sqrt[n]{(X_1)(X_2)(X_3)\ldots(X_n)}$$

Ejemplo

Sean los números: 6 y 54 hallar la media geométrica:

$$MG = \sqrt[2]{6\ X\ 54} = \sqrt[2]{324} = 18$$

Ejemplo

Sean los números 3, 9 y 27 hallar la media geométrica:

$$MG = \sqrt[3]{3\ X\ 9\ X\ 27} = \sqrt[3]{729} = 9$$

LA MEDIANA EN DATOS AGRUPADOS

Se encuentra en el intervalo donde la frecuencia acumulada llega hasta la mitad de la suma de las frecuencias absolutas, FA ≥ = N/2

$$Med = L_I + \left[\frac{\frac{N}{2} - F}{\ell}\right](C)$$

Dónde:

- L_I = Límite inferior de la clase que contiene a la mediana
- F = Frecuencia acumulada que precede a la clase de la mediana.
- ℓ = Frecuencia de clase de la mediana.
- C = Intervalo de clase.
- N = Número de datos

Ejemplo

Calcular la Mediana en Datos Agrupados

PRESIÓN	ℓ	F
50 - 59	3	3
60 - 69	7	10
70 - 79	18	28
80 - 89	12	40
90 - 99	8	48
100 - 109	2	50

De la tabla:

$N/2 = 50/2 = 25$
$L_I = 70$
$N = 50$
$\ell = 18$
$C = 10$
$F = 10$

$$Med = 70 + \left[\frac{\frac{50}{2} - 10}{18} \right] (10) = 78.33$$

MODA EN DATOS AGRUPADOS

Primeramente hay que ubicar en el cuadro de frecuencias, cual es la que tiene mayor valor, la cual indica que hi se encuentra la moda

Se calcula con la siguiente formula:

$$M_o = L_i + \left[\frac{D_1}{D_1 + D_2} \right] i$$

L_i = Límite Inferior de la clase modal
D_1 = Diferencia Modal entre la Frecuencia Modal y la Pre-modal
D_2 = Diferencia Modal entre la Frecuencia Modal y la Post-modal
i = Valor del Intervalo de clase

Ejemplo

Se tiene el siguiente cuadro de frecuencias:

INTERVALO	FRECUENCIA	
150 - 190	10	
190 - 230	20	
230 - 270	30	Moda
270 - 310	25	
310 - 350	15	

L_i = Límite Inferior de la clase modal = 230

D_1 = Diferencia Modal entre la Frecuencia Modal y la Pre-modal
= 30 – 20 = 10

D_2 = Diferencia Modal entre la Frecuencia Modal y la Post-modal
= 30 – 25 = 5

i = Valor del Intervalo de clase = 270 – 230 = 40

Reemplazando:

$$M_o = 230 + \left[\frac{10}{10+5}\right]40 = 256.6$$

EJERCICIOS

1. Fardos de lana reportados por la Empresa Michel & Compañía de la Ciudad de Arequipa

100	60	15	30	80	90	110	40	80	70	100
50	60	100	70	80	60	110	15	100	65	80
50	40	40	120	100	80	80	80	30	100	50
50	50	100	120	80	40	60	65	100	100	60
90	100	80	60	100	60	40	100	100	100	100
100	100	80	120	80	100	80	100	45	100	40
80	80	100	100	100	70	80	80	40	100	55
65	70	100	60	50	120	70	50	80	80	70
100	10	100	50	35	100	35	80	75	80	100
70	50	70	110	80	50	100	30	40	100	60
50	100	70	110	110	50	100	100	30	100	100

Elaborar y calcular:

a) Un cuadro de distribución de frecuencias
b) Calcular el promedio aritmético simple en datos agrupados
c) Calcular el Decil 30, el Percentil 75
d) Calcular la Mediana y la Moda en datos agrupados

2. Los datos corresponden a una muestra de la venta semanal en soles de una microempresa de calzados entre los años de 2010-2015

806	290	1108	529	968	505	3135	850	308	732	768	508	656	611
466	882	950	769	899	737	524	428	2148	799	520	364	938	430
303	195	722	775	824	314	478	565	1847	560	461	843	1091	394
674	307	1314	417	969	701	1217	1010	661	1032	1443	868	2117	877
750	40	1083	2757	455	2426	1464	574	2090	1208	779	791	1808	2757
459	29	2217	974	1153	485	1005	405	847	339	323	462	4641	438
885	63	608	601	626	1186	571	512	1078	712	622	703	1113	560

Elaborar y calcular:

a) Un cuadro de distribución de frecuencias
b) Calcular el promedio aritmético simple en datos agrupados
c) Calcular el Decil 10, el Percentil 95
d) Calcular la Mediana y la Moda en datos agrupados

3. Los datos corresponden a una muestra de precios de papa en soles por kilogramo, en diferentes mercados de la feria de los lunes de la ciudad de Arequipa, durante el primer trimestre del año 2014

0.45	**0.50**	**0.40**	**0.60**	**0.55**	**0.55**	**0.50**	**0.65**	**0.38**	**0.63**
0.46	**0.88**	**0.95**	**0.76**	**0.89**	**0.73**	**0.52**	**0.42**	**0.48**	**0.43**
0.30	**0.75**	**0.72**	**0.77**	**0.84**	**0.34**	**0.47**	**0.56**	**0.84**	**0.39**
0.67	**0.37**	**0.34**	**0.47**	**0.69**	**0.70**	**0.71**	**0.81**	**0.66**	**0.87**
0.75	**0.49**	**0.83**	**0.77**	**0.45**	**0.46**	**0.44**	**0.57**	**0.90**	**0.77**
0.49	**0.95**	**0.72**	**0.74**	**0.53**	**0.45**	**0.52**	**0.45**	**0.87**	**0.38**
0.88	**0.63**	**0.68**	**0.61**	**0.66**	**0.86**	**0.57**	**0.52**	**0.78**	**0.56**
0.49	**0.83**	**0.77**	**0.45**	**0.56**	**0.56**	**0.78**	**0.62**	**0.57**	**0.51**
0.95	**0.72**	**0.74**	**0.53**	**0.75**	**0.45**	**0.48**	**0.56**	**0.78**	**0.45**

Elaborar y calcular:

a) Un cuadro de distribución de frecuencias
b) Calcular el promedio aritmético simple en datos agrupados
c) Calcular el Decil 20, el Percentil 80
d) Calcular la Mediana y la Moda en datos agrupados

CAPITULO IV

MEDIDAS DE DESVIACIÓN ó DISPERSIÓN

CAPITULO IV

MEDIDAS DE DESVIACIÓN ó DISPERSIÓN

Son usadas para medir la variación entre los datos y son las siguientes:

1. Desviación Estándar
2. Varianza
3. Coeficientes de Variabilidad

Desviación estándar

Notación

σ **=** desviación estándar de población
s = desviación estándar de muestra

Definición

a) Para una población

Está definida por:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - u)^2}{N}}$$

Esta fórmula se simplifica para su cálculo y queda escrita de la manera siguiente:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{N} - u^2}$$

Ejemplo:

Los datos corresponden a las edades de los trabajadores de una microempresa de confección de ropa en la ciudad de Arequipa: 33, 40, 28, 32, 25, 37, 42, 31, 34, 29, 30, y 41

a) Calcular el promedio

Estos datos corresponden a una población por lo tanto la fórmula que se debe utilizar es:

$$u = \frac{\sum_{i=1}^{n} X_i}{N} = \frac{sumatoria\ de\ valores\ de\ observaciones}{numero\ de\ observaciones\ de\ la\ muestra}$$

$$u = \frac{\sum_{i=1}^{n} X}{N} = \frac{33+40+28+32+25+37+42+31+34+29+30+41}{12}$$

u = 33.5

b) Calcular la desviación estándar

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{N} - u^2}$$

$$\sigma = \sqrt{\frac{33^2 + ...41^2}{12} - 33.5^2} \qquad \sigma = \sqrt{\frac{13794}{12} - 33.5^2}$$

$$\sigma = 5.22$$

b) Para una muestra

En este caso se debe hacer una corrección, restando en uno el denominador de modo que la fórmula de desviación estándar es:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}}$$

Para efectos del cálculo la fórmula puede ser simplificada de la siguiente forma:

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

Ejemplo.

Una empresa comercializadora de productos de limpieza para el hogar realiza un pequeño sondeo del gasto promedio semanal por familia en soles, en la ciudad de Arequipa, obteniendo los siguientes resultados:

110	**50**	**85**	**90**	**50**
100	**40**	**80**	**120**	**35**
45	**58**	**75**	**110**	**45**
75	**70**	**90**	**85**	**58**

a) Calcular el promedio

$$u = \frac{\sum_{i=1}^{n} X_i}{N} = \frac{sumatoria\ de\ valores\ de\ observaciones}{numero\ de\ observaciones\ de\ la\ muestra} = \frac{1471}{20} = 73.55$$

b) Calcular la desviación estándar

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

$$s = \sqrt{\frac{120403 - 1471^2 / 20}{19-1}} = 25.35$$

VARIANZA

Notación

σ = Varianza de población
S² = Varianza de muestra

Definición

a) Para una Población

La varianza es el valor que se obtienen de elevar al cuadrado la desviación estándar. Todas las fórmulas de la desviación estándar, son también válidas para la varianza, elevando todas ellas al cuadrado

$$\sigma^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{N} = \frac{\sum_{i=1}^{n} X_i^2}{N} - \mu^2$$

Ejemplo 1

Dados los siguientes datos: 13, 10, 8, 12, 12, 11, 14, 9, 10 y 11 que corresponden a la nota final del curso de Matemáticas

a) Calcular el promedio

$$u = \frac{\sum_{i=1}^{n} X_i}{N} = \frac{110}{10} = 11$$

b) Calcular la Varianza

$$\sigma^2 = \frac{\sum_{i=1}^{n} X_i^2}{N} - \mu^2 = \frac{1240}{10} - 11^2 = 3$$

b) Para una muestra

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

Ejemplo 2

En una muestra de precios de papa se obtuvo los siguientes resultados en soles por kilogramo: 0.45, 0.50, 0.40, 0.60, 0.55, 0.55, 0.60, calcular

$$s^2 = \frac{\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2 / n}{n-1} = \frac{1.9375 - 3.65^2 / 7}{7-1}$$

$$s^2 = 0.0057$$

Ejemplo 3

En una encuesta realizada en un centro educativo estatal a través de una muestra tomada al azar, se obtienen los siguientes datos sobre el número de hermanos que tiene el estudiante:

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

a) Calcular el promedio aritmético

$$u = \frac{\sum_{i=1}^{n} X_i}{N}$$

$$u = \frac{\sum_{i=1}^{n} X_i}{N} = \frac{190}{45} = 4.222$$

b) Calcular la desviación estándar

$$s = \sqrt{\frac{\sum_{i=1}^{n} n_i X_i^{\,2} - (\sum_{i=1}^{n} n_i X_i)^2 / n}{n-1}}$$

$$s = \sqrt{\frac{1004 - 190^2 / 45}{45-1}} = 2.141462$$

c) Calcular la varianza

$s^2 = (2.141462)^2 = 4.585859$

DESVIACIÓN ESTÁNDAR EN DATOS AGRUPADOS

a) Para una población

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} n_i (X_i - \mu)^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} n_i X_i}{n} - u^2}$$

b) Para una muestra

$$s = \sqrt{\frac{\sum_{i=1}^{n} n_i (X_i - \overline{X})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} n_i X_i^{\,2} - (\sum_{i=1}^{n} n_i X_i)^2 / n}{n-1}}$$

Ejemplo 1

En la escuela primaria de menores N° 40140 de Pampacolca Arequipa la asistencia a clases en el sexto grado por alumno para el año 2016 fue:

121	159	151	137	155	162	140	108	150	159	104
136	127	102	144	129	147	128	159	142	134	138
122	160	101	163	158	145	106	153	126	147	156
104	108	105	135	142	125	118	140	109	120	161
142	142	108	149	132	134	144	133	140	147	126
146	118	104	124	125	113	160	141	139	163	
158	164	135	133	122	148	125	157	123	165	
103	101	161	136	138	128	124	134	162	163	
154	110	114	146	141	152	155	116	139	145	
149	105	146	151	149	125	128	151	162	147	

a) Elaborar un cuadro de distribución de frecuencias

1. Determinación del número de clases
   N = 50
   K = 1+3.3. Log (n) = 1 + 3.3 log(105)
   K = 7.6699
   K ~ 8.0

2. Determinación del rango (R)
   $R = X_{MAX} – X_{min}$
   $X_{MAX} = 165$
   $X_{min} = 101$
   R = 64.0

3. Determinación del intervalo de clase (I)

$$I = \frac{R}{k} = \frac{64}{7.7}$$

I = 8.3

4. Determinación de la amplitud de la tabla (A)

A = K x I = 8 x 8.3
A = 66.4

5. Determinación del límite inferior de la primera clase (L_i)

$$Li = X_{MIN} - \frac{A-R}{2} = 101 - \frac{66.4-64}{2}$$

$$Li = 99.8$$

6. Determinación de los limites restantes

Se le agrega el intervalo de clase (8.3) al límite inferior de la primera clase (99.8) y así sucesivamente hasta completar los límites de las 8 clases:

a<X≤ b	
99.8	108.1
108.1	116.4
116.4	124.7
124.7	133
133	141.3
141.3	149.6
149.6	157.9
157.9	166.2

7. Conteo para determinar la frecuencia (n_i)

a<X≤ b		n_i	X_i	X_i^2	n_iX_i	$n_iX_i^2$
	108.1	13	103.95	10805.6	1351.35	140472.83
108.1	116.4	5	112.25	12600.06	561.25	63000.313
116.4	124.7	9	120.55	14532.3	1084.95	130790.72
124.7	133	12	128.85	16602.32	1546.2	199227.87
133	141.3	17	137.15	18810.12	2331.55	319772.08
141.3	149.6	18	145.45	21155.7	2618.1	380802.65
149.6	157.9	13	153.75	23639.06	1998.75	307307.81
157.9	166.2	18	162.05	26260.2	2916.9	472683.65
TOTAL		**105**			**14409.05**	**2014057.9**

b) Calcular el promedio aritmético simple en datos agrupados

$$\overline{X} = \frac{\sum_i^n n_i X_i}{n}$$

$$\overline{X} = \frac{14409.05}{105}$$

$$\overline{X} = 137.2$$

c) Calcular la desviación estándar en datos agrupados

$$s = \sqrt{\frac{\sum_{i=1}^n n_i X_i^{\ 2} \ -(\sum_{i=1}^n n_i X_i)^2 / n}{n-1}}$$

$$s = \sqrt{\frac{2014057.9 - 14409.05^2 / 105}{105-1}}$$

$$s = 18.79$$

VARIANZA EN DATOS AGRUPADOS

a) Para una población

$$\sigma^2 = \frac{\sum_{i=1}^{n} N_i (X_i - \mu)^2}{N} = \frac{\sum_{i=1}^{n} N_i X_i^{\,2}}{N} - \mu^2$$

b) Para una muestra

$$s^2 = \frac{\sum_{i=1}^{n} n_i (X_i - \overline{X})^2}{n-1} = \frac{\sum_{i=1}^{n} n_i X_i^{\,2} - (\sum_{i=1}^{n} n_i X_i)^2 / n}{n-1}$$

Ejemplo

En una muestra tomada al azar en la ciudad de Arequipa, sobre el ingreso familiar en soles se obtuvo los siguientes resultados:

Ingreso en S/.	$\mathbf{n_i}$	$\mathbf{X_i}$	$\mathbf{X_i^2}$	$\mathbf{n_iX_i}$	$\mathbf{N_iX_i^2}$
120 - 326	22	223	49729	4906	1094038
327 - 533	50	430	184900	21500	9245000
534 - 740	51	637	405769	32487	20694219
741 - 947	47	844	712336	39668	33479792
948 - 1154	16	1051	1104601	16816	17673616
1155 - 1361	9	1258	1582564	11322	14243076
1362 -1568	3	1465	2146225	4395	6438675
1569 -1775	1	1672	2795584	1672	2795584
1776 - 1982	1	1879	3530641	1879	3530641
TOTAL	**200**			**134645**	**109194641**

a) Calcular el promedio

$$\overline{X} = \frac{\sum_{i}^{n} n_i X_i}{n} = \frac{134645}{200}$$

$$\overline{X} = 673.225$$

b) La Varianza

$$s^2 == \frac{\sum_{i=1}^{n} n_i X_i^{\,2} - (\sum_{i=1}^{n} n_i X_i)^2 / n}{n-1} = \frac{109194641 - 134645^2 / 200}{200-1} = 93207.3411$$

c) La Desviación estándar

$s = \sqrt{93207.3411}$

s = 305.3

COEFICIENTE DE VARIABILIDAD

Notación

CV = Coeficiente de variabilidad

Definición

Es el cociente entre la desviación estándar y el promedio, por lo general se le expresa en porcentaje, pero también puede ser utilizada en tanto por uno

$$CV = \frac{s}{\overline{X}} x \quad 100$$

Si el CV ≤ 30 Población Homogénea (Se puede utilizar para fines experimentales o investigación).

Si el CV >30 Población Heterogénea (No sirve)

Ejemplo

En una empresa, la distribución de utilidades en nuevos soles para el año 2006 fue de: 35000, 38000, 29000, 48000, 23000, 59000 y 36000. Calcular:

a) El Promedio

$$\mu = \frac{\sum_i^n X_i}{N}$$

$$\mu = \frac{268000}{7}$$

$$\mu = 38285.7143$$

b) La desviación estándar

$$\sigma = \sqrt{\frac{\sum_{i=1}^{X} X_i^2}{N} - \mu^2}$$

$$\sigma = \sqrt{\frac{11120000000}{7} - 38285.7143^2}$$

$$\sigma = 11080.411$$

c) Calcular el coeficiente de variabilidad

$$CV = \frac{\sigma}{\mu} x100$$

$$CV = \frac{11080.4111}{38285.7143} x100$$

$$CV = 28.94\%$$

Ejemplo

En el siguiente cuadro se muestra el rendimiento de papa en cien comunidades campesinas en toneladas métricas por hectárea.

Rendimiento En TM/Há	**n_i**	**X_i**	**X_i^2**	**n_iX_i**	**$n_iX_i^2$**
4.6 – 4.7	2	4.65	21.6225	9.3	43.25
4.7 – 4.8	8	4.75	22.5625	38.0	180.50
4.8 – 4.9	14	4.85	23.5225	67.9	329.32
4.9 – 5.0	27	4.95	24.5025	133.65	661.57
5.0 – 5.1	25	5.05	25.5025	126.25	637.56
5.1 – 5.2	16	5.15	26.5225	82.40	424.36
5.2 – 5.3	7	5.25	27.5625	36.75	192.94
5.3 – 5.4	1	5.35	28.6225	5.35	28.62
TOTAL	**100**			**499.6**	**2498.11**
PROMEDIO	**4.996**		**DESV. ESTANDAR**		**0.1459**
VARIANZA	**0.0213**		**C. VARIACIÓN**		**2.92**

Calcular:

a) calcular el promedio

$$\overline{X} = \frac{\sum_{1}^{n} n_i X_i}{n} = \frac{499.60}{100}$$

$$\overline{X} = 4.996$$

b) Calcular la varianza

$$s^2 == \frac{\sum_{i=1}^{n} n_i X_i^{\ 2} - (\sum_{i=1}^{n} n_i X_i)^2 / n}{n-1} = \frac{2498.11 - 499.6^2 / 100}{100-1} = 0.0213$$

c) Calcular la desviación estándar

$$\sigma = \sqrt{0.0213}$$
$$\sigma = 0.1459$$

d) Calcular el coeficiente de variabilidad

$$CV = \frac{\sigma}{\overline{X}} x100$$
$$CV = \frac{0.1459}{5} x100$$
$$CV = 2.92\%$$

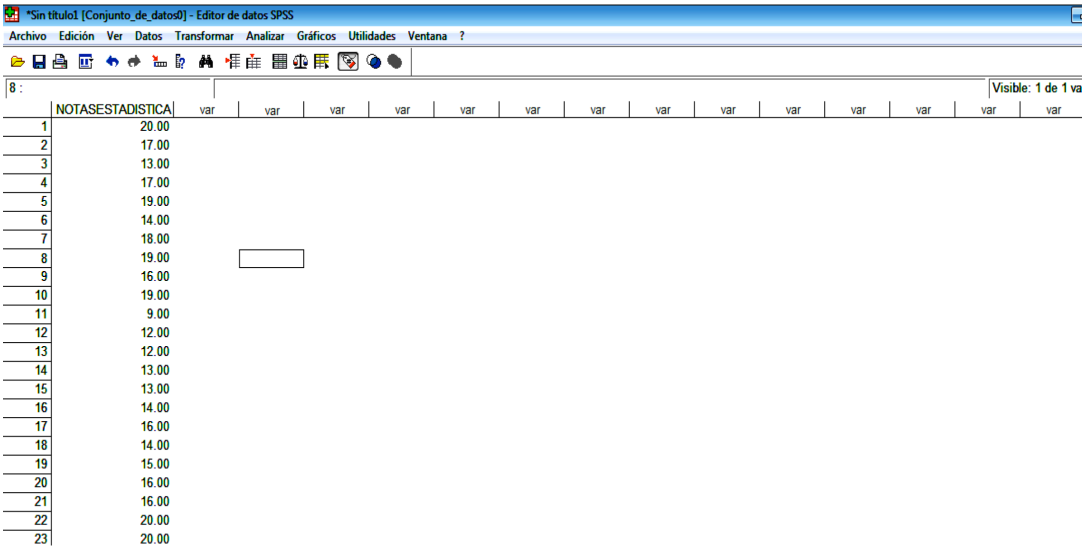
APLICACIÓN DEL SOFTWARE ESTADISTICO SPSS

Ejemplo.

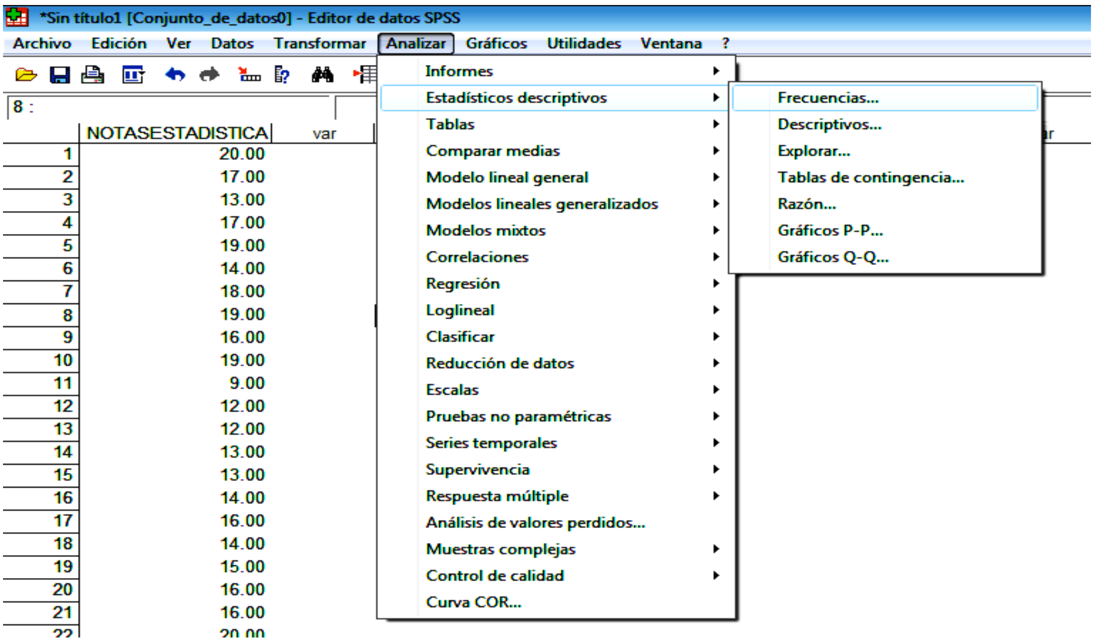
Se tienen las notas del curso de estadística se pide calcular el promedio, la mediana, la moda, la desviación estándar y la varianza

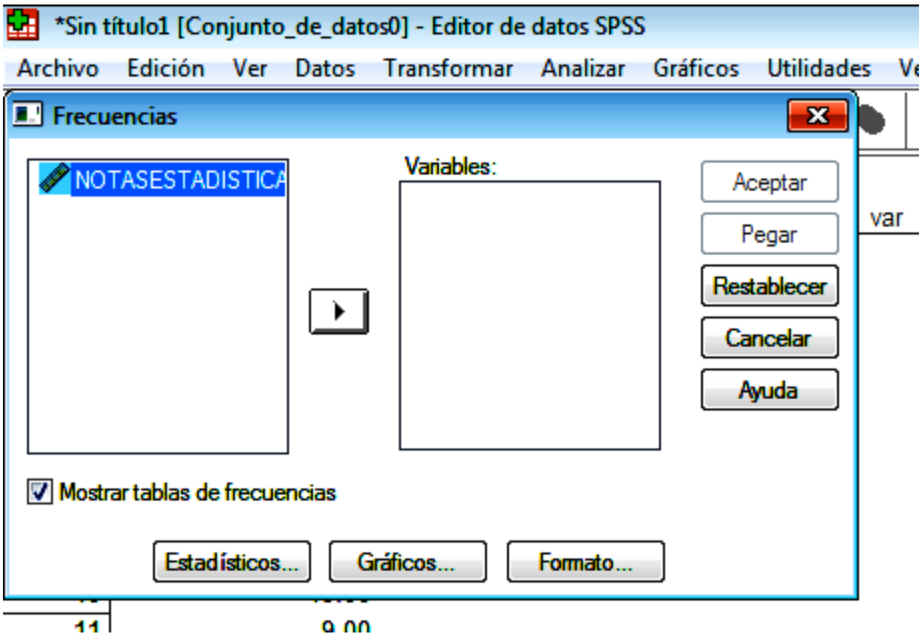
20	14	09	14	16	11	18	08	10	10
17	18	12	16	20	12	19	08	11	13
13	19	12	14	20	13	20	12	13	13
17	16	13	15	17	14	15	13	13	14
19	19	13	16	17	14	13	14	15	15

Se colocan los datos en el SPSSS



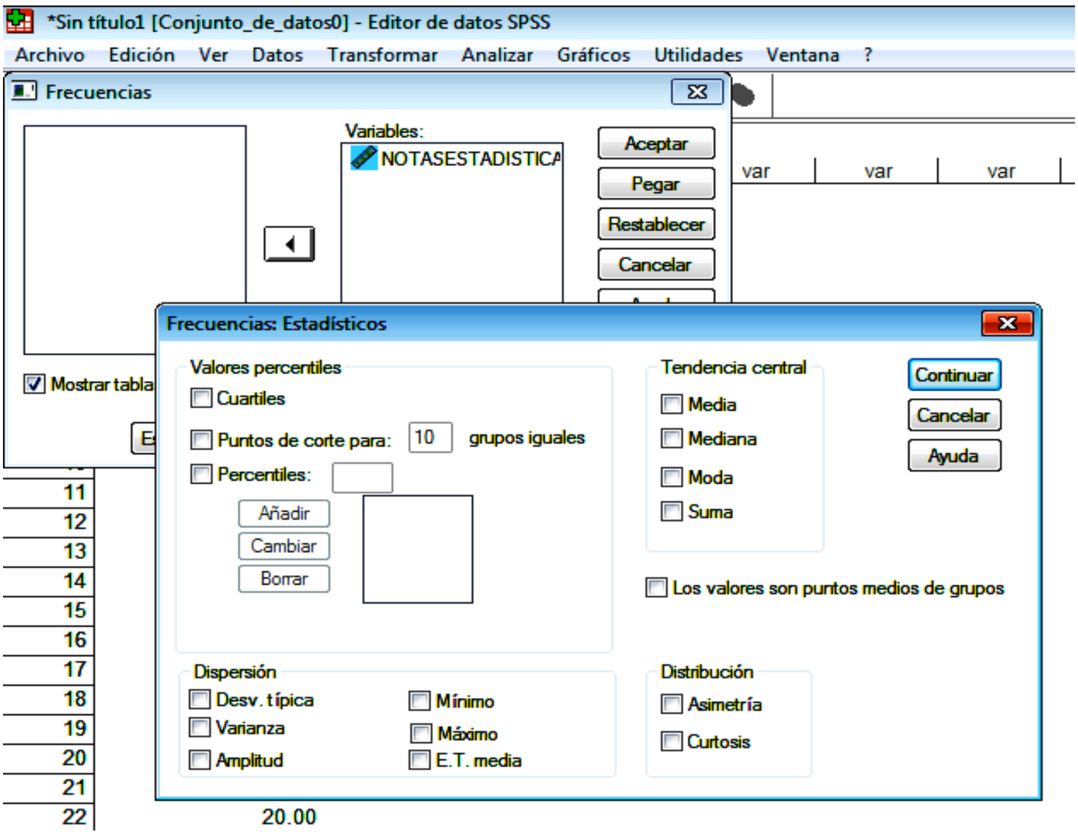
Luego Analizar – Estadisticos descriptivos - Frecuencias

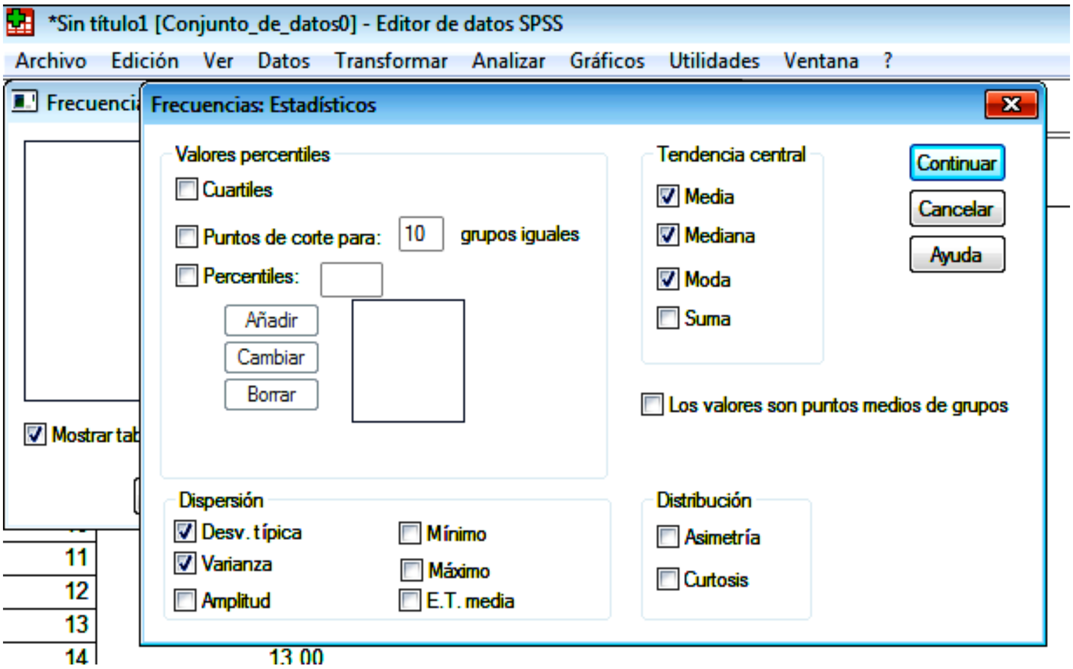


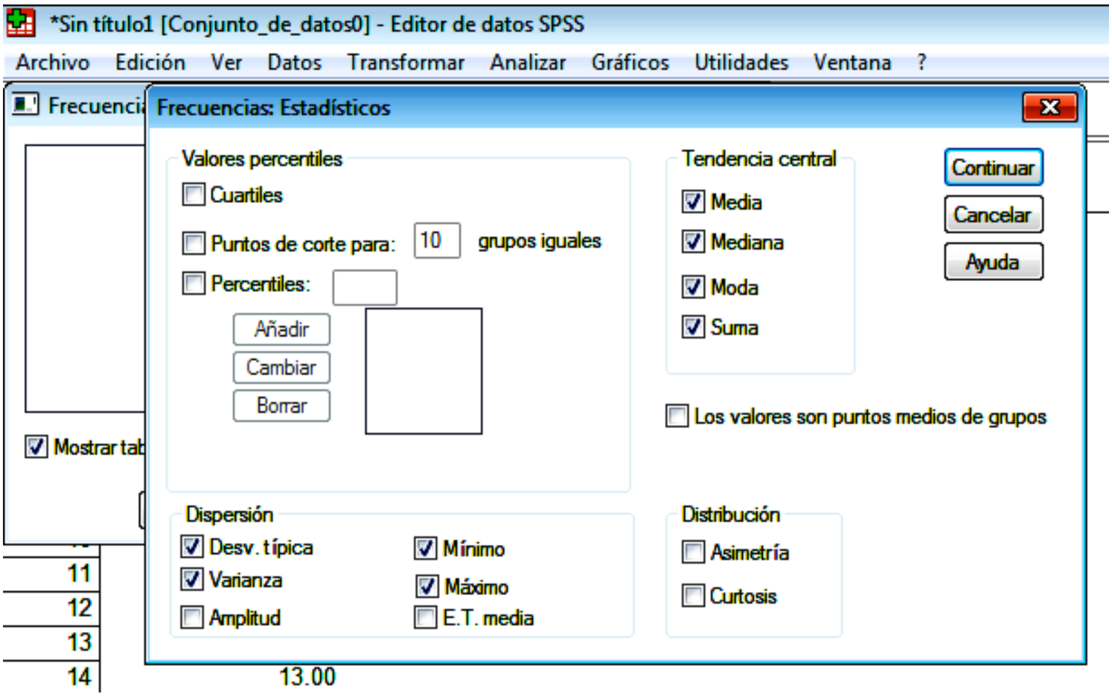


Pasar Notas estadistica a variables

Clic en Estadisticos y dar check a la Medida de Tendencia Central y Medidas de Dispersión

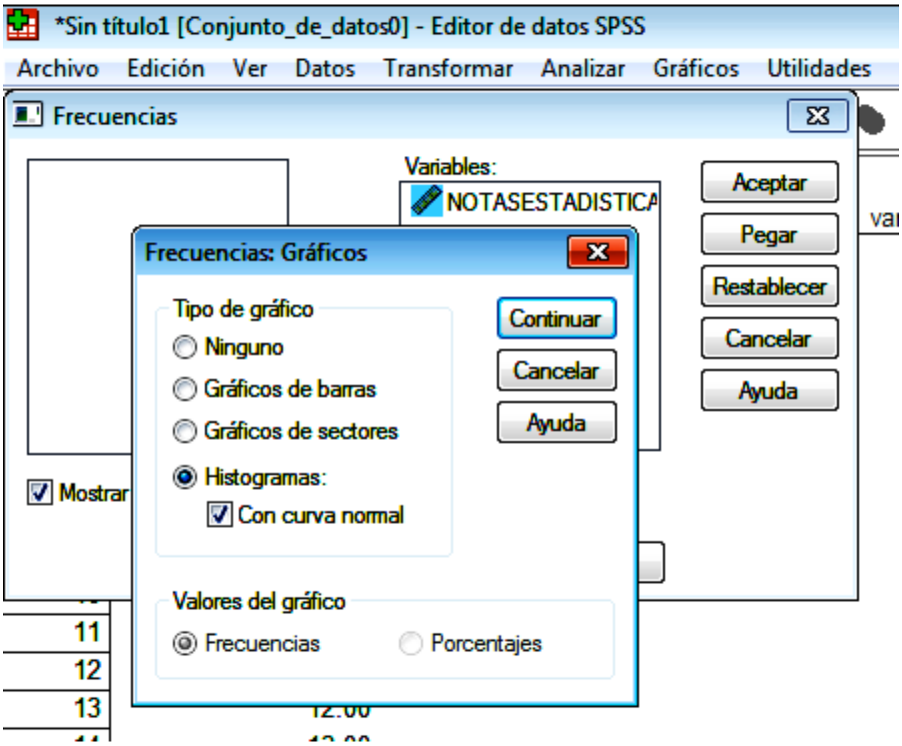






Continuar

Dar a graficos



Continuar – Aceptar

Frecuencias

[Conjunto_de_datos0]

Estadísticos

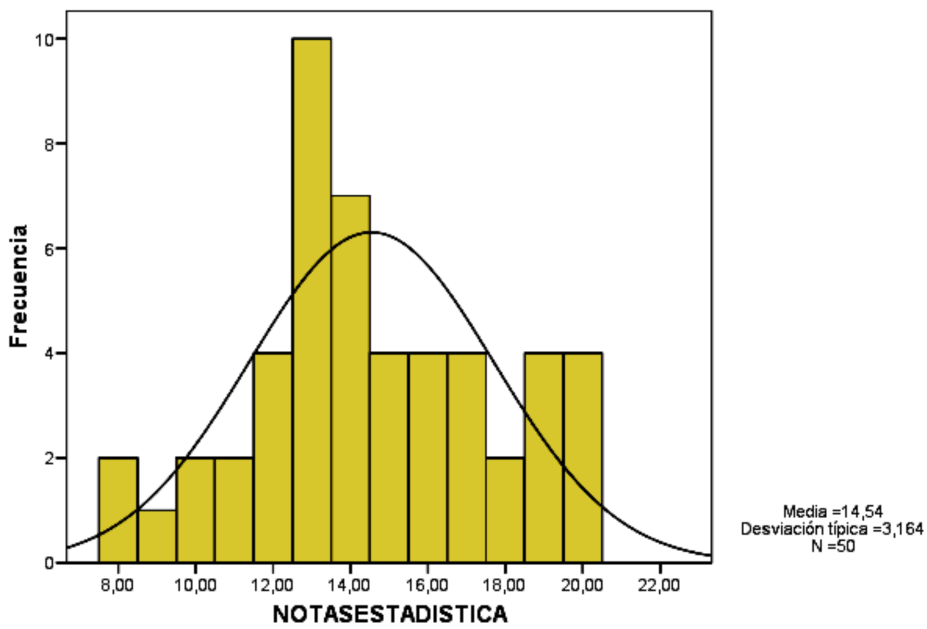
NOTASESTADISTICA

N	Válidos	50
	Perdidos	0
Media		14.5400
Mediana		14.0000
Moda		13.00
Desv. típ.		3.16363
Varianza		10.009
Mínimo		8.00
Máximo		20.00
Suma		727.00

NOTASESTADISTICA

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	8.00	2	4.0	4.0	4.0
	9.00	1	2.0	2.0	6.0
	10.00	2	4.0	4.0	10.0
	11.00	2	4.0	4.0	14.0
	12.00	4	8.0	8.0	22.0
	13.00	10	20.0	20.0	42.0
	14.00	7	14.0	14.0	56.0
	15.00	4	8.0	8.0	64.0
	16.00	4	8.0	8.0	72.0
	17.00	4	8.0	8.0	80.0
	18.00	2	4.0	4.0	84.0
	19.00	4	8.0	8.0	92.0
	20.00	4	8.0	8.0	100.0
	Total	50	100.0	100.0	

Histograma



EJERCICIOS

1. En un torneo de tiro con escopeta el equipo estrella consiguió los siguientes puntos: 50, 24, 30, 46, 48, 78, 26, 62, 38, 32.
Calcule:

a) La desviación estándar.
b) La Varianza.
c) El Coeficiente de Variabilidad

2. Se tienen los siguientes resultados de las notas promedio del Curso de Investigación de la maestría de la escuela de Postgrado

INTERVALO	FRECUENCIA
10 -- 14	5
14 -- 16	9
16 -- 18	15
18 -- 20	11

Calcule:

a) La desviación estándar.
b) La Varianza.
c) El Coeficiente de Variabilidad

3. Se tiene el peso en Kg, de 100 barras de acero, distribuidos en una tabla de frecuencias.

INTERVALO	FRECUENCIA
120 -- 124	10
125 -- 129	36
130 -- 134	84
135 -- 139	54
140 -- 144	16

Calcular:

a) La desviación estándar.
b) La Varianza.
c) El Coeficiente de Variabilidad

CAPÍTULO V

DISTRIBUCIONES DE PROBABILIDAD

CAPÍTULO V

DISTRIBUCIONES DE PROBABILIDAD

DISTRIBUCIÓN BINOMIAL

La fórmula que se usa es:

$$B(x;n;p) = C_n^x \, p^x \, q^{(n-x)}$$

n = número de intentos independientes que se realizan
p = probabilidad de éxito
q = probabilidad de fracaso
p + q = 1
x = número de éxitos que se desean tener

Ejemplo.

Una prueba de inteligencia está compuesta de 10 preguntas, cada una de las cuales tiene 4 respuestas, siendo sólo una de ellas correcta. Si se contesta aleatoriamente se pide:

a) Probabilidad de acertar 4 preguntas.
b) Probabilidad de no acertar ninguna

a) Probabilidad de acertar 4 preguntas.

Se trata de una distribución binomial con:

$p = \frac{1}{4}$, n=10 y r=4

$$p(x = 4) = \binom{10}{4} \cdot \left(\frac{1}{4}\right)^4 \cdot \left(\frac{3}{4}\right)^6 = 0{,}1460$$

b) Probabilidad de no acertar ninguna

En este caso r= 0

$$p(x=0)=\binom{10}{0}\cdot\left(\frac{3}{4}\right)^{10}=0,0563$$

Ejemplo.

En una fábrica se realiza el control de calidad de los productos, para ello se seleccionan 10 de los artículos elaborados y se inspeccionan con el fin de determinar si son defectuosos o no. La probabilidad de que un producto sea defectuoso es del 12%.
Determinar la probabilidad de que de los 10 artículos seleccionados 3 artículos sean defectuosos.

$$B(x;n;p)=C_n^x\, p^x q^{(n-x)}$$

n = 10
p = 12% = 0.12
q = 1 – p = 1 - 0.12 = 0.88
x = 3

Reemplazando:

$$B(3;10;0.12)=C_{10}^3\,(0.12)^3(0.88)^{(10-3)}=0.0847$$

La probabilidad de que de los 10 artículos seleccionados 3 sean defectuosos es de 0.0847, es decir es del 8.47%

DISTRIBUCIÓN DE POISSON

Es otra distribución discreta de probabilidad, llamada así en memoria de Simeon Denis Poisson(1781-1840), quien la descubrió. Se usa en muchas situaciones que se refieren a fenómenos que ocurren en un intervalo continuo de espacio o de tiempo, aunque dicha continuidad puede dividirse en intervalos más pequeños.
Y se calcula con la siguiente formula:

$$p(X,\lambda)=\frac{\lambda^x . e^{-\lambda}}{X!}$$

X = 0, 1, 2,…….
X = es el número de éxitos
λ = representa el promedio de éxitos esperados

Ejemplo.

En un supermercado, una cajera atiende en promedio 7 personas en una hora de trabajo.
Determinar la probabilidad de que en una hora determinada la cajera atienda 10 personas

$$p(X,\lambda) = \frac{\lambda^{x} . e^{-\lambda}}{X!}$$

X = 10
λ = 7

$$p(10;7) = \frac{(7)^{10}(e)^{-7}}{10!} = 0.0710$$

La probabilidad de que en un día determinado la cajera atienda 10 personas es de 0.0710, es decir, del 7.10%

Ejemplo.

El 8% de los registros contables de una empresa presentan algún problema, si un auditor toma una muestra de 40 registros ¿Calcular la probabilidad de que existan 5 registros con probemas?

n = 40
P = 0.08
Lambda = 40 * 0.08 = 3.2
X = 5
e = 2.71828

Entonces: **$P(X=5) = (3.2^{5}) (e^{-3.2}) / 5! = 0.1139793$**

DISTRIBUCIÓN HIPERGEOMÉTRICA

La fórmula que se usa es:

$$h(x;n;k;N) = \frac{C_{k}^{x} C_{(N-k)}^{(n-x)}}{C_{N}^{n}}$$

x = 0, 1, 2, 3,….,n son los valores que se sacan de la muestra
n = tamaño de muestra
N = total resultados posibles
k = número de éxitos
N - k = número de fracasos

Ejemplo.

En una empresa hay 28 empleados en el departamento administrativo y 43 en el departamento de ventas. Se desea seleccionar un comité de 5 empleados para qe asistan a un evento. Determinar la probabilidad de que de los 5 empleados que se seleccionan 3 pertenezcan al departamento administrativo

x = 3
n = 5
N = 28 + 43 = 71
k = 28
N - k = 71 – 28 = 43

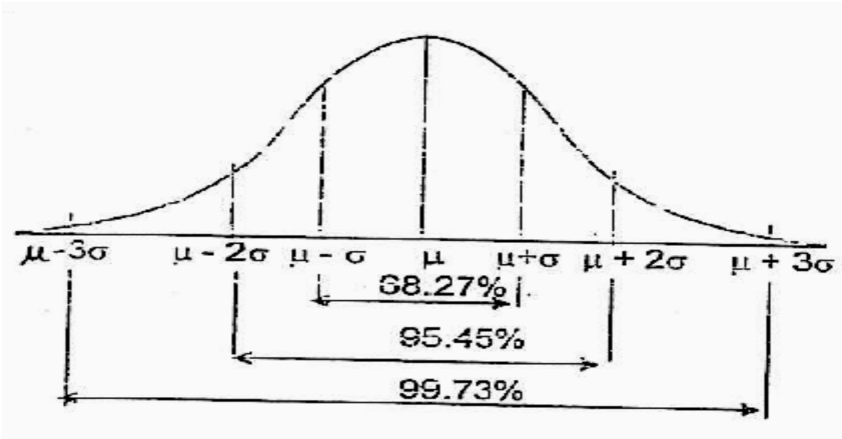
$$h(3;5;28;71) = \frac{C_{28}^{3} C_{(71-28)}^{(5-3)}}{C_{71}^{5}} = 0.2272$$

La probabilidad de que de los 5 empleados seleccionados en la empresa 3 pertenezcan al departamento administrativo es de 0.2272, es decir, del 22.72%

DISTRIBUCIÓN DE PROBABILIDAD NORMAL

Esta es una distribución de probabilidad continua, simétrica, mesocúrtica y gráficamente representada tiene la forma de una campana.
Como la distribución normal es una distribución continua, el número de casos que ella incluye puede ser infinitamente grande.
La probabilidad de ocurrencia de un cierto evento, es medida de acuerdo con la proporción del área que dicho evento representa bajo la curva normal..
Vale la pena aclarar, que en una distribución normal se cumple lo siguiente:

a) El 68.27% de los datos caen entre $\mu \pm 1\sigma$.
b) El 95.45% de los datos caen entre $\mu \pm 2\sigma$.
c) El 99.73% de los datos caen entre $\mu \pm 3\sigma$.



La ecuación de densidad normal es:

$$y = f(X) = \frac{1}{\sqrt{2\Pi\sigma^2}}\ e^{\frac{1}{2}(\frac{x-\mu}{\sigma})dx}$$

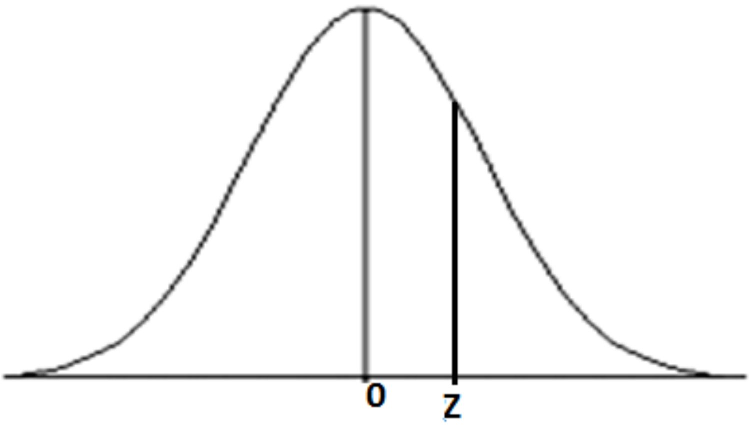
La probabilidad de que la variable X obtenga un valor en a y b es:

$$P(a \leq X \leq b) = \int_a^b f(X) = \frac{1}{\sqrt{2\Pi\sigma^2}}\ e^{\frac{1}{2}(\frac{x-\mu}{\sigma})dx}$$

Por lo tanto se requiere interacción numérica. Es por esto que para facilitar el uso de esta distribución, debemos estandarizar la variable X.

La notación X ˜ N (μ, σ^2) indica que la variable X tiene una distribución normal con media μ y varianza σ^2.

El rango de μ es $-\infty < \mu < \infty$ y el de σ^2 es $\sigma^2 \geq 0$. Por lo tanto el número posible de distribuciones normales es infinito: como no es posible contar con una tabla para que cada distribución, es que afortunadamente se ha determinado que cualquier variable X ˜ N (μ, σ^2) puede ser transformada en una variable Z ˜ N (0,1); la misma que si tiene una tabla, que puede encontrarse en cualquier texto de estadística o en una computadora.



La tabla de la probabilidad de que una variable se encuentre entre 0 y cualquier valor positivo Z.

$P(0 < Z < 1) = 0.3413$
$P(0 < Z < 2.56) = 0.4948$
$P(-1 < Z < 0) = 0.3413$
$P(-1 < Z < 2.56) = 0.3413 + 0.4948 = 0.8361$
$P(1 < Z < 2.56) = 0.4948 – 0.3413 = 0.1535$

El proceso sería si $P(0 < Z < Z^*) = 0.4929$. Hallar el valor de Z^* si $Z^* = 0$. Para esto recurrimos a la tabla y observamos a que valor Z le corresponde la probabilidad de 0.4929: de donde $Z^* = 2.45$.

Para la realización de los respectivos cálculos de las probabilidades se deben de tener en cuenta las siguientes propiedades.

- Si el problema dice mayor o menor que el valor obtenido de Z de la tabla se deberá de restar de 0.5

- En los intervalos si los dos valores se encuentran a la derecha de la media se deben de restar.

- En los intervalos si los dos valores se encuentran a la izquierda de la media se deben de restar.

- En los intervalos si los valores uno se encuentra a la izquierda y el otro valor se encuentra a la derecha se deben de sumar.

ESTANDARIZACIÓN DE X

El proceso de transformación de “X” a “Z” se llama la estandarización de X y la distribución N (0,1); se llama distribución normal estándar. La estandarización de “X” se realiza a través de la ecuación:

$$Z = \frac{X-\mu}{\sigma}$$

Es decir, si X ˜ N (μ, σ²), entonces $Z = \frac{X-\mu}{\sigma}$ ˜ (0,1)

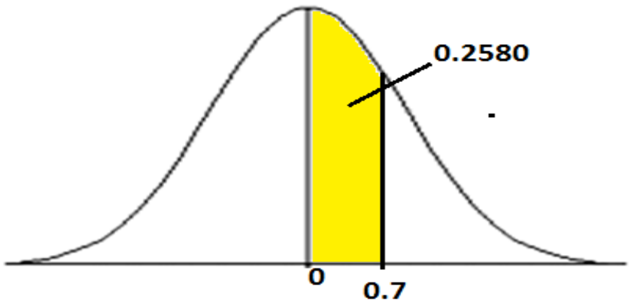
Ejemplo

Dado X ˜ N (4, 25), determinar:

a) P (4 < X < 7.5) $p = (\frac{4-4}{5}) < (\frac{X-4}{5}) < (\frac{7.5-4}{5})$

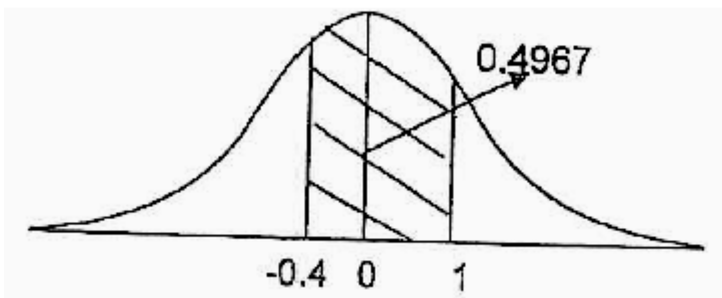
= P (0 < Z < 0.7) = 0.2580

(Este valor se encuentra en la tabla, calculadora o computadora)



b) P (2 < X < 9) = $p = (\frac{2-4}{5}) < Z < (\frac{9-4}{5})$

= P (-0.4 < Z < 1) = 0.4967



Ejemplo

La media del peso de barras de fierro con distribución normal para un lote de producción es de 10,000 kg. La desviación estándar es de 1,000 kg.
Calcular:

a) ¿Cuál es la probabilidad de que el peso de las barras de fierro este entre 7,900 kg. y 11,000 kg.?
b) ¿Cuál es la probabilidad de que el peso de las barras de fierro sea mayor de 12,000 kg.?
c) ¿Cuál es la probabilidad de que el peso de las barras de fierro sea menor de 8,500 kg?
d) ¿Cuál es la probabilidad de que el peso de las barras de fierro este entre 11,000 kg y 12,000 kg?

<u>SOLUCIÓN</u>

a) $$Z = \frac{X_i - \mu}{\sigma}$$

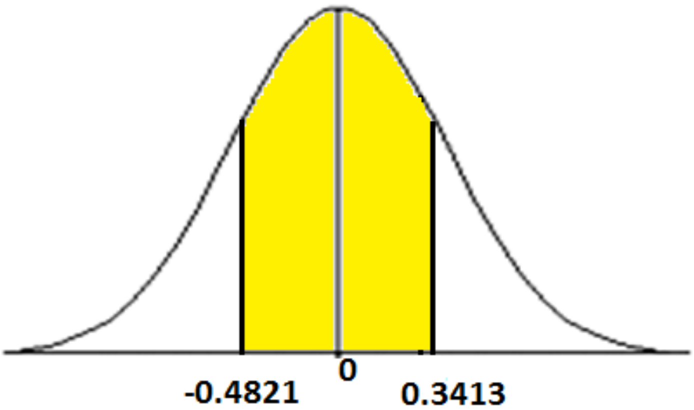
$$Z_1 = \frac{7{,}900 - 10{,}000}{1{,}000} = -2.10$$

De tablas $Z_1 = 0.4821$

$$Z_2 = \frac{11{,}000 - 10{,}000}{1{,}000} = 1$$

De tablas $Z_2 = 0.3413$

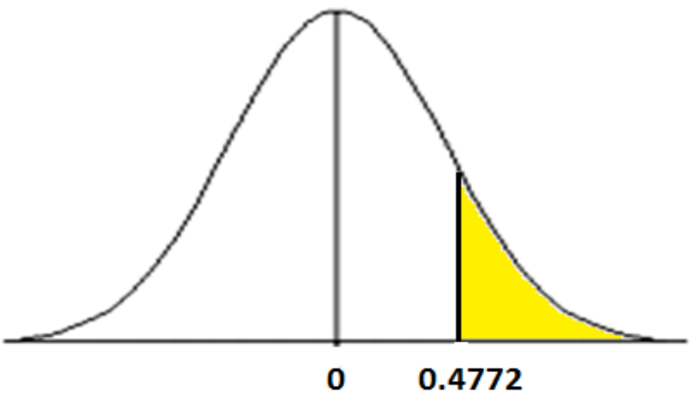
$$P(Z_1 \langle X \rangle Z_2) = 0.4821 + 0.3413 = 0.8234 \Rightarrow 82.34\%$$



b) $Z = \dfrac{12{,}000 - 10{,}000}{1{,}000} = 2$

De tablas $Z = 0.4772$

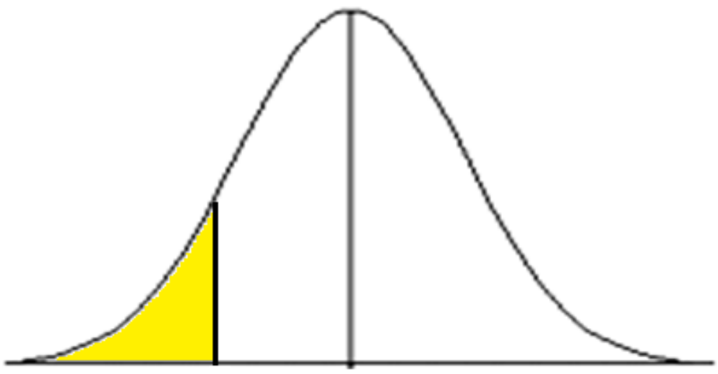
$P(X \rangle Z) = 0.5000 - 0.4772 = 0.0228 \Rightarrow 2.3\%$



c) $Z = \dfrac{8{,}500 - 10{,}000}{1{,}000} = -1.50$

De tablas $Z = 0.4332$

$P(X \langle Z) = 0.5000 - 0.4332 = 0.0668 \Rightarrow 6.7\%$



d) $$Z = \frac{X_i - \mu}{\sigma}$$

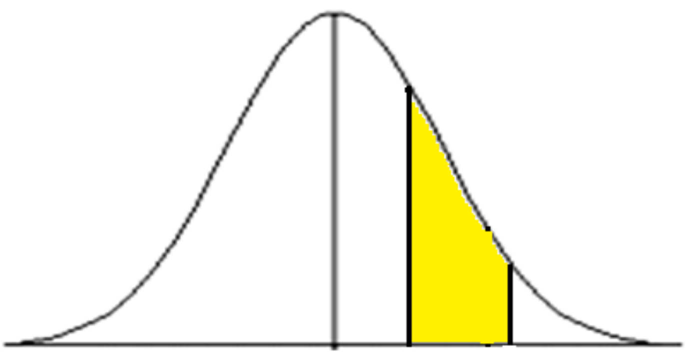
$$Z_1 = \frac{11{,}000 - 10{,}000}{1{,}000} = 1$$

De tablas $Z_1 = 0.3413$

$$Z_2 = \frac{12{,}000 - 10{,}000}{1{,}000} = 2$$

De tablas $Z_2 = 0.4772$

$$P(Z_1 \langle X \rangle Z_2) = 0.4772 – 0.3413 = 0.1359 \Rightarrow 13.59\%$$



PRINCIPALES APLICACIONES DE LA DISTRIBUCIÓN NORMAL

Selección de poblaciones

La distribución normal puede ser utilizada para realizar la selección de poblaciones, con fines de mejoramiento u otras aplicaciones.

Ejemplo

Las notas obtenidas en un examen fueron 17.6, 17.3, 17.2, 13.4, 16.7, 14.8, 13.4, 16.4, 15.1, 15.0, 17.7, 12.8, 15.2, 15, 12.3 ¿Cuál es la nota mínima para seleccionar el cuarto superior?

Calculando, se tiene:

μ, = 15.33, y

$\sigma = 1.73$

Despejando la probabilidad se tiene:

$P\ (X \geq a) = 0.25$

$$p = \left(Z \geq \frac{a-15.33}{1.73} \right) = 0.25$$

En la tabla se busca Z para una probabilidad de 0.25 y se obtiene:

$Z = 0.675$

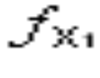
Por lo tanto:

$$Z = \frac{a-15.33}{1.73} \quad \Longrightarrow \quad Z \times 1.73 = a - 15.33$$

$a = Z \times 1.73 + 15.33$
$a = 0.675 \times 1.73 \div 15.33$
$a = 16.50$

Otra forma de resolver el mismo ejercicio

Es usando el programa Excel de la computadora, donde se ingresa al icono



se busca DISTR.NORM.INV y luego aparecerán tres líneas que deberán ser llenados de la manera siguiente:

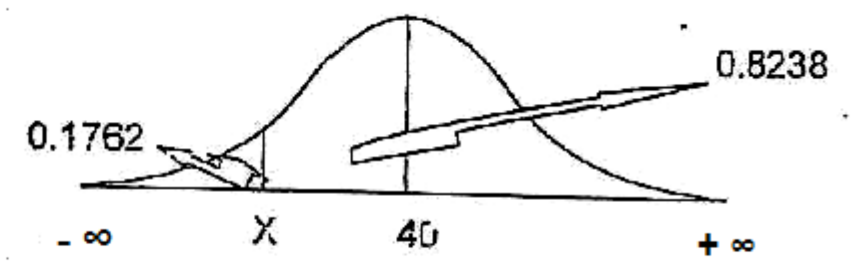
Probabilidad	0.75
Media	15.33
Desv_estandar	1.73

Acepta

Luego se y aparecerá el valor correspondiente a X, que en este caso es 16.50.
Por lo tanto para seleccionar el cuarto superior se debe hacer a partir de 16.5

Ejemplo

Si X ˜ N (40,36). ¿Cuál será el valor de X si se pretende seleccionar al 82.38% superior?



$P(X \geq a) = 0.8238$

La computadora nos da los valores acumulados de -∞ hasta el valor de X, por lo tanto la probabilidad queda expresada de la manera siguiente:

$P(X < a) = 1 - 0.8238 = 0.1762$

Con estos datos se recurre al Programa Excel de una computadora y se ingresa al **icono**

f_x

Donde se busca DISTR.NORM.INV y luego aparecerán tres líneas que deberán ser llenados de la manera siguiente:

Probabilidad	0.1762
Media	40
Desv_estandar	6

Acepta

Luego se y aparecerá el valor correspondiente a X, que en este caso es 34.4. Por lo tanto el valor de X que se pretende seleccionar a 82.38% superior es 34.4

Ejemplo

Si X ˜ N (37,14.44). Calcular:

a) P (X ≥ 33.2) = 1 – P (x < 33.2)

La computadora nos da los valores acumulados de -∞ hasta el valor de X, por lo tanto la probabilidad a obtener en la computadora es:

P (X < 33.2) =?

En este caso se procede casi de manera similar en la computadora, con la diferencia que en este caso se ingresa al **icono**

f_{X_1}

Donde se busca DISTR.NORM y luego aparecerán cuatro líneas que deberán ser llenados de la manera siguiente:

X	33.2
Media	37
Desv_estandar	3.8

| Verdadero | Acum |

Acepta

Luego se y aparecerá el valor correspondiente a la probabilidad, que en este caso es 0.1587. Por lo tanto el valor de la probabilidad es:

$P (X \geq 33.2) = 1 - 0.1587 = 0.8413$

b) $P (38.9 \leq X \leq 42.7) =$

Para calcular esta probabilidad se debe ejecutar el siguiente procedimiento:

$P (38.9 \leq X \leq 42.7) = P (X \leq 42.7) - P (38.9 \leq X) =$

En el icono f_{X_1} se busca DISTR.NORM, donde aparecerán cuatro líneas que deberán ser llenadas de la manera siguiente:

X	38.9
Media	37
Desv_estandar	3.8
Acum	Verdadero

Acepta

Luego se y aparecerá el valor correspondiente a la probabilidad, que en este caso es 0.6915

Nuevamente en el **icono** f_{X_1} se busca DISTR, NORM, donde aparecerán cuatro líneas que deberán ser llenados de la manera siguiente:

X	42.7
Media	37
Desv_estandar	3.8
Acum	Verdadero

Acepta

Luego se y aparecerá el valor correspondiente a la probabilidad, que en este caso es 0.9332

Con estos resultados se obtiene la probabilidad deseada que es:

P (X ≤ 42.7) – P (38.9 ≤ X) = 0.9332 – 0.6915

P (38.9 ≤ X ≤ 42.7) = 0.2417

Intervalos de confianza

Los intervalos espacios comprendidos entre dos límites, en el caso de la estimación de parámetros, son los límites entre los cuales están comprendidos los estimadores con una probabilidad determinada:

P (L_i ≤ μ ≤ L_s) = 1 -∞ Donde los valores de los límites son:

$$p(\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{\frac{1-\sigma}{2}} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{\frac{1-\sigma}{2}}) = 1 - \infty$$

Ejemplo

Se toma 49 tomates de un lote, obtenemos el promedio de 85 gr. Por antecedentes conocemos la varianza del lote de 10.24 gr. Calcular al 95% un rango.

Asumiendo que X ˜ N (μ, 3.2^2), se tiene que:

$$p(\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{\frac{1-\sigma}{2}} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{\frac{1-\sigma}{2}}) = 0.95$$

En la tabla de distribución Z se encuentra que $\mathbf{Z_{0.475}}$ = 1.96

$$p(85 - \frac{3.2}{\sqrt{49}} \times 1.96 \leq \mu \leq 85 + \frac{3.2}{\sqrt{49}} \times 1.96) = 0.95$$

P (84.1 ≤ μ ≤ 85.9) = 0.95

Esto indica que en una muestra de tomates el 95% tendrán pesos comprendidos entre 84.1 y 85.9 gr.

Dimensión muestra

Poblaciones infinitas:

$$n = \left(\frac{Z\sigma}{d}\right)^2$$

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

Poblaciones infinitas:

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

Ejemplo

Calcular cuántos tomates se deben tomar del ejercicio anterior del peso de los tomates, para un a = 0.05 y un error del 4% con respecto al promedio.

$Z_{0.475} = 1.96$
d = 4% del $\overline{X}$
d = 0.02 x 85 =3.4
$\sigma = 3.2$

$$n = \left(\frac{1.96 \text{ x } 3.2}{1.7}\right)^2$$

n = 13.6 = 14

El tamaño de muestra requerido es de 14 tomates

Ejemplo

Un organismo de desarrollo quiere instalar una casa de partos en una zona rural, obteniendo el número de partos mensuales a través de una muestra piloto en 10 unidades campesinas, con los siguientes resultados: 5, 3, 4, 3, 5, 4, 3, 4, 2, 5. Además se conoce por trabajos anteriores el valor de la varianza de la población de 1.061

a) Calcule el promedio y la desviación estándar

$$\sum X_i = 38 \qquad \sum X_i^2 = 154 \qquad n = 10$$

$$\overline{X} = \frac{38}{10} \qquad \overline{X} = 38$$

b) Calcula un intervalo de confianza del 99% para el promedio

$\sigma = 1.03$ $\qquad$ $Z_{0.495} = 2.575$

$$p(3.8 - \frac{1.03}{\sqrt{10}} \times 2.575 \leq \mu \leq 3.8 + \frac{1.03}{\sqrt{10}} \times 2.575) = 0.99$$

$$P\ (2.96 \leq \mu \leq 4.64) = 0.99$$

c) Calcular el tamaño de muestra, si se conoce que esa zona se tiene 75 comunidades campesinas, para un $a = 0.05$ y un error del 10% con respecto al promedio poblacional.

$Z_{0.475} = 1.96$
$d = 10\%$ del $\overline{X}$
$d = 0.10 \times 3.8 = 0.38$
$s = 1.03$

$$n = (\frac{1.96 \times 1.03}{0.38})^2 \qquad n = 28.2 = 28$$

El estudio deberá realizarse en 28 comunidades

DISTRIBUCIÓN T DE STUDENT

En la mayoría de los casos se desconoce la varianza de la población, por lo cual no es posible utilizar la distribución Z, pero gracias a W.S. Gosset, quien en el año 1908 dedujo una distribución que lo denomino " distribución t de Student", la misma que fue perfeccionada por R.A. Fisher en el año 1928.

Esta distribución revoluciono prácticamente la estadística de las pequeñas muestras. La formula es:

$$t = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$

Donde “s” es la desviación estándar de la muestra.

Esta distribución se ofrece en tablas, conocidas como las “tablas de t”, también se tiene en calculadoras y computadoras

Principales aplicaciones de la distribución “t de Student”

Intervalos de confianza

Cuando no se conoce el valor de la varianza poblacional es posible estimar los intervalos de confianza, utilizando la varianza muestral y la distribución de “t”, de la manera siguiente:

$$P(\bar{X} - \frac{s}{\sqrt{n}} t_{n-1\alpha} \leq \mu \leq \bar{X} + \frac{s}{\sqrt{n}} t_{n-1\alpha}) = 1 - \infty$$

Ejemplo

Conociendo que el rendimiento de gas domestico, en una muestra de 10 familias, es: 20, 23, 15, 17, 18, 19, 20, 18, 20 y 15 días/balón de gas mensual. Estimar los limites de confianza para la media de la población con un a = 0.05.

$n=10$ $X = 18.5$ $s^2 = 6.06$ $s = 2.46$ $t_{(9,\ 0.05)} = 2.262$

$$P(18.5 - \frac{2.46}{\sqrt{10}} 2.262 \leq \mu \leq 18.5 + \frac{2.46}{\sqrt{10}} 2.262) = 0.95$$

$P\ (16.74 \leq \mu \leq 20.26) = 0.95$

De una muestra de 100 balones de gas, 95 de estos balones tendrán una duración que oscila entre 16.74 y 20.26 días.

Intervalos de confianza para una proporción

$$s\bar{X} = \sqrt{\frac{s^2}{n}} \text{ por } \sqrt{\frac{pq}{n}}$$

$$p\left(p - \sqrt{\frac{pq}{n}}\, t_{n-1\alpha} \leq \mu \leq p + \sqrt{\frac{pq}{n}}\, t_{n-1\alpha}\right) = 1 - \infty$$

Ejemplo

En una muestra a 100 personas se les encuesta sobre un tipo de sangre rara, obteniéndose que 10 de ellos tengan este tipo de sangre. Calcular un margen al 90%.

$p = \frac{10}{100} = 0.1$ q = (1 – 0.1) = 0.9 $t_{99,\,0.1} = 1.66$

$$p\left(0.1 - \sqrt{\frac{0.1x0.9}{100}}\, x\, 1.66 \leq \mu \leq 0.1 + \sqrt{\frac{0.1x0.9}{100}}\, x\, 1.66\right) = 0.90$$

P (0.0502 ≤ P ≤ 0.1498) = 0.90

Tamaño de muestra

La varianza de promedios es:

$$s^2\bar{X} = \frac{s^2}{n}$$

Por lo tanto el tamaño de muestra, cuando no se conoce la varianza poblacional, se estima con la varianza de promedios de la muestra y la distribución de "t", de la manera siguiente:

$$t = \frac{X - \mu}{s / \sqrt{n}} \text{ ó } n\left(\frac{S.t_{n-1\alpha}}{\mathbf{d}}\right)^2$$

Ejemplo

En una empresa de producción de chompas de fibra de alpaca, se quiere seleccionar tres grupos, de acuerdo a su productividad diaria; para la cual se toma de los registros de la empresa el rendimiento promedio de chompas/día/obrero, con los siguientes resultados:

Calcular el tamaño de muestra para d = 5% $\bar{y}$a = 0.03

Promedio	D. Estándar	n – 1	$t_{34,0.03}$	D= 5%Prom.	N
6.0	1.35	34	2.27	0.30	106

Tamaño de muestra para proporciones

Para el cálculo del tamaño de muestra para proporciones se utiliza la misma fórmula del tamaño de muestra de promedios, donde se remplaza $\bar{X}$ por “p” y la varianza s^2 por “pq”, de tal manera que la formula queda de la manera siguiente:

$$n = \frac{pq\, t^2_{n-1\alpha}}{d^2}$$

4.3	6.5	5.2	3.8	7.1	8.0	5.7	6.4	7.3	5.9	6.1	4.6
5.1	6.9	5.4	6.5	7.3	8.4	4.6	3.6	8.2	5.7	7.7	6.4
3.5	5.7	7.3	6.9	5.7	7.0	5.3	3.7	5.1	6.9	4.6	

Ejemplo

En una encuesta sobre un tipo de sangre, se tomo una muestra a 100 personas, de estas, 10 tuvieron el tipo de sangre de interés. Calcular el tamaño de muestra para un a = 0.05 y d= 0.03.

$$p = \frac{10}{100} = 0.1$$

q = 0.9 $\quad$ $t_{99,\,0.05}$ = 1.98 $\quad$ d = 0.03

n = 3937 = 394

Ejemplo

En la Carrera Profesional de Ingeniería la proporción de estudiantes mujeres es de 0.25. Calcular el tamaño de muestra para a = 0.05 y una d= 0.10.

P = 25 $\quad$ p = 0.75 $\quad$ $Z_{0.05}$ = 1.98 $\quad$ d = 0.10

$$n = \frac{pqZ^2 \frac{1-\alpha}{2}}{d^2} = \frac{0.25 \times 0.75 \times 1.64^2}{0.10^2}$$

n = 50.7 = 51

DISTRIBUCIÓN DE CHI CUADRADO

5.4.1 Notación

X^2 = Chi cuadrado

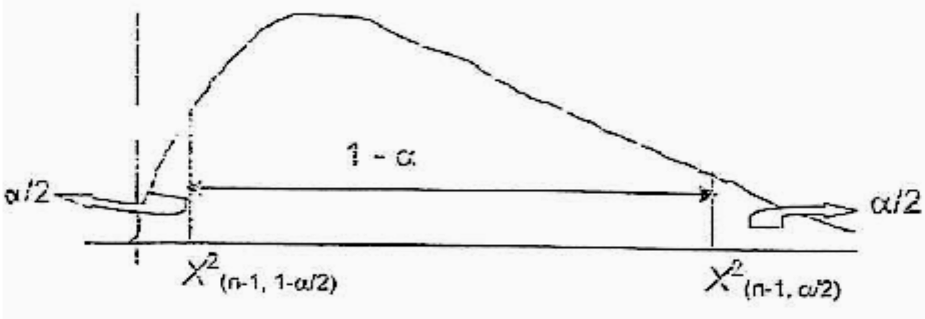
5.4.2 Definición

Si s^2 es un estimador insesgado de σ^2, entonces:

$$\frac{(n-1)}{\sigma^2} = X^2$$

La distribución de $X^2_{(n-1)}$ es una distribución asimétrica, donde:

- el límite inferior es $X^2_{(n-1,\ 1-\alpha/2)}$, y
- el límite superior es $X^2_{(n-1,\ \alpha/2)}$



Límites de confianza para σ^2

Una de las aplicaciones de la distribución de Chi Cuadrado es el cálculo de intervalo de confianza para la varianza poblacional, con la formula siguiente:

$$p\left(\frac{(n-1)s^2}{X^2_{n-1.\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)s^2}{X^2_{n-1.1-\frac{\alpha}{2}}}\right) = 1-\alpha$$

Ejemplo

En un Centro Educativo con 85 docentes, se toma una muestra de 12 profesores para revisar las faltas por diversos motivos durante el año 2004 y los resultados fueron los siguientes:

8	5	18	7	20
12	16	90	9	12
3	5	5	8	15
1	8	30	4	6

Calcular un intervalo de confianza para la varianza con una a = 0.10

$$p\left(\frac{19x367.15}{30.14} \leq \sigma^2 \leq \frac{19x367.15}{10.12}\right) = 0.90$$

P (231.42 ≤ σ^2 ≤ 689.51) = 0.90

Ejemplo

Conociendo que el rendimiento de gas doméstico, en una muestra de 10 familias, es 20, 23, 15, 17, 18, 19, 20, 18, 20 y 15 días / balón de gas. Estimar los límites de confianza para la varianza de la población con un a = 0.05.

$$\sum X_i^2 = 3477 \qquad \sum X_i = 185 \qquad n = 10$$

$$s^2 = \frac{3477 - 185^2 / 10}{9} = 6.06$$

$$p\left(\frac{9x6.06}{19.02} \leq \sigma^2 \leq \frac{9x6.06}{2.70}\right) = 0.95$$

P (2.86 ≤ σ^2 ≤ 20.18) = 0.90

Ejemplo

En el cuadrado de distribución de frecuencia se muestra el número de lectores de una revista de circulación nacional por edades.

Calcular:

a) La varianza

$$s^2 = \frac{\sum_{i=1}^{n} n_i X_i^2 - (\sum_{i=1}^{n} n_i X_i)^2 / n}{n-1} = \frac{554362.5012375 \ / 325}{324}$$

$$s^2 = 26$$

b) Intervalos de confianza para la varianza, para un a=0.05

Edad	Nº	X_0	x^2_i	$n_i x_i$	$n_i x^2_i$
10 – 19	59	14.5	210.25	855.5	12404.75
20 – 34	79	27.0	729	2133.0	57591.00
35 – 44	60	39.5	1560.25	2370.0	93615.00
45 – 54	54	49.5	2450.25	2673.0	132313.50
55 – 64	73	59.5	3540.25	4343.5	258438.25
Total	**325**			**12375.0**	**554362.50**

$$p\left(\frac{(n-1)s^2}{X^2_{n-1.\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)s^2}{X^2_{n-1.1-\frac{\alpha}{2}}}\right) = 1-\alpha$$

$$p\left(\frac{324x26}{376.84} \leq \sigma^2 \leq \frac{324x26}{276.03}\right) = 0.95$$

P (22.33 ≤ σ² ≤ 30.49) = 0.95

EJERCICIOS

1. En un salón de baile hay 100 personas, 25 personas usan lentes, calcular la probabilidad de que al extraer una muestra de 30 personas 6 de ellos usen lentes.

2. El 10% de las cajas de cereal de un lote de producción presentan algún defecto, si el departamento de control de calidad toma una muestra de 30 cajas. ¿Calcular la probabilidad de que existan 8 cajas de cereal con problemas?

3. Se lanza una moneda 10 veces y se desea hallar la probabilidad de obtener exactamente 4 caras.

4. Se realiza el control de calidad en un proceso productivo en cierto período del turno de trabajo, los cuales siguen una distribución de Poissón, con un promedio de 3 defectos por hora. Se quiere obtener la probabilidad de que transcurran 4 horas sin defectos.

5. Se está estudiando la cantidad de cromo en las emisiones de una empresa textil para lo cual se toman 20 muestras en el lapso de los tres turnos. Si tres de las muestras dan altos índices de cromo. ¿Cuál es la probabilidad que dos muestras estén contaminadas?

6. Una compañía de seguros sabe que el 0.06% fallece por ataque cardiacos ¿Cuál es la probabilidad de que la compañía tenga que pagar a sus 5,000 afiliados?
   a) A más de 3 afiliados
   b) Exactamente a 3 afiliados

CAPÍTULO VI

PRUEBAS DE HIPÓTESIS

CAPÍTULO VI

PRUEBAS DE HIPÓTESIS

Las pruebas de hipótesis son comparaciones que se hacen con los parámetros de una o más poblaciones

PARTES DE UNA PRUEBA DE HIPÓTESIS

Las pruebas de hipótesis tienen en su planteamiento dos partes.

a) Hipótesis nula (H_0)

Es la parte de la hipótesis que se asume como cierta.

b) Hipótesis alterna (H_1)

Es la parte de la hipótesis que se trata de probar.

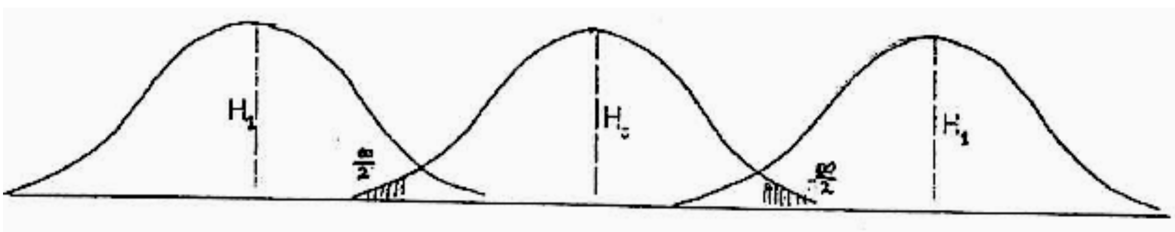
Debe anotarse que cuando se rechaza H_0, entonces se acepta H_1. Pero si no se rechaza H_0 esto no quiere decir que H_0 sea cierta; sino que solamente se puede decir que no hay la evidencia suficiente para rechazarse.

FORMA DE LAS HIPÓTESIS

De acuerdo al tipo de hipótesis que se quiera probar, estas pueden ser de dos o una cola.

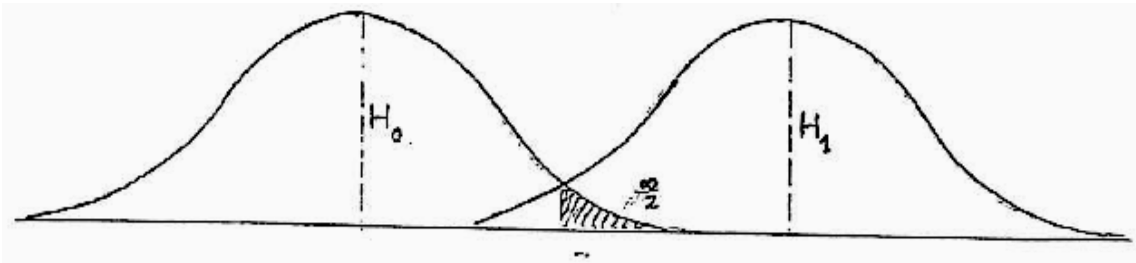
a) De dos colas

H_0: parámetro = a vs. H_1: parámetro = a

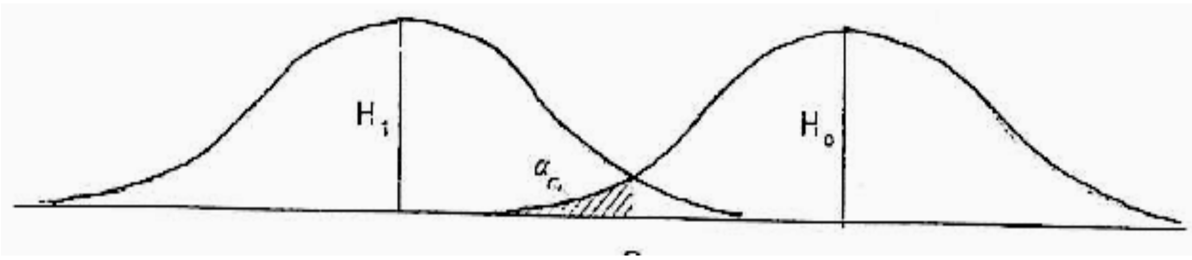


b) De una cola

H_0: parámetro ≤ a vs. H_1: parámetro > a



H_0: parámetro ≤ a vs. H_1: parámetro < a



DEFINICIONES PRELIMINARES

a) Error I

Es el error que consiste en rechazar H_0, cuando esta es cierta.

b) Error II

Es el error que consiste en aceptar H_0, cuando esta es falsa.

c) Potencia (1- β)

Probabilidad de rechazar H_0, cuando esta es cierta.

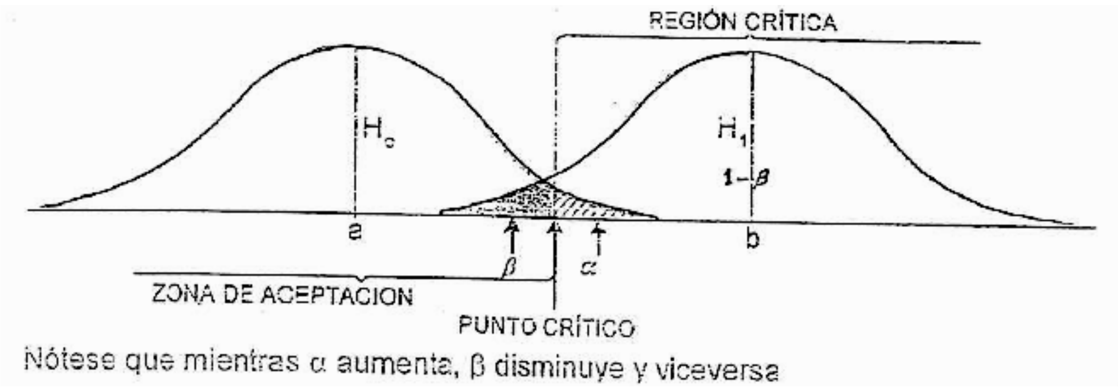
d) Región critica

Es el conjunto de valores que sustentan la Hipótesis alterna.

Todos estos conceptos pueden ser sintetizados en el cuadro siguiente:

SITUACION DE LA HIPOTESIS	**CIERTO**	**FALSO**
Se rechaza H_0	ERROR I	Decisión correcta
No se rechaza H_0	Decisión correcta	ERROR II

Gráficamente podría ser mostrado de la manera siguiente:



Nótese que mientras a aumenta, β disminuye y viceversa.

ETAPAS DE LA PRUEBA DE HIPÓTESIS

a) Plantear la H_o y H_1
b) Escoger el nivel de significación
c) Seleccionar el estadístico adecuado su distribución: así como el punto y región crítica
d) Se toma los datos para calcular el estadístico correspondiente y luego comparar con el punto crítico.
e) Conclusión, interpretando los resultados en función de la hipótesis planteada.

Ejemplo

Se señala que el peso de los chipis es de 10 oz; en una muestra al azar de 18 paquetes se obtiene los siguientes resultados: $\overline{X} = 9.73$ oz, S^2 = 0.18. Probar al nivel del 5%, si el anuncio es cierto o si en realidad los paquetes de chips pesan menos de 10 oz

a) Ho: $\upsilon \geq 10$

$H_1 : u < 10$

b) A = 0.05

c) Estadístico es $\overline{X}$ N(u, $\frac{S}{n}$) la región critica si $t_c < (-t_{17.0.010} =$ -1.74). se rechaza H_o

d) $t = \dfrac{9.73 - 10}{\sqrt{\dfrac{0.18}{18}}} = -2.7$

e) Conclusión

(T_c = -2.7) < ($-t_{17.0.010}$ = -1.74), por lo tanto se rechaza Ho lo que nos permite aceptar H_1; es decir que el peso promedio de los paquetes de chips es menor a las 10 oz

Ejemplo

En una empresa agraria previo a la ejecución de un trabajo de investigación sobre peso de vellón se decide tomar una muestra piloto, obteniendo los siguientes resultados.

3.6	4.7	4.0	3.5	3.0	3.9	3.7	3.6	4.0	3.0	3.2	3.1	3.7	3.6	4.3
4.4	4.7	4.3	3.7	3.8	3.7	3.5	3.9	3.4	4.0	4.3	4.1	3.2	3.8	3.4

Calcular

a) El tamaño demuestra para a=0.05

$$n = \left[\frac{S.t\ n - L\alpha}{N}\right]^2$$

$\sum_{i=1}^{n} X = 113.1$ n = 30

$\sum_{i=1}^{n} X1^2 = 432.47$ $t_{29,\ 0.05} = 2.045$

$X = \frac{113.1}{30} = 3.77$ d = 5% $\bar{X}$ d = 0.1885

$$S^2 \frac{432.47 - 113.1^2/30}{29} = 0.2098$$

s = 0.457994127

$$n = \left[\frac{0.4580\ x\ 2.045}{0.1885}\right]^2 = 24.7 \approx 25$$

b) Un intervalo de confianza del 99% para el promedio

$$P\left[x - \frac{s}{\sqrt{n}}\, t\, n-1,\alpha \leq \mu \leq x - \frac{s}{\sqrt{n}}\, t\, n-1,\alpha\right] = 1 - \infty$$

$t_{29.0.01}$ = 2.756

$$P\left[3.77 - \frac{0.4580}{\sqrt{30}} x 2.756 \leq \mu \leq 3.77 + \frac{S0.4580}{\sqrt{30}}\, x\, 2.756\right] = 0.99$$

$$P\,(3.54 \leq \mu \leq 4.00) = 0.99$$

c) Probar la hipótesis de que el promedio es menor que 3.8 Kg., para un a=0.01

1) H_0: μ ≥ 3.8
   H_1: μ < 3.8

2) .a=0.01

3) El estadístico es X ~ N (μ, $\frac{\sigma^2}{n}$). La región critica si t_c < (-t29.0.01 = -2.756), se rechaza H_o

4) $.t = \frac{3.77-3.8}{0.0836} = -0.36$

5) Conclusión

   (t_c = -0.36) > (-t 29.0.01 = -2.756); por lo tanto no se rechaza H_0, es decir que el peso promedio de vellón es mayor o igual que 3.8 Kg.

Ejemplo

En una empresa de producción, se conoce que el peso promedio de un producto es de 25.3 Kg, con una varianza de 3.3. Con la finalidad de mejorar el peso de la producción se contrata a un técnico con alta calificación y luego de un tiempo se toma una muestra de los pesos; con los siguientes resultados: 25, 27, 29, 32, 25, 24, 26, 29, 28, 27, 28, 30, 31, 34, 32, 31 y 33. ¿Se justifico la presencia del técnico?

1) H_0: $\mu \leq 25.3$
   H_1: $\mu > 25.3$

2) $\alpha = 0.05$

3) El estadístico es $X \sim N(\mu,\alpha^2)$ La región critica si $Z_c > (Z_{0.45} = 1.645)$; se rechaza H_0

4) $$Zc = \frac{28.2-25.3}{\sqrt{3.3/17}} = 8.1$$

5) Conclusión.

   $(Z_c = 8.1) > (Z_{0.475} = 1.645)$; por lo tanto se rechaza H_0, es decir que se justifico la presencia del técnico.

CHI CUADRADO EN FORMA MANUAL

Ejemplo

El servicio nacional de salud desea verificar si a distribución proporcional del estado nutricional de los niños no varía en tres ciudades de la región para lo cual toma una muestra de niños de las tres ciudades y los clasifica según estado nutricional obteniendo la siguiente tabla:

		CIUDAD		
		LIMA	**CAJAMARCA**	**AREQUIPA**
ESTADO NUTRICIONAL	**OBESO**	**82**	**70**	**62**
	SOBREPESO	**93**	**62**	**67**
	NORMAL	**25**	**18**	**21**
	FLACO	**16**	**15**	**18**

a)

H_0 : El Estado Nutricional depende de las Ciudades en las que habita.

H_1 : El Estado Nutricional no depende de las Ciudades en las que habita

b) Grados Libertad = (Nro Filas -1)(Nro Columnas -1)

Grados Libertad = (4 – 1)(3 – 1) = (3) (2) = 6

c) De Tablas **$X^2_{(0.05,\,6)}$** = 12.592

d) Regla de decisión:

Si **X^2_R ≤ 12.592** no se rechaza **H_0**

Si **X^2_R > 12.592** se rechaza **H_0**

e) El Coeficiente Chi cuadro en forma manual se calcula con la siguiente formula:

f) $$X^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

		CIUDAD		
		LIMA	CAJAMARCA	AREQUIPA
ESTADO NUTRICIONAL	OBESO	82	70	62
	SOBREPESO	93	62	67
	NORMAL	25	18	21
	FLACO	16	15	18

Se procederá a calcular los valores esperados

Los valores esperados se calculan así:

Ejemplo para el valor observado 80, el valor esperado será:

$$V.E. = \frac{\left(Vtotalfila * VtotalColumna\right)}{VTotalG}$$

$$V.E. = \frac{\left(214 * 216\right)}{549} = 84.20$$

Y así se calcularan los demás valores esperados, tal como se muestra en la tabla siguiente:

		CIUDAD			TOTAL
		LIMA	CAJAMARCA	AREQUIPA	
ESTADO NUTRICIONAL	OBESO	82 (84.20)	70 (64.32)	62 (65.49)	214
	SOBREPESO	93 (87.34)	62 (66.72)	67 (67.93)	222
	NORMAL	25 (25.18)	18 (19.23)	21 (19.58)	64
	FLACO	16 (19.28)	15 (14.73)	18 (14.99)	49
TOTAL		216	165	168	549

Luego aplicando la fórmula para calcular Chi cuadrado, obtenemos la siguiente tabla:

O	E	O-E	((O-E)^2)/E
82	84.20	-2.20	0.06
93	87.34	5.66	0.37
25	25.18	-0.18	0.00
16	19.28	-3.28	0.56
70	64.32	5.68	0.50
62	66.72	-4.72	0.33
18	19.23	-1.23	0.08
15	14.73	0.27	0.01
62	65.49	-3.49	0.19
67	67.93	-0.93	0.01
21	19.58	1.42	0.10
18	14.99	3.01	0.60
		TOTAL	2.81

Entonces el Chi Cuadrado es de **2.81**

Decisión y justificación:

Como el valor de 2.81 es menor que el de la Tabla 12.592, **por lo tanto no se rechaza H_0, y se concluye que el Estado Nutricional no depende de la Ciudad de cada habitante.**

APLICACIÓN DEL SOFTWARE ESTADISTICO SPSS PARA OBTENER EL VALOR DE CHI CUADRADO

Ejemplo

El servicio nacional de salud desea verificar si a distribución proporcional del estado nutricional de los niños no varía en tres ciudades de la región para lo cual toma una muestra de niños de las tres ciudades y los clasifica según estado nutricional obteniendo la siguiente tabla:

		CIUDAD		
		LIMA	CAJAMARCA	AREQUIPA
ESTADO NUTRICIONAL	OBESO	82	70	62
	SOBREPESO	93	62	67
	NORMAL	25	18	21
	FLACO	16	15	18

a)

H_0 : El Estado Nutricional depende de las Ciudades en las que habita.

H_1 : El Estado Nutricional no depende de las Ciudades en las que habita

b) Grados Libertad = (Nro Filas -1)(Nro Columnas -1)

Grados Libertad = (4 – 1)(3 – 1) = (3) (2) = 6

c) De Tablas $X^2_{(0.05,\ 6)} = 12.592$

d) Regla de decisión:

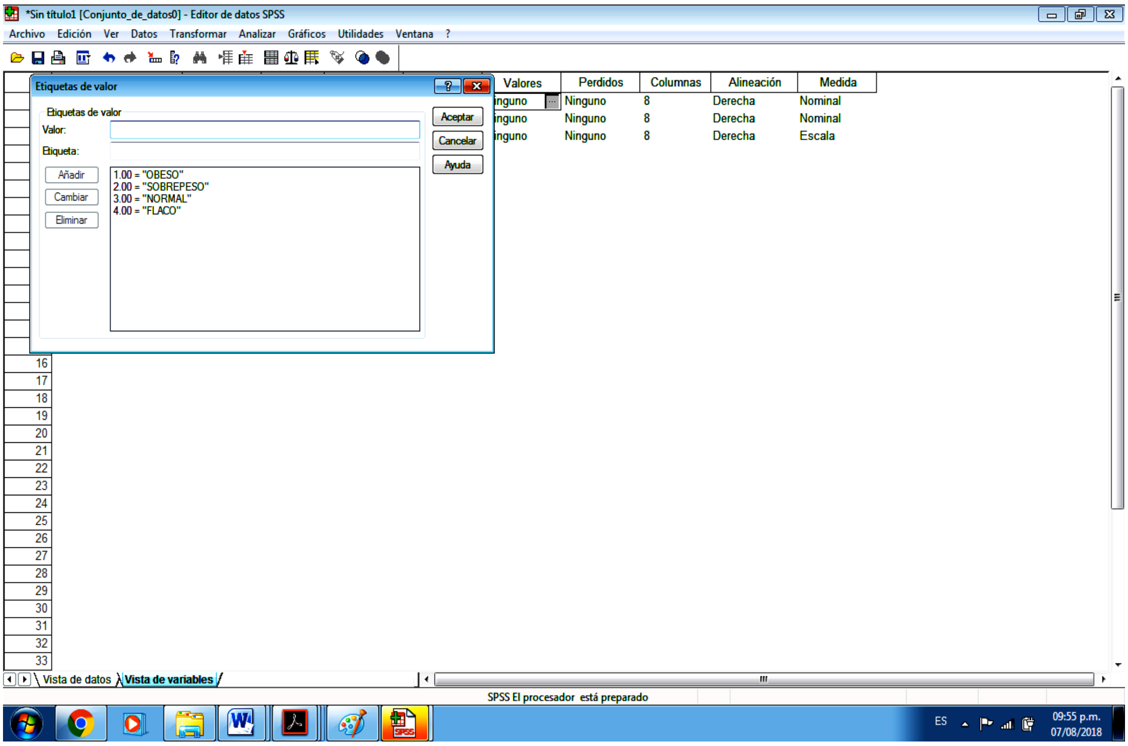
Si $X^2_R \leq 12.592$ no se rechaza **H_0**

Si $X^2_R > 12.592$ se rechaza **H_0**

Primeramente vaciaremos los datos al Programa Estadístico SPSS

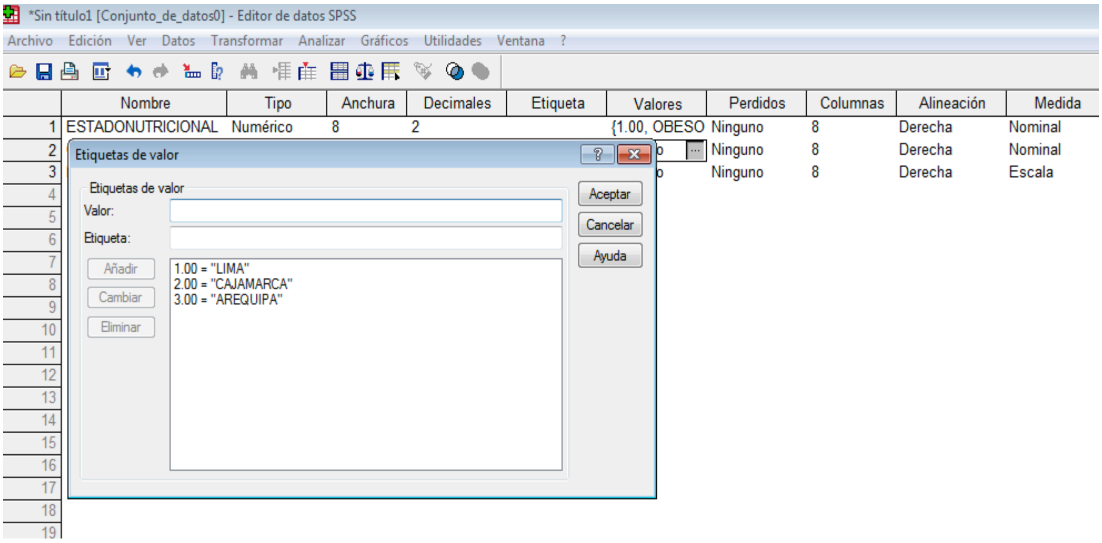
Ir a vista de variables, escribir Estado Nutricional, Ciudades y Frecuencia

Hacer clic en valores para Estado Nutricional



Aceptar

Hacer clic en valores para Ciudades



Aceptar

a) Luego ir a vistas de datos

b) En la Variable Estado Nutricional, colocar:

   1 (Obeso) repetido tres veces por cada ciudad.

   2 (Sobrepeso) repetido tres veces por cada ciudad.

   4 (Normal) repetido tres veces por cada ciudad.

   5 (Flaco) repetido tres veces por cada ciudad.

c) En la Variable Ciudad, colocar:

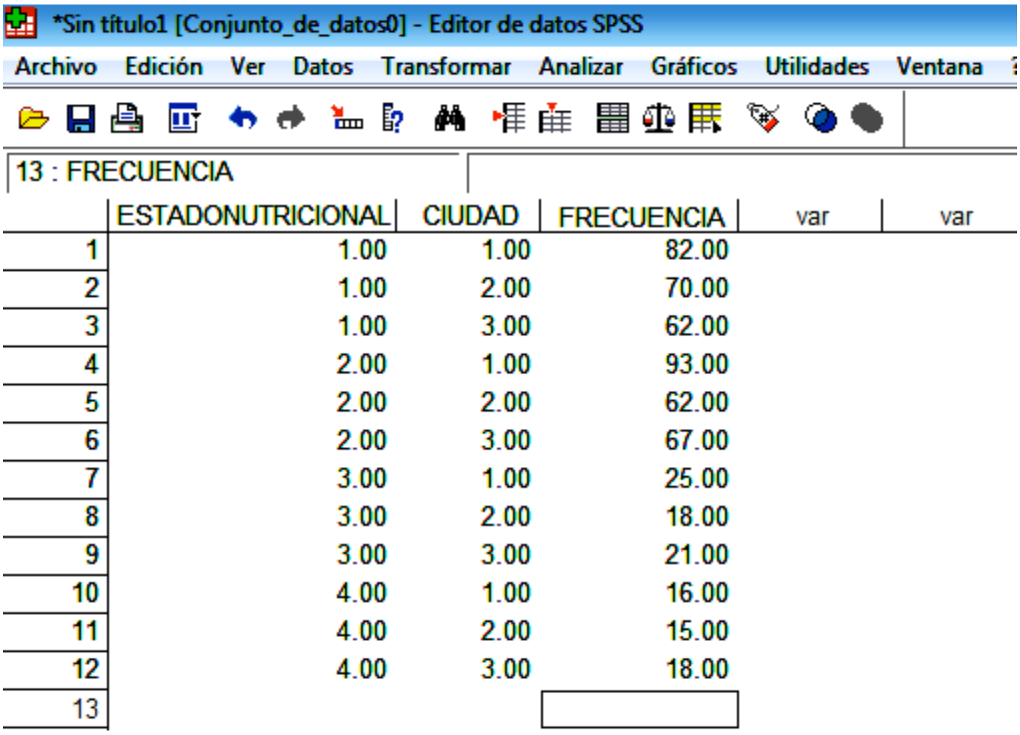
   1 , 2, 3 por cada categoria de Estado Nutricional (Obeso).

   1 , 2, 3 por cada categoria de Estado Nutricional (Sobrepeso).

   1 , 2, 3 por cada categoria de Estado Nutricional (Normal).

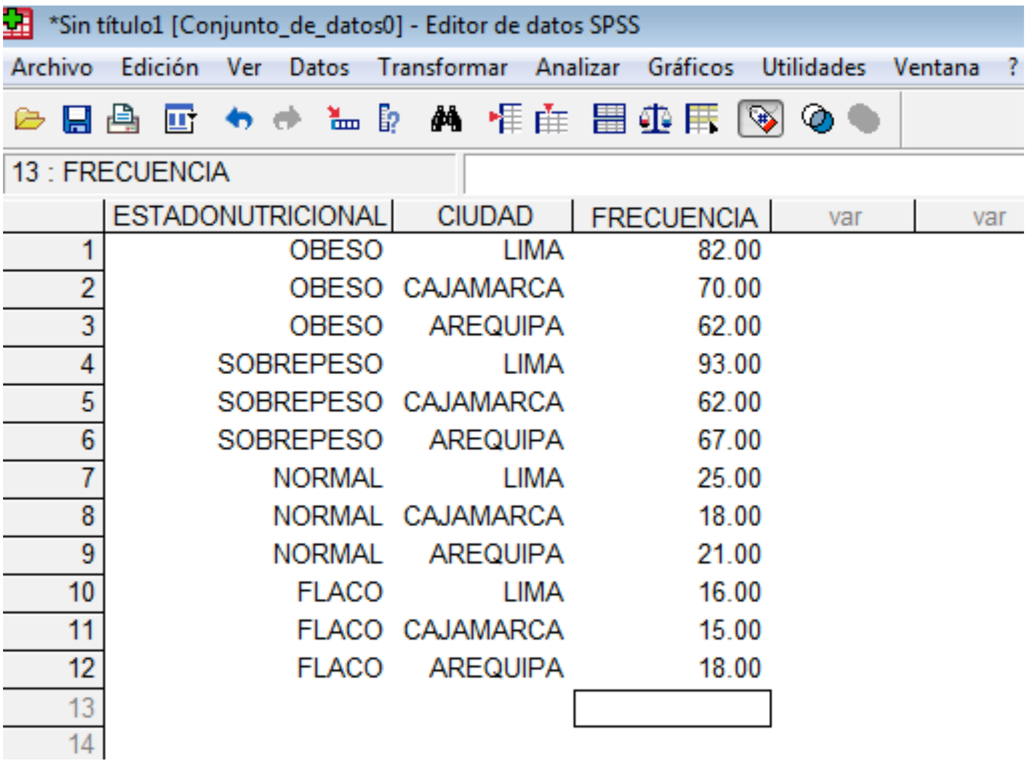
   1 , 2, 3 por cada categoria de Estado Nutricional (Flaco).

d) En la Variable Frecuencia, colocar los valores númericos fila por fila, comenzando en 82 …….18



	ESTADONUTRICIONAL	CIUDAD	FRECUENCIA	var	var
1	1.00	1.00	82.00		
2	1.00	2.00	70.00		
3	1.00	3.00	62.00		
4	2.00	1.00	93.00		
5	2.00	2.00	62.00		
6	2.00	3.00	67.00		
7	3.00	1.00	25.00		
8	3.00	2.00	18.00		
9	3.00	3.00	21.00		
10	4.00	1.00	16.00		
11	4.00	2.00	15.00		
12	4.00	3.00	18.00		
13					

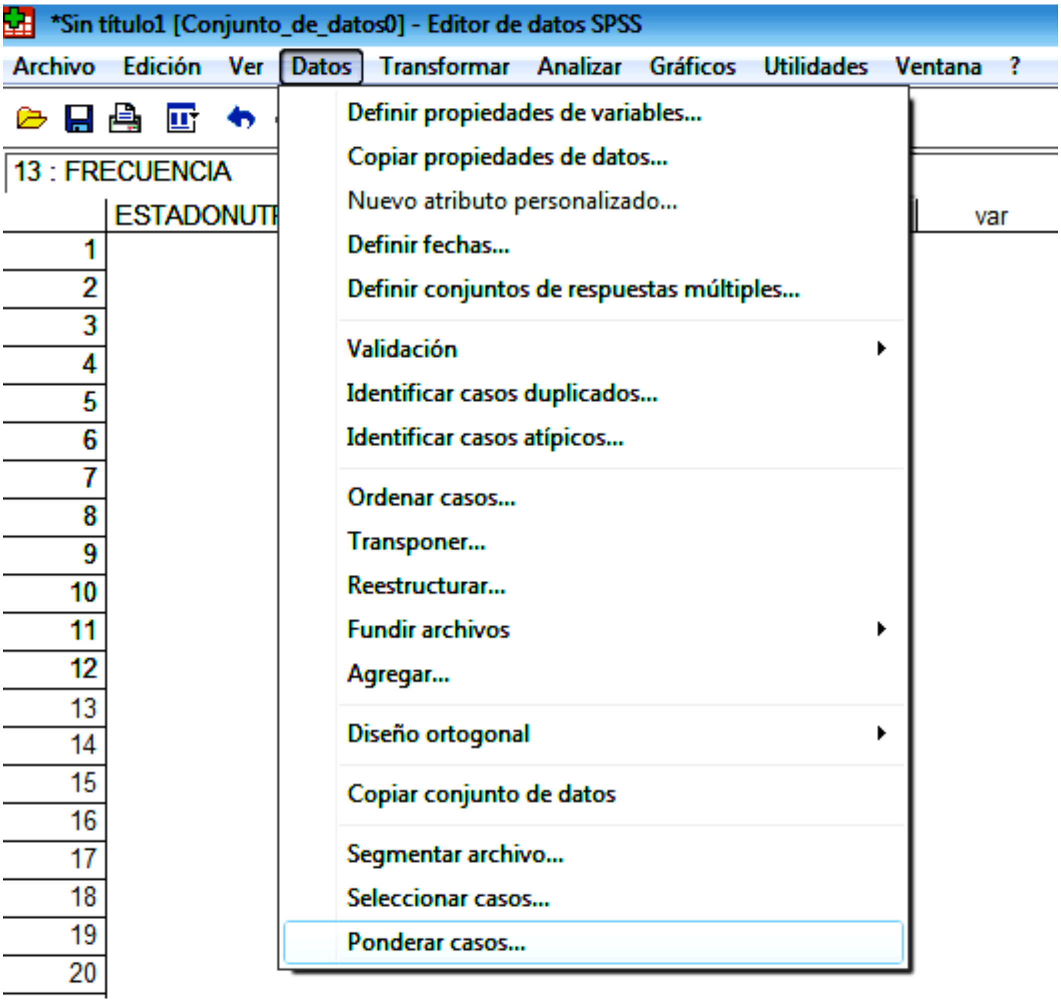
Posesionandose en cualquier lugar de las variables, hacer clic k en la banderita y aparecera



	ESTADONUTRICIONAL	CIUDAD	FRECUENCIA	var	var
1	OBESO	LIMA	82.00		
2	OBESO	CAJAMARCA	70.00		
3	OBESO	AREQUIPA	62.00		
4	SOBREPESO	LIMA	93.00		
5	SOBREPESO	CAJAMARCA	62.00		
6	SOBREPESO	AREQUIPA	67.00		
7	NORMAL	LIMA	25.00		
8	NORMAL	CAJAMARCA	18.00		
9	NORMAL	AREQUIPA	21.00		
10	FLACO	LIMA	16.00		
11	FLACO	CAJAMARCA	15.00		
12	FLACO	AREQUIPA	18.00		
13					
14					

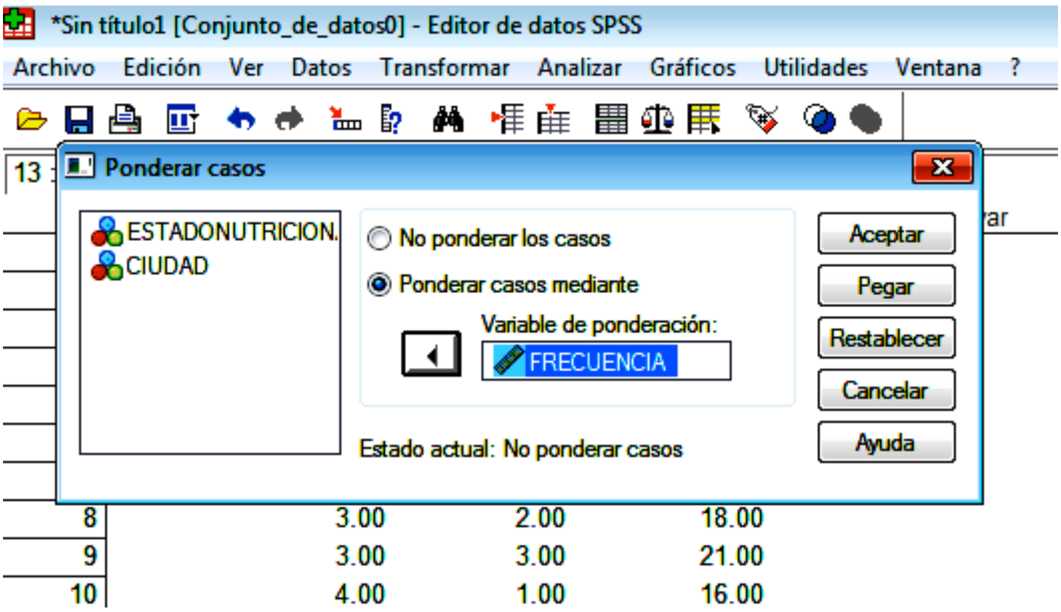
Dar clic a la banderita nuevamente y volverá al cuadro anterior.

Continuando con el proceso ir a Datos – Ponderar casos



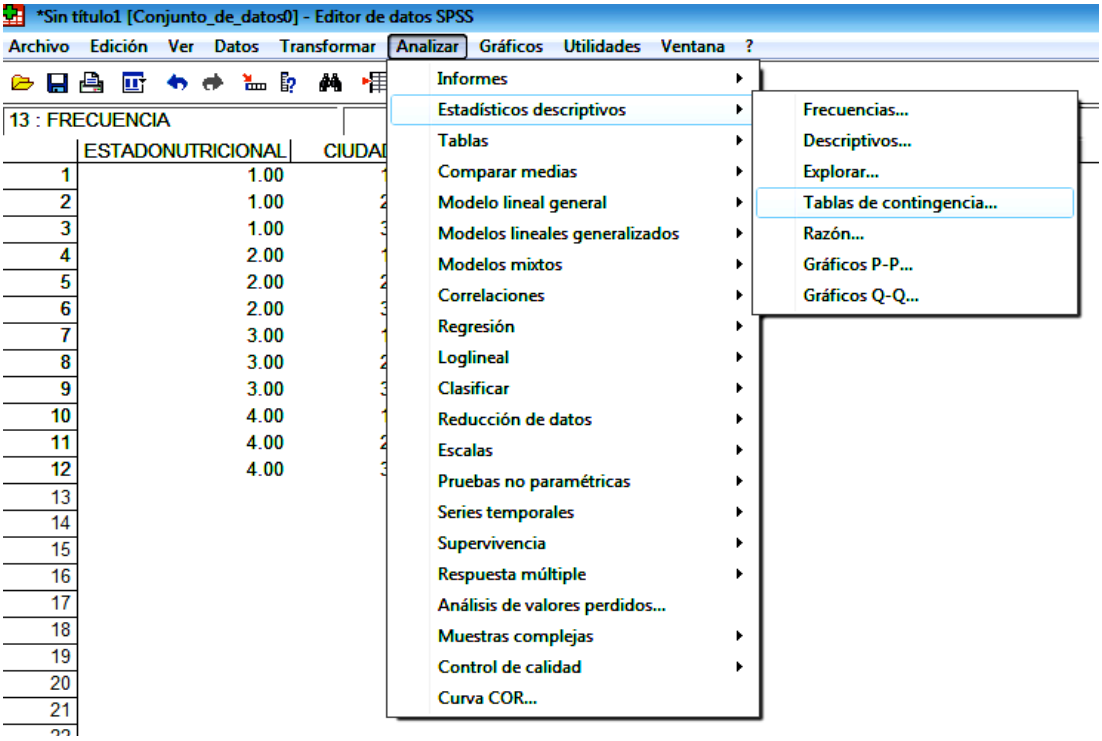
Clic en ponderar casos.

Hacer check en ponderar casos, con la variable frecuencia



Aceptar

Analizar – Estadísticos descriptivos – Tablas de contingencia

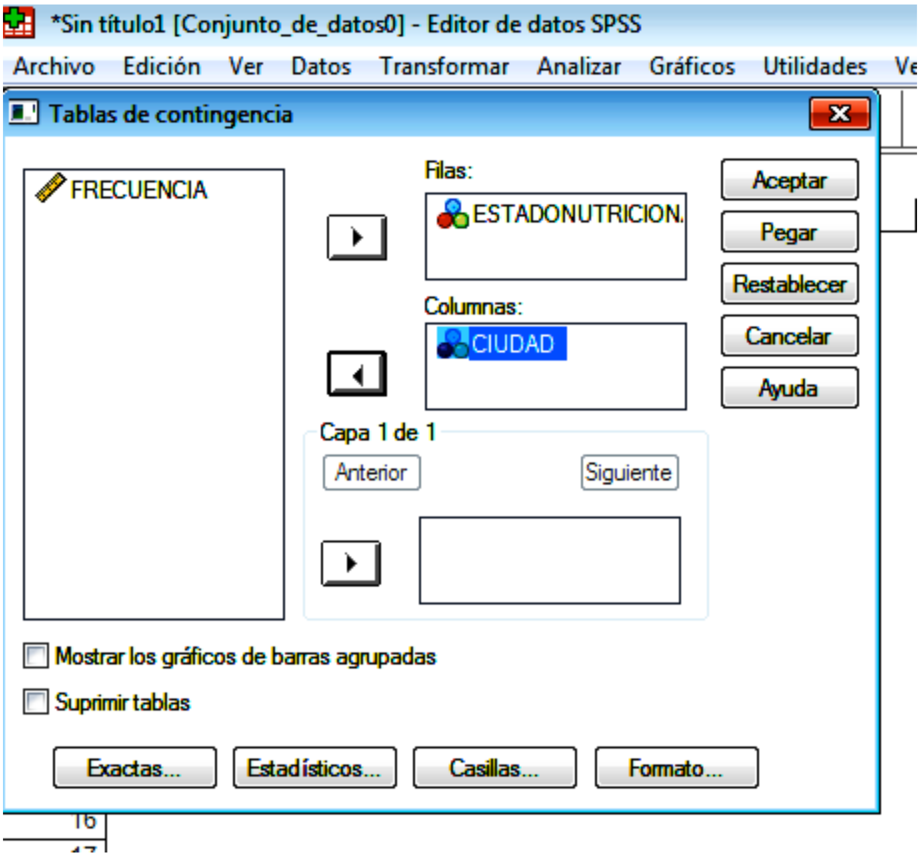


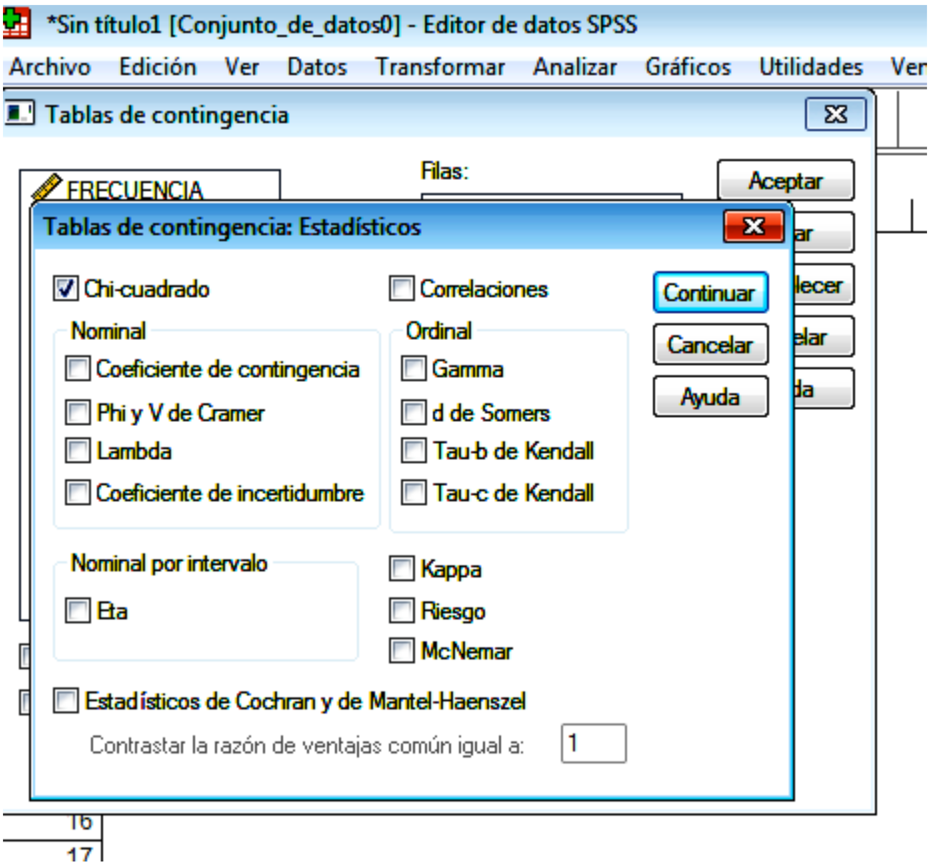
Hacer clic

Luego pasar a las Filas Satisfacción en el trabajo y a la Columna el Rango

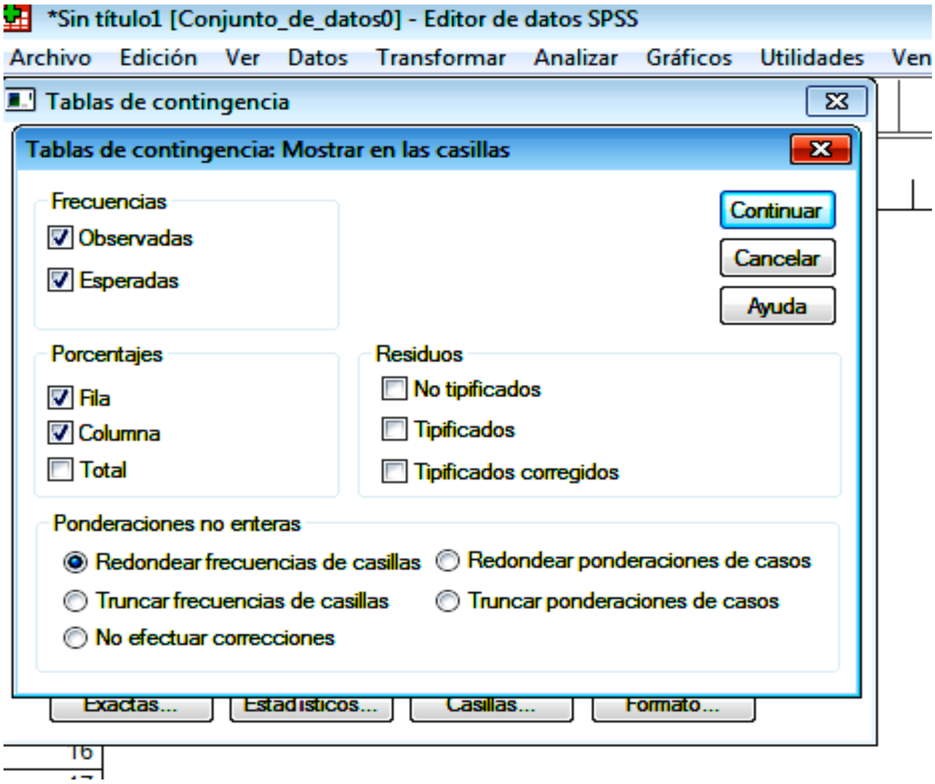
Clic en estadísticos, hacer check en Chi Cuadrado

Clic en Casillas y check en frecuencias esperadas,% en fila y columna

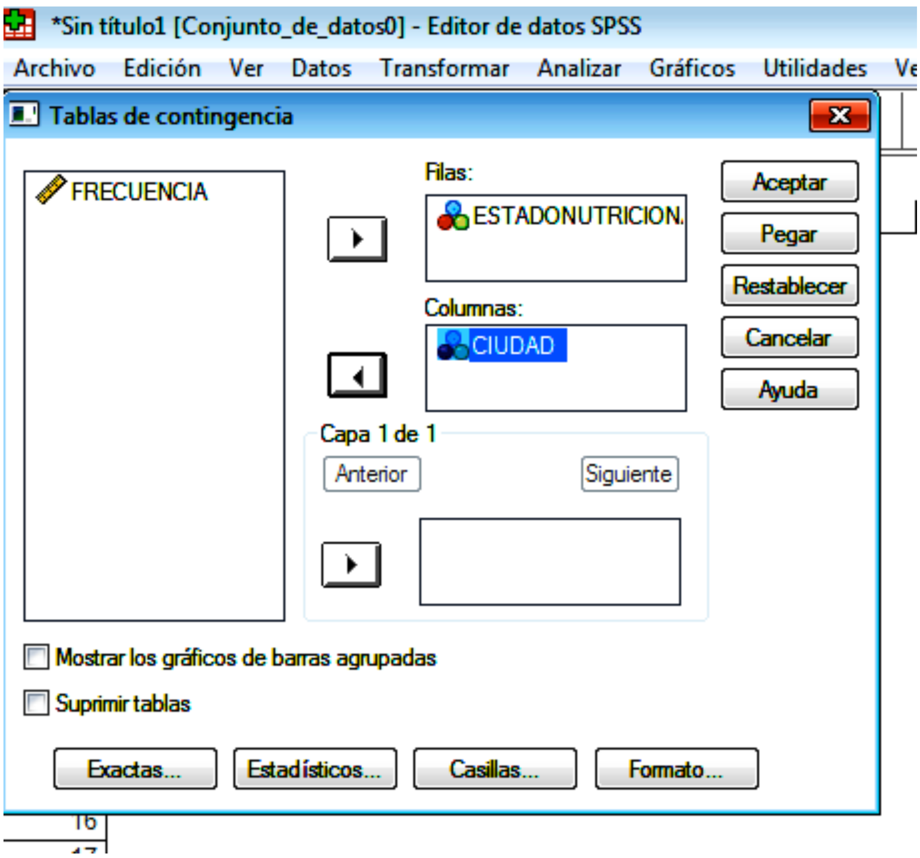

*Sin título1 [Conjunto_de_datos0] - Editor de datos SPSS
Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades
Tablas de contingencia
FRECUENCIA
Filas:
ESTADONUTRICION
Columnas:
CIUDAD
Capa 1 de 1
Anterior
Siguiente
Aceptar
Pegar
Restablecer
Cancelar
Ayuda
Mostrar los gráficos de barras agrupadas
Suprimir tablas
Exactas...
Estadísticos...
Casillas...
Formato...


*Sin título1 [Conjunto_de_datos0] - Editor de datos SPSS
Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades
Tablas de contingencia
Tablas de contingencia: Estadísticos
Chi-cuadrado
Correlaciones
Continuar
Cancelar
Ayuda
Nominal
Coeficiente de contingencia
Phi y V de Cramer
Lambda
Coeficiente de incertidumbre
Ordinal
Gamma
d de Somers
Tau-b de Kendall
Tau-c de Kendall
Nominal por intervalo
Eta
Kappa
Riesgo
McNemar
Estadísticos de Cochran y de Mantel-Haenszel
Contrastar la razón de ventajas común igual a:
1

Continuar



Contnuar



Aceptar

Tablas de contingencia

[Conjunto_de_datos0]

Resumen del procesamiento de los casos

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
ESTADONUTRICIONAL * CIUDAD	549	100.0%	0	.0%	549	100.0%

Tabla de contingencia ESTADONUTRICIONAL * CIUDAD

			CIUDAD			Total
			LIMA	CAJAMARCA	AREQUIPA	
ESTADONUTRICIONAL	OBESO	Recuento	82	70	62	214
		Frecuencia esperada	84.2	64.3	65.5	214.0
		% de ESTADONUTRICIONAL	38.3%	32.7%	29.0%	100.0%
		% de CIUDAD	38.0%	42.4%	36.9%	39.0%
	SOBREPESO	Recuento	93	62	67	222
		Frecuencia esperada	87.3	66.7	67.9	222.0
		% de ESTADONUTRICIONAL	41.9%	27.9%	30.2%	100.0%
		% de CIUDAD	43.1%	37.6%	39.9%	40.4%
	NORMAL	Recuento	25	18	21	64
		Frecuencia esperada	25.2	19.2	19.6	64.0
		% de ESTADONUTRICIONAL	39.1%	28.1%	32.8%	100.0%
		% de CIUDAD	11.6%	10.9%	12.5%	11.7%
	FLACO	Recuento	16	15	18	49
		Frecuencia esperada	19.3	14.7	15.0	49.0
		% de ESTADONUTRICIONAL	32.7%	30.6%	36.7%	100.0%
		% de CIUDAD	7.4%	9.1%	10.7%	8.9%
Total		Recuento	216	165	168	549
		Frecuencia esperada	216.0	165.0	168.0	549.0
		% de ESTADONUTRICIONAL	39.3%	30.1%	30.6%	100.0%
		% de CIUDAD	100.0%	100.0%	100.0%	100.0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	2.806[a]	6	.833
Razón de verosimilitudes	2.794	6	.834
Asociación lineal por lineal	.735	1	.391
N de casos válidos	549		

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 14.73.

Decisión y justificación:

Como el valor de 2.806 es menor que el de la Tabla 12.592, **por lo tanto no se rechaza H_0, y se concluye que el Estado Nutricional no depende de la Ciudad de cada habitante.**

COMPARACIÓN DE DOS POBLACIONES.

Hasta ahora era solo ha sido de nuestro interés la probabilidad, intervalos de confianza y pruebas de hipótesis para los parámetros de una sola población. Ahora nos ocuparemos de comparar el comportamiento de una misma característica en dos poblaciones o el comportamiento de una misa característica en dos poblaciones o el comportamiento de dos grupos de una misma población frente a la aplicación de dos tratamientos. Pero antes de señalar las comparaciones es necesario conocer la distribución de las variables en forma conjunta.

$X \sim N(\mu x, \sigma^2 x)$ y $Y \sim N(\mu y, \sigma^2 y)$, donde X y Y son independientes, entonces de acuerdo al teorema del límite central, se tiene que:

$X - Y \sim N(\mu x, \mu y, \sigma^2 x + \sigma^2 y)$;

Sin embargo, como:

$$\bar{X} \sim N\left(\mu x, \frac{\sigma^2 x}{n_x}\right) \text{ y } \bar{Y} \sim N\left(\mu x, \frac{\sigma^2 Y}{n_Y}\right)$$

Entonces:

$$\bar{X} - \bar{Y} \sim N\left(\mu x - \mu y, \frac{\sigma^2 x}{n_x} + \frac{\sigma^2 x}{n_Y}\right)$$

Conociendo la distribución, es fácil calcular la probabilidad, intervalo de confianza o probar una hipótesis. A continuación trataremos de verificar esta afirmación, a través de algunos ejemplos.

Ejemplo

Si $X \sim N(3, 9)$ y $Y \sim N(4, 8)$. Calcular la probabilidad de que un valor X sea mayor que un valor Y.

Por lo tanto: $(X - Y) \sim N(-1, 17)$

$$P(X > Y) = P(X - Y > 0) = P\left(Z > \frac{1}{\sqrt{17}}\right) = P(Z > 0.24) = 1 - 0.5948 = 0.4052$$

Ejemplo

Con los datos anteriores, si se toman muestras de tamaño 15 y 10 respectivamente. ¿Cuál es la probabilidad de que $\overline{X}$ sea $\overline{Y}$ mayor ?

SI X~N (3,9) y Y~N (4, 8); entonces:

$$(\bar{X} - \bar{Y}) \sim N\left(-1, \frac{9}{15} + \frac{8}{10}\right) \text{ ó } (\bar{X} - \bar{Y}) \sim N\left(-1, \frac{7}{5}\right)$$

$$= P\,(X - Y > 0) = P\left(Z > \frac{1}{\sqrt{\frac{7}{5}}}\right) = P\left(Z > \sqrt{\frac{7}{5}}\right) =$$

P (Z > 0.85) = 1 – 0.8023 = 0.1977

Ejemplo

Se supone que en las muestras del ejemplo 2. $\overline{X}$ = 3.5 y $\overline{Y}$ = 4.2. Estimar un intervalo de confianza del 95 para µx - µy

$$\left(\bar{X} - \bar{Y} - t_{nx+ny-2}\sqrt{\frac{\sigma^2 X}{n_X} + \frac{\sigma^2 Y}{n_Y}} \leq \mu x - \mu y \leq \bar{X} - \bar{Y} - t_{nx+ny-2}\sqrt{\frac{\sigma^2 X}{n_X} + \frac{\sigma^2 Y}{n_Y}}\right) = 1-\alpha$$

$$P\left((3.5\text{-}4.2) - \sqrt{\frac{7}{5}}\,x1.96 \leq \mu x - \mu y \leq (3.5\text{-}4.2) + \sqrt{\frac{7}{5}}\,x1.96\right) = 0.95$$

P (-3.019 ≤ µx - µy ≤ 1.619) = 0.95

Ejemplo

Si X˜N (µx, 9) y Y˜N (µy, 8); se toma una muestra de 15 para la primera población y una muestra de tamaño 10 para segunda, donde los promedios son $\overline{X}$ = 3.5 y Y = 4.2. Probar que µx ≠ µy.

a) Ho: µx = µy
   H1: µx ≠ µy

b) a = 0.05

c) El estadístico es : $\bar{X} - \bar{Y} \sim N\left(\mu x - \mu y, \frac{\sigma^2 x}{n x} + \frac{\sigma^2 Y}{n_Y}\right)$

Punto crítico: $Z_{0.975} = 1.96$

Región critica: Si $|Z_c| > (Z_{0.975} = 1.96)$, se rechaza H_0

d)

$$Z = \frac{(3.5 - 4.2) - 0}{\sqrt{\frac{9}{15} + \frac{8}{10}}} = 0.5916$$

e) Conclusión.

$|Z_c| = 0.59 < (Z_{0.975} = 1.96)$,); por lo tanto no se rechaza H_0, es decir que estadísticamente no existe diferencia entre estos dos promedios.

EJERCICIOS

1. El peso de las cajas de cereal tiene una distribución normal con un peso promedio de 20 Kg. y una varianza de 4 kg^2. Si se saca al azar una caja de cereal. ¿Cuál es la probabilidad de que su peso sea de 18 kg.?

2. La salida de las combis de transporte público tiene una distribución normal con un promedio de 20 minutos y una varianza de 5 $minutos^2$. Hallar la probabilidad que la salida de las combis de transporte este entre 30 y 40 minutos

3. Ciertos albañiles de construcción hacen una pared en promedio de 20 días y una varianza de 9 $días^2$, Si se elige al azar un albañil. ¿Cuál es la probabilidad de que haga la pared en 14 días?

4. Las ventas de cuatro tiendas de repuestos mecánicos, tienen una distribución normal con un promedio de 50,000 dólares mensuales y una varianza de 16,000 $dólares^2$. Si se elige al azar una tienda, hallar la probabilidad que sus ventas sean por lo menos de 45,000 dólares

5. El tiempo promedio de parada por mantenimiento de los tractores de un fundo es de 120 horas y un desviación estándar de 25 horas. Si el tractor Nro 3 se descompone, hallar la probabilidad de que el tiempo de parada este entre 60 y 90 días.

6. Use la tabla Z para calcular las siguientes probabilidades:
   a) $P(z < -2.96)$
   b) $P(z < 1.97)$
   c) $P(z > 2.94)$
   d) $P(z > -1.88)$
   e) $P(2.43 < z < 2.69)$
   f) $P(-1.34 < z < 2.71)$
   g) $P(-2.07 < z < -1.68)$

7. Una empresa produce pavos de tres meses y su desviación estándar es de 250 grs. Después de 6 meses se realiza un estudio a 30 pavos de 3 seleccionados aleatoriamente y da como resultado una desviación estándar de 225 grs. Con un nivel de significación del 5% comprobar si hay variación o no en el peso.

8. Se tiene dos líneas de producción de cereales, se toma una muestra de la línea Nro 01 de 36 cajas y tienen un promedio de 14.1 kg. y desviación estándar de 2.11 kg., de la línea Nro 02 se toma una muestra de 4 cajas y tienen un promedio de 15.6 kg. y una desviación estándar de 1.95 kg. Con un nivel de significación de 5%.

Se puede decir que hay diferencia significativa entre las dos líneas de producción.

.

CAPITULO VII

ANÁLISIS DE REGRESIÓN

CAPITULO VII

ANÁLISIS DE REGRESIÓN

Los métodos de regresión estudian la construcción de modelos para explicar o representar la dependencia entre una variable respuesta o dependiente (Y) y la(s) variable(s) explicativa(s) o dependiente(s), X

REGRESIÓN LINEAL

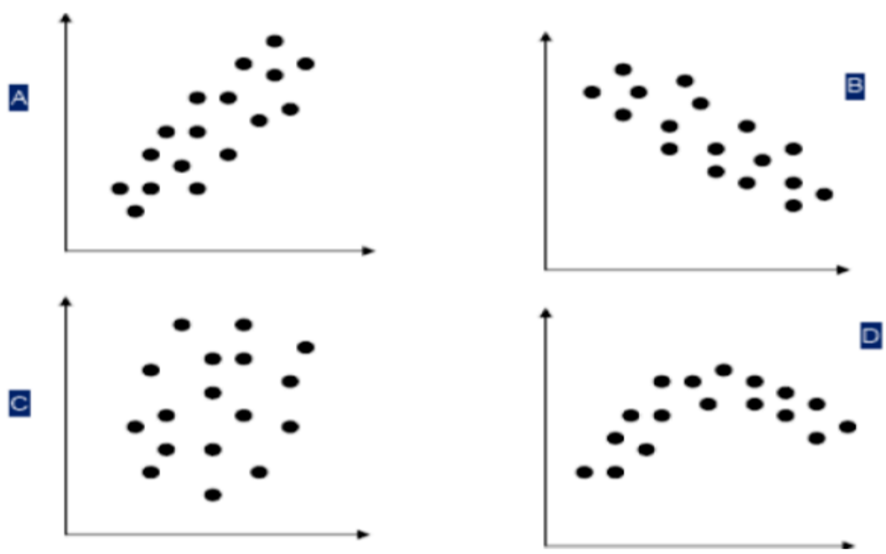
Conocer el efecto que una o varias variables pueden causar sobre otra, e incluso predecir en mayor o menor grado valores en una variable a partir de otra.

DIAGRAMA DE DISPERSIÓN

Es la gráfica donde se encuentran todos los puntos de las observaciones, tanto de la variable dependiente como la variable independiente.

El diagrama de dispersión nos da dos tipos de información:

a) Relación entre las variables.
b) Tipo de línea o ecuación de estimación.



ANÁLISIS DE CORRELACIÓN

Es la herramienta estadística de que nos valemos para describir el grado de relación que hay entre las dos variables.
Su interpretación es:

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

Y se representa:

0.00	$\longrightarrow$	Ninguna correlación
0.25	$\longrightarrow$	Correlación débil
0.50	$\longrightarrow$	Correlación Moderada
0.75	$\longrightarrow$	Correlación Intensa
1.00	$\longrightarrow$	Correlación perfecta

Mientras más cercano el valor a uno los datos son más precisos

MÉTODO DE LOS MÍNIMOS CUADRADOS

Permite ajustar los datos observados a la línea recta, este ajuste se obtendrá minimizando el error entre los puntos estimados (Y) y los observados (Y).
Con este método debe calcularse la pendiente de la estimación (b) y la intersección en el eje Y, (a). Al conocer estos dos factores conocemos la ecuación lineal, en donde al darle el valor de X, obtendremos el valor de Y.
Entonces:

$$Y = a + b(X)$$

También se pueden calcular a, b y r con las siguientes formulas:

a) Cálculo de "a"

$$a = \frac{\sum X^2 \sum Y - \sum X \sum Y}{N \sum X^2 - \left(\sum X\right)^2}$$

b) **Cálculo de "b"**

$$b = \frac{N \sum X.Y - \sum X \sum Y}{N \sum X^2 - (\sum X)^2}$$

c) **Cálculo de "R"**

$$R = \frac{N \sum X.Y - \sum X \sum Y}{\sqrt{N \sum X^2 - (\sum X)^2}\sqrt{N \sum Y^2 - (\sum Y)^2}}$$

Ejemplo

En la siguiente tabla se recogen los costos y el número de pasajeros de una línea aérea.

Calcular:

a) Calcular el Coeficiente de Correlación.
b) Obtener la Recta de Regresión para predecir el costo en función del número de pasajeros.
c) Estimar el costo de 103 pasajeros.

N° PASAJEROS X	COSTO($ 1000) Y	X^2	Y^2	X.Y
61	4.28	3,721	18.32	261.08
63	4.08	3,968	16.65	257.04
67	4.42	4,489	19.54	296.14
69	4.17	4,761	17.39	287.73
70	4.48	4,900	20.07	313.60
74	4.30	5,476	18.49	318.20
76	4.82	5,776	23.23	366.32
81	4.70	6,561	22.09	380.70
86	5.11	7,396	26.11	439.46
91	5.13	8,281	26.39	466.83
95	5.64	9,025	31.80	535.80
97	5.56	9,409	30.91	539.32
930	56.69	73,764	270.98	4,462.22

a) Procedemos calcular las columnas: X^2, Y^2 y X.Y.
b) Luego procedemos a sumar cada una de las columnas.
c) Y aplicamos las fórmulas para calcular los factores: a, b y R

1. **Cálculo de "a"**

$$a = \frac{(73,764)(56.69) - (930)(4,462.22)}{12(73,764) - (930)^2} = 1.57$$

2. **Cálculo de "b"**

$$b = \frac{12(4,462.22) - (930)(56.69)}{12(73,764) - (930)^2} = 0.0407$$

3. **Cálculo de "R"**

$$R = \frac{12(4,462.22) - (930)(56.69)}{\sqrt{(12)(73,764) - (930)^2}\sqrt{12(270.99) - (56.69)^2}} = 0.94$$

<u>Respuestas</u>.

El Coeficiente de Correlación es:

R^2 = 0.8836

La Ecuación de la Recta es:

Y = 1.57 + 0.0407(X)

El costo para 103 pasajeros es:

Y = 1.57 + 0.0407 (103)
Y = 5.76

APLICACIÓN DEL SOFTWARE ESTADÍSTICO SPSS

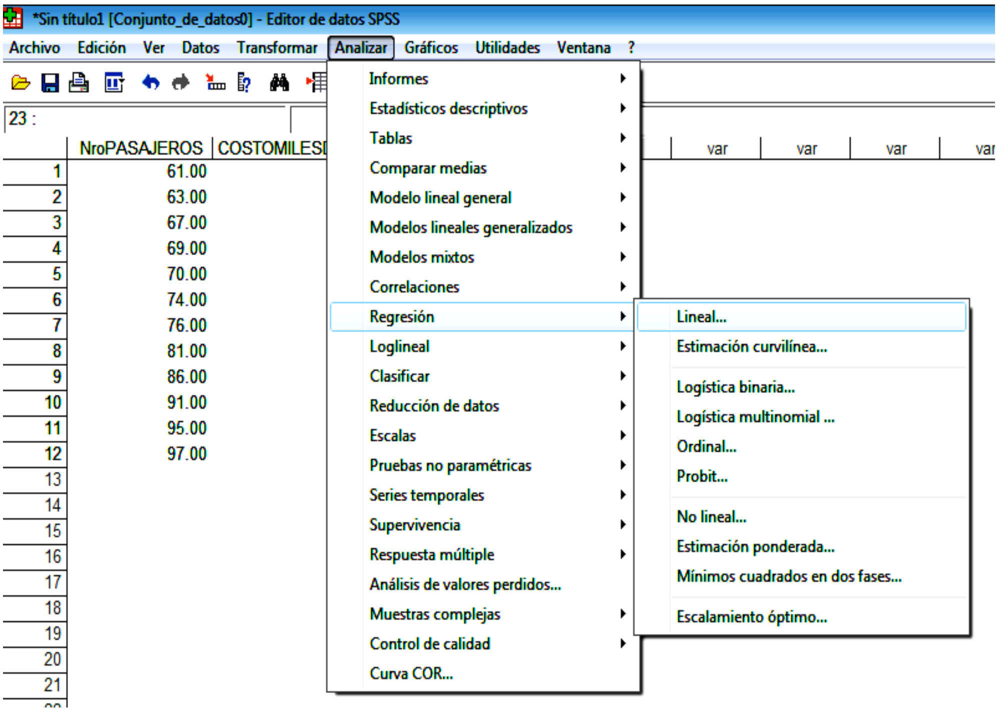
1. Abrir software SPSS y ingresar a vista variables, digitar variables

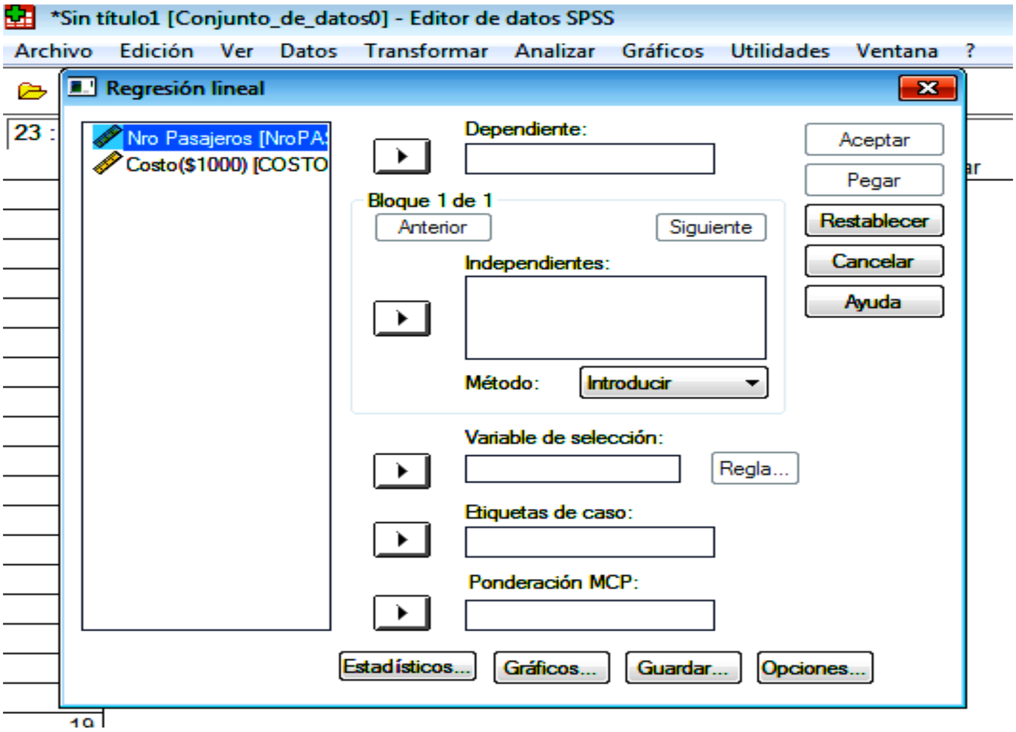
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	NroPASAJEROS	Numérico	8	2	Nro Pasajeros	Ninguno	Ninguno	8	Derecha	Escala
2	COSTOMILESDOLARES	Numérico	8	2	Costo($1000)	Ninguno	Ninguno	8	Derecha	Escala
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										

2. Luego ir a vista de datos y llenar los valores
   Elegir:
   a) Variable dependiente "Y" Costo Miles de Dolares
   b) Variable Independiente "X" Nro Pasajeros

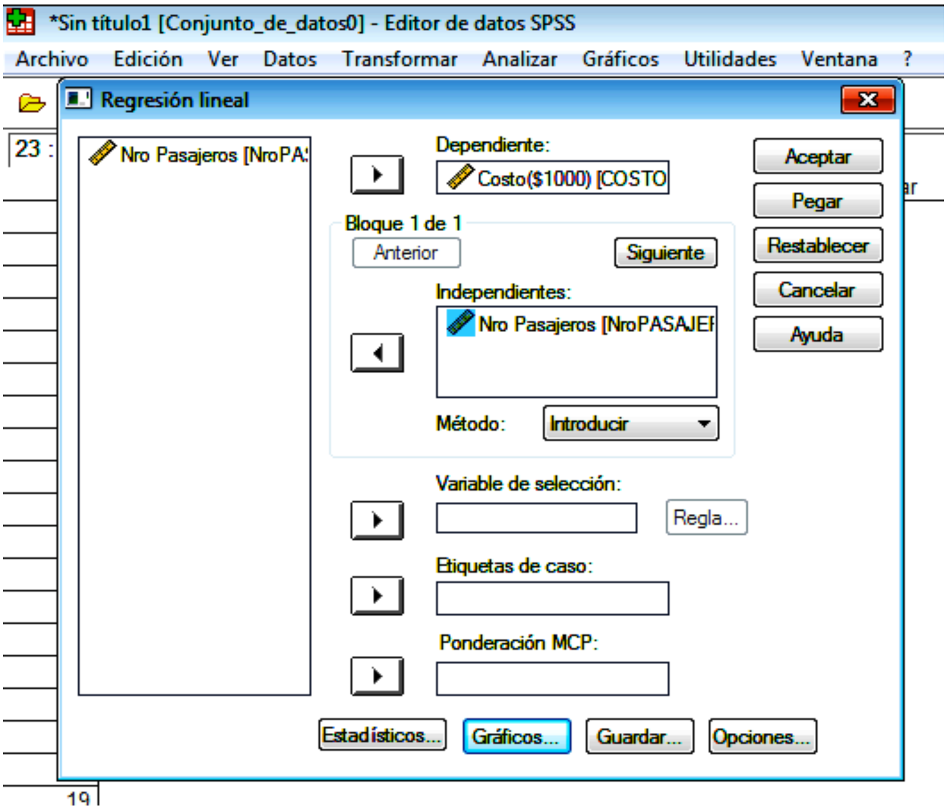
	NroPASAJEROS	COSTOMILESDOLARES	var	var	var	var
1	61.00	4.28				
2	63.00	4.08				
3	67.00	4.42				
4	69.00	4.17				
5	70.00	4.48				
6	74.00	4.30				
7	76.00	4.82				
8	81.00	4.70				
9	86.00	5.11				
10	91.00	5.13				
11	95.00	5.64				
12	97.00	5.56				

3. Luego ir a Analizar – Regresión - Lineal





4. Pasar La variable Dependiente (Y) Costo ($1000) y la Variable Independiente (X) Nro Pasajeros



5. Aceptar

Regresión

[Conjunto_de_datos0]

Variables introducidas/eliminadas[b]

Modelo	Variables introducidas	Variables eliminadas	Método
1	Nro Pasajeros[a]	.	Introducir

a. Todas las variables solicitadas introducidas

b. Variable dependiente: Costo($1000)

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	.948[a]	.899	.889	.17722

a. Variables predictoras: (Constante), Nro Pasajeros

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	2.798	1	2.798	89.092	.000[a]
	Residual	.314	10	.031		
	Total	3.112	11			

a. Variables predictoras: (Constante), Nro Pasajeros

b. Variable dependiente: Costo($1000)

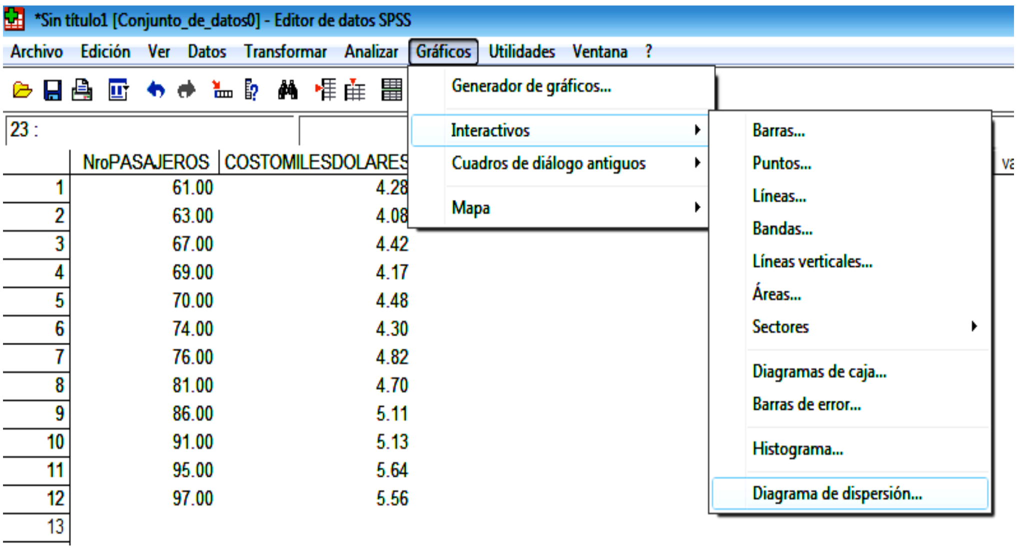
Coeficientes[a]

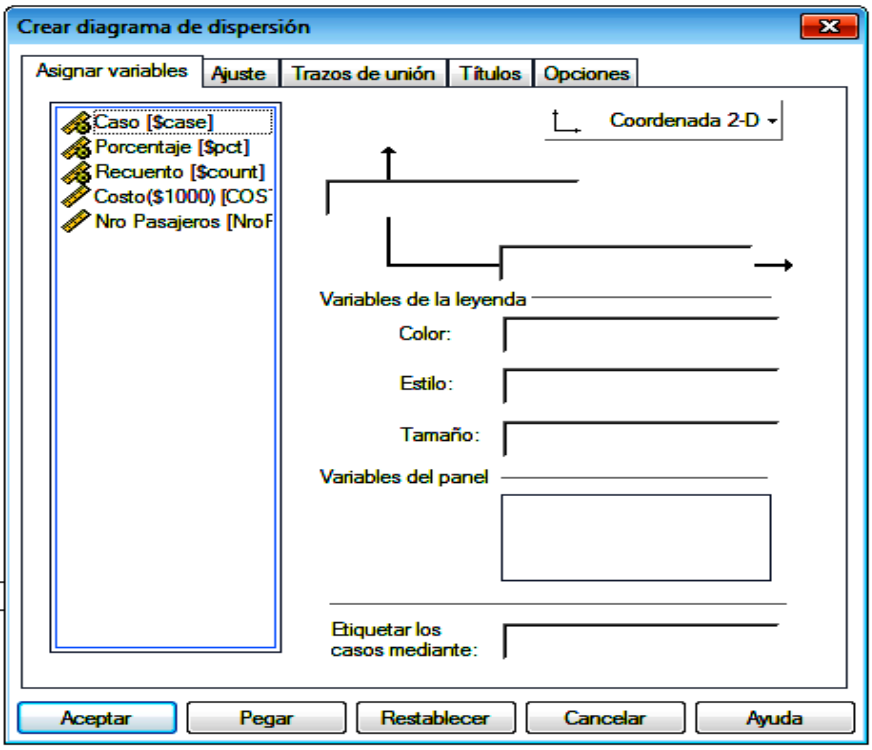
Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	1.570	.338		4.643	.001
	Nro Pasajeros	.041	.004	.948	9.439	.000

a. Variable dependiente: Costo($1000)

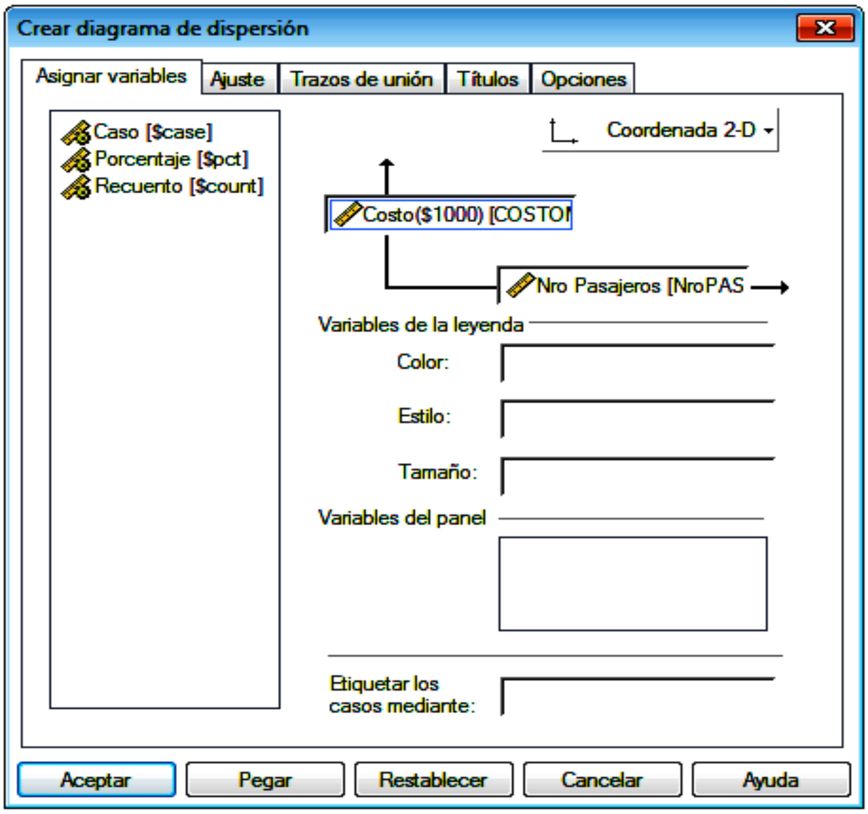
CREAR DIAGRAMA DE DISPERSIÓN

Luego ir a Gráficos – Interactivos – Diagrama de Dispersión

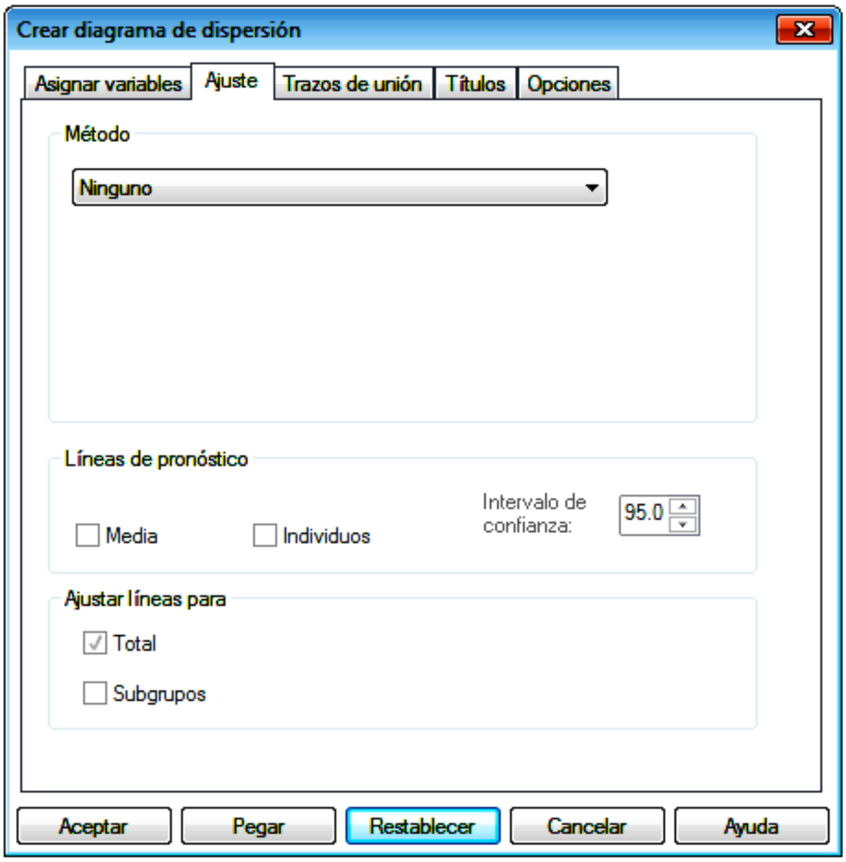




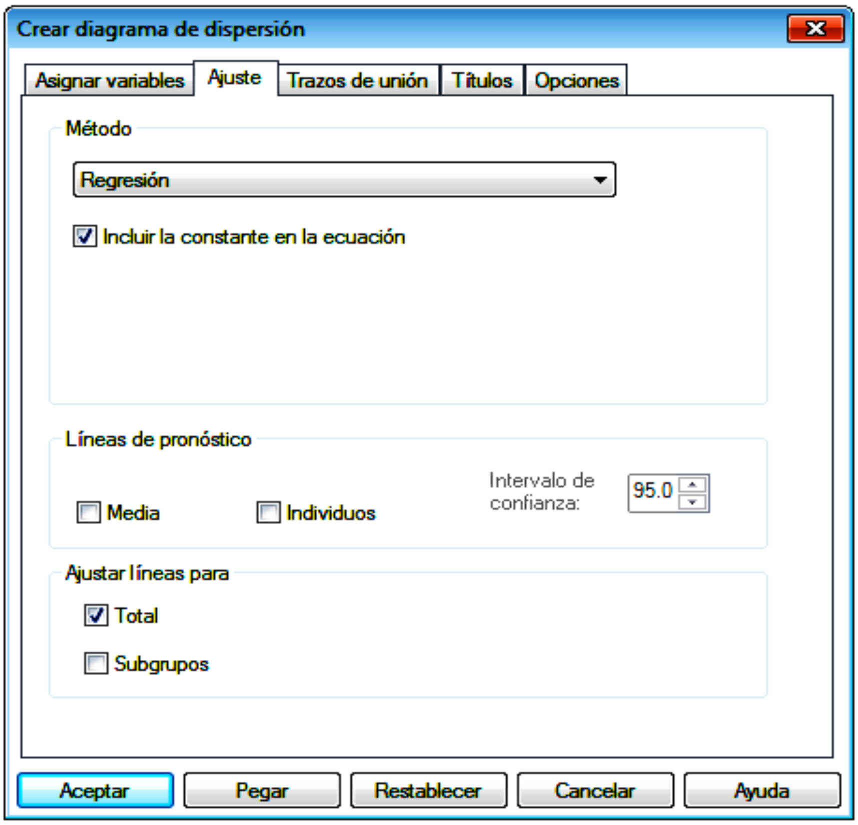
6. Pasar las variables



7. Ir a Ajuste



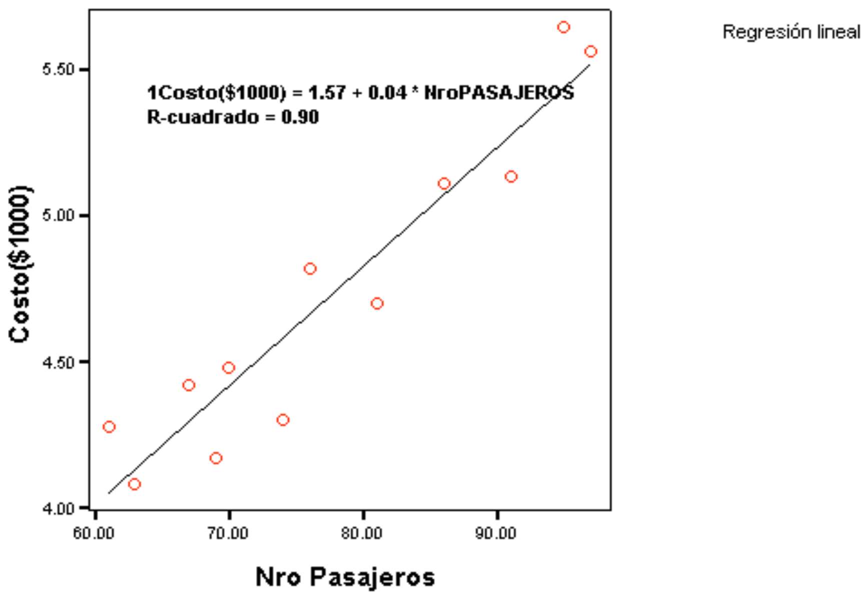
8. Escoger regresión lineal



9. Aceptar

Gráfico interactivo

[Conjunto_de_datos0]



REGRESIÓN MULTIPLE

El objetivo del análisis de regresión múltiple consiste en dar a conocer aquellas variables que son de utilidad en la predicción del valor de una variable dependiente. Cuando existe una variable que no resulta de ayuda en la predicción del análisis, esta variable pude ser eliminada del modelo de regresión múltiple y así resultaría un modelo más fácil de utilizar.
Entonces la Ecuación de la Recta es:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3$$

Ejemplo

En la facultad de Ingeniería de Procesos de la UNSA, se quiere entender las capacidades de aprendizaje de los alumnos que llevan la asignatura de Diseño de Reactores, y que tienen como prerrequisitos los cursos de Balance de Materia y Energía, Ingeniería de Reacciones y Fenómenos de transferencia, como se muestra en el siguiente cuadro:

DISEÑO DE REACTORES Y	BALANCE DE MATERIA Y ENERGIA X_1	INGENIERÍA DE REACCIONES X_2	FENOMENOS DE TRANSFERENCIA X_3
13	15	15	13
13	14	13	12
13	16	13	14
15	20	14	16
16	18	18	17
15	16	17	15
12	13	15	11
13	16	14	15
13	15	14	13
13	14	13	10
11	12	12	10
14	16	11	14
15	17	16	15
15	19	14	16
15	13	15	10

ANALISIS DE DATOS

1. **Variable Independiente:**

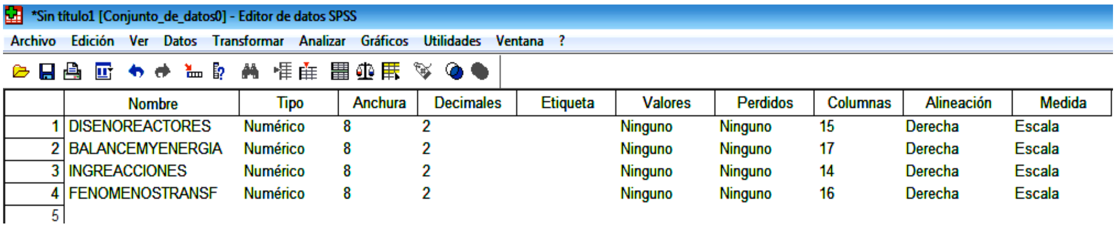
   a) Balance de Materia y Energía.
   b) Ingeniería de Reacciones.
   c) Fenómenos de Transferencia

2. **Variable Dependiente:**

   Diseño de Reactores

APLICACIÓN DEL SOFTWARE ESTADÍSTICO SPSS

1. Ingresamos la variable dependiente y las variables Independientes.
   Vamos a vista de variables

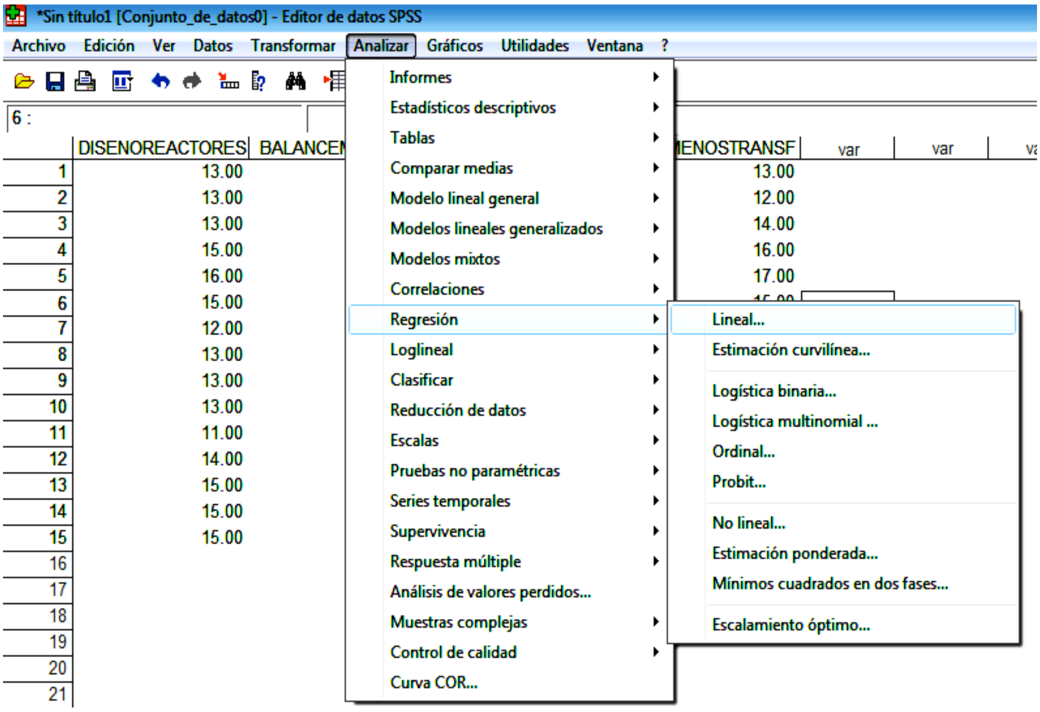


	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	DISENOREACTORES	Numérico	8	2		Ninguno	Ninguno	15	Derecha	Escala
2	BALANCEMYENERGIA	Numérico	8	2		Ninguno	Ninguno	17	Derecha	Escala
3	INGREACCIONES	Numérico	8	2		Ninguno	Ninguno	14	Derecha	Escala
4	FENOMENOSTRANSF	Numérico	8	2		Ninguno	Ninguno	16	Derecha	Escala
5										

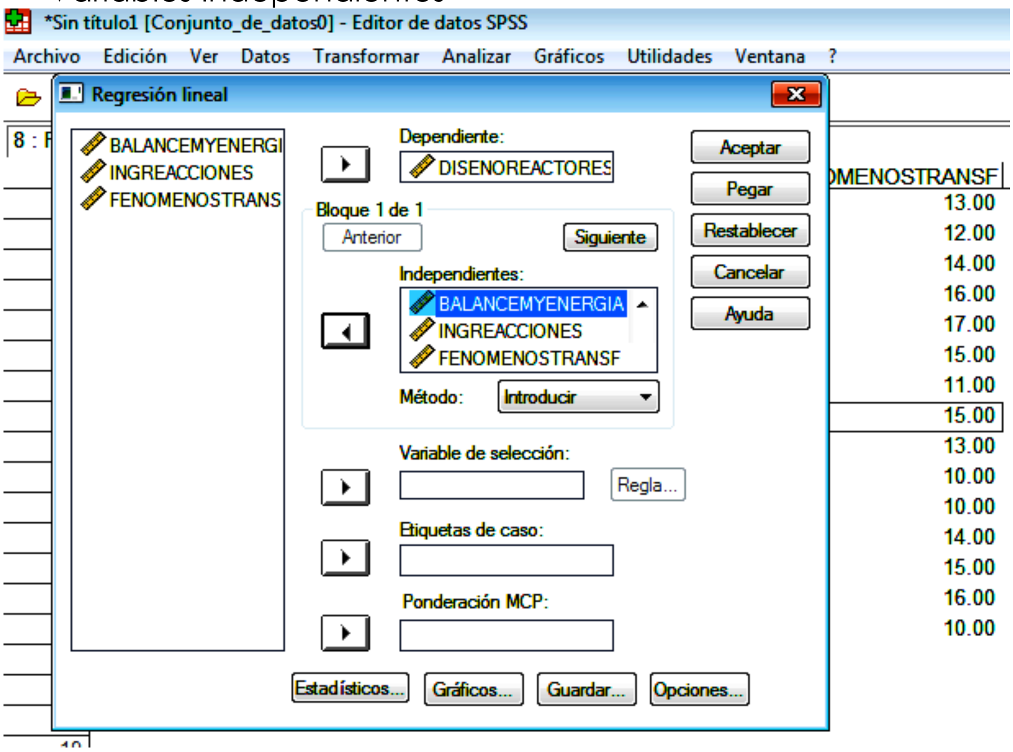
2. Luego vamos a vista de datos y ingresamos las notas respectivas

	DISENOREACTORES	BALANCEMYENERGIA	INGREACCIONES	FENOMENOSTRANSF
1	13.00	15.00	15.00	13.00
2	13.00	14.00	13.00	12.00
3	13.00	16.00	13.00	14.00
4	15.00	20.00	14.00	16.00
5	16.00	18.00	18.00	17.00
6	15.00	16.00	17.00	15.00
7	12.00	13.00	15.00	11.00
8	13.00	16.00	14.00	15.00
9	13.00	15.00	14.00	13.00
10	13.00	14.00	13.00	10.00
11	11.00	12.00	12.00	10.00
12	14.00	16.00	11.00	14.00
13	15.00	17.00	16.00	15.00
14	15.00	19.00	14.00	16.00
15	15.00	13.00	15.00	10.00
16				

3. Vamos a Analizar – Regresión – Lineal



4. Luego proceder a pasar la Variable Dependiente y las Tres Variables Independientes



5. Aceptar y se darán los siguientes resultados:

Regresión

[Conjunto_de_datos0]

Variables introducidas/eliminadas[b]

Modelo	Variables introducidas	Variables eliminadas	Método
1	FENOMENOSTRANSF, INGREACCIONES, BALANCEMYENERGIA[a]	.	Introducir

a. Todas las variables solicitadas introducidas

b. Variable dependiente: DISENOREACTORES

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	.835[a]	.697	.614	.86126

a. Variables predictoras: (Constante), FENOMENOSTRANSF, INGREACCIONES, BALANCEMYENERGIA

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	18.774	3	6.258	8.436	.003[a]
	Residual	8.160	11	.742		
	Total	26.933	14			

a. Variables predictoras: (Constante), FENOMENOSTRANSF, INGREACCIONES, BALANCEMYENERGIA

b. Variable dependiente: DISENOREACTORES

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	2.551	2.369		1.077	.305
	BALANCEMYENERGIA	.583	.267	.950	2.186	.051
	INGREACCIONES	.373	.144	.493	2.589	.025
	FENOMENOSTRANSF	-.242	.270	-.410	-.893	.391

a. Variable dependiente: DISENOREACTORES

LA ECUACIÓN DE LA RECTA ES:

$$Y = 2.551 + 0.583(X_1) + 0.373(X_2) - 0.242(X_3)$$

De estos resultados se puede decir:

1. De acuerdo al valor del coeficiente de correlación múltiple, podemos afirmar que las variables X_1 (Balance de Materia y Energía) , X_2 (Ingeniería de Reacciones) y X_3 (Fenómenos de Transferencia), se encuentran asociadas en forma directa de una manera muy fuerte con la variable dependiente Diseño de Reactores, en un 83.5%.
2. De acuerdo al Coeficiente de determinación R^2, podemos decir que el 69.70% de la nota de Diseño de Reactores pueden ser explicadas por las notas de Balance de Materia y Energía, Ingeniería de Reacciones y Fenómenos de Transferencia.

VERIFICACIÓN DE LA VALIDEZ DEL MODELO DE REGRESIÓN MULTIPLE

Para lo cual haremos cruce de la Variable Dependiente con cada una de las Variables Independientes. SPSS, aplicando regresión simple. Dando los siguientes resultados:

a) **NOTA DISEÑO DE REACTORES VS. NOTA BALANCE DE MATERIA Y ENERGÍA.**

Variables introducidas/eliminadas[b]

Modelo	Variables introducidas	Variables eliminadas	Método
1	BALANCEMY ENERGIA[a]	.	Introducir

a. Todas las variables solicitadas introducidas

b. Variable dependiente: DISENOREACTORES

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	.715[a]	.511	.474	1.00625

a. Variables predictoras: (Constante), BALANCEMYENERGIA

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	13.770	1	13.770	13.600	.003[a]
	Residual	13.163	13	1.013		
	Total	26.933	14			

a. Variables predictoras: (Constante), BALANCEMYENERGIA

b. Variable dependiente: DISENOREACTORES

b) NOTA DISEÑO DE REACTORES VS. NOTA INGENIERÍA DE REACCIONES.

Variables introducidas/eliminadas[b]

Modelo	Variables introducidas	Variables eliminadas	Método
1	INGREACCI ONES[a]	.	Introducir

a. Todas las variables solicitadas introducidas

b. Variable dependiente: DISENOREACTORES

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	.593[a]	.351	.301	1.15949

a. Variables predictoras: (Constante), INGREACCIONES

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	9.456	1	9.456	7.034	.020[a]
	Residual	17.477	13	1.344		
	Total	26.933	14			

a. Variables predictoras: (Constante), INGREACCIONES

b. Variable dependiente: DISENOREACTORES

c) NOTA DISEÑO DE REACTORES VS. NOTA FENÓMENOS DE TRANSFERENCIA.

Variables introducidas/eliminadas[b]

Modelo	Variables introducidas	Variables eliminadas	Método
1	FENOMENOSTRANSF[a]	.	Introducir

a. Todas las variables solicitadas introducidas

b. Variable dependiente: DISENOREACTORES

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	.669[a]	.448	.406	1.06939

a. Variables predictoras: (Constante), FENOMENOSTRANSF

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	12.066	1	12.066	10.551	.006[a]
	Residual	14.867	13	1.144		
	Total	26.933	14			

a. Variables predictoras: (Constante), FENOMENOSTRANSF

b. Variable dependiente: DISENOREACTORES

De acuerdo a los cálculos desarrollado: se puede concluir:

CRUCE NOTAS	R^2
NOTA DISEÑO DE REACTORES VS. NOTA BALANCE DE MATERIA Y ENERGÍA.	**51.10%**
NOTA DISEÑO DE REACTORES VS. NOTA INGENIERÍA DE REACCIONES.	**35.10%**
NOTA DISEÑO DE REACTORES VS. NOTA FENÓMENOS DE TRANSFERENCIA	**44.80%**

Por lo que se puede concluir:

1) La variable que tiene mayor relación con la variable dependiente es la Nota de Balance de Materia y Energía con un 51.10%
2) La variable que tiene mediana relación con la variable dependiente es la Nota de Fenómenos de Transferencia con un 44.80%
3) La variable que tiene menor relación con la variable dependiente es la Nota de Ingeniería de Reacciones con un 35.10%

EJERCICIOS

1. De un equipo de baloncesto, de los cuales se mide su talla y su peso, se tienen los siguientes resultados:

Talla X (m.)	1.8	1.6	1.8	1.5	1.7	1.8	1.9	1.7	1.6	1.7	1.7	1.7
Peso Y (kg.)	70.4	68.1	81.7	61.3	70.8	76.3	80.8	72.6	59.9	65.8	63.1	69.0

   Se pide lo siguiente:
   a) Obtener el diagrama de dispersión de estos datos.
   b) Trazar una recta que se aproxime a los datos.
   c) Estimar el peso de un estudiante cuya estatura es 1.9 m.
   d) Estimar la estatura de un estudiante cuyo peso es 76.7 kg.

2. Con los siguientes datos calcular la recta de regresión y el coeficiente de correlación, utilizando el programa SPSS

GASTO Y	INGRESO X1	FAMILIA X2
430	2100	3
310	1100	4
320	900	5
460	1600	4
1250	6200	4
440	2300	3
520	1800	6
290	1000	5
1290	8900	3
350	2400	2
350	1200	4
780	4700	3
430	3500	2
470	2900	3
380	1400	4

3. Con los siguientes datos calcular la recta de regresión y el coeficiente de correlación, utilizando el programa SPSS

PESO Y	TALLA X1	CINTURA X2	EDAD X3
75	1.69	0.620	26
93	1.97	0.750	32
64	1.71	0.600	30
73	1.76	0.710	65
59	1.62	0.660	45
79	1.70	0.620	42
86	1.91	0.790	38
86	1.87	0.740	36
74	1.77	0.700	35
63	1.71	0.660	30
81	1.77	0.710	20
73	1.80	0.690	51

CAPITULO VIII

ANÁLISIS FACTORIAL

CAPITULO VIII

ANÁLISIS FACTORIAL

EXPERIMENTOS CON UN SOLO FACTOR

APLICACIÓN EN FORMA MANUAL

Ejemplo 1

La siguiente información muestral sobre el tratamiento de acero, verifique la hipótesis de que las medias de tratamiento son iguales. Utilice un nivel de significancia del 0.05

T_1	T_2	T_3
8	3	3
6	2	4
10	4	5
9	3	4

Se siguen los siguientes pasos:

1. La Hipótesis
   H_0 : Todas las medias son iguales
   H_1 : Al menos una media es distinta

2. **α = 0.05**
3. Calcular:
   3.1 Sacar media por cada grupo

   $$\bar{X}_1 = 8.25, \quad \bar{X}_2 = 3, \quad \bar{X}_3 = 4$$

   3.2 Calcular
   Nro Columnas = C = 3
   Nro Filas por grupo = r = 4
   n = C * r = 3 * 4 = 12

3.3 Calcular el promedio, de las medias obtenidas (la gran media)

$$\Pr omedio = \frac{8.25 + 3 + 4}{3} = 5.08$$

3.4 SUMA DE CUADRADOS TOTALES **(SCT)**

$$SCT = \sum\sum\left(X_{ij} - \Pr omedio\right)^2$$

Aplicando lá formula anterior a cada uno de los datos se obtiene lós siguientes resultados

	T_1	T_2	T_3	TOTAL
	8.53	**4.33**	**4.33**	**17.19**
	0.85	**9.49**	**1.17**	**11.1**
	24.20	**1.17**	**0.006**	**25.38**
	15.37	**4.33**	**1.17**	**20.87**
TOTAL	**48**	**19.32**	**6.68**	**74.95**

3.5 SUMA DE CUADRADOS DE TRATAMIENTO **(SCTR)**

$$SCTR = \sum rj\left(\overline{X}_j - \Pr omedio\right)^2$$

Donde r=4 (Nro Datos) y se trabaja la media de cada grupo con el promedio y elevandolo al cuadrado y multiplicando por el Nro de datos de cada tratamento.
Aplicando la formula anterior se tiene los siguientes resultados.

40.20	**17.31**	**4.67**	**62.28**

3.6 SUMA CUADRADOS DE ERROR **(SCE)**

$$SCE = \sum\sum\left(X_{ij} - \overline{X}_j\right)^2$$

Se aplica cada uno de los valores del ejercicio restandole la media de cada tratamiento elevandolo al cadrado

	T_1	T_2	T_3	TOTAL
	0.06	0	1	1.06
	5.06	1	0	6.06
	3.06	1	1	5.06
	0.56	0	0	0.56
TOTAL	8.74	2	2	12.74

3.7 CUADRADO MEDIO TOTAL **(CMT)**

$$CMT = \frac{SCT}{n-1} = \frac{74.95}{12-1} = 6.81$$

3.8 CUADRADO MEDIO DEL TRATAMIENTO **(CMTR)**

$$CMTR = \frac{SCTR}{c-1} = \frac{62.18}{3-1} = 31.09$$

3.9 CUADRADO MEDIO DEL ERROR **(CME)**

$$CME = \frac{SCE}{n-c} = \frac{12.74}{12-3} = 1.42$$

4. PRUEBA DE FISHER **(F)**

$$F = \frac{CMTR}{CME} = \frac{31.09}{1.42} = 21.89$$

BUSCAR EN LA TABLA FISHER

$$Grado\ Libertad = \frac{c-1}{n-c} = \frac{3-1}{12-3} = \frac{2}{9}$$

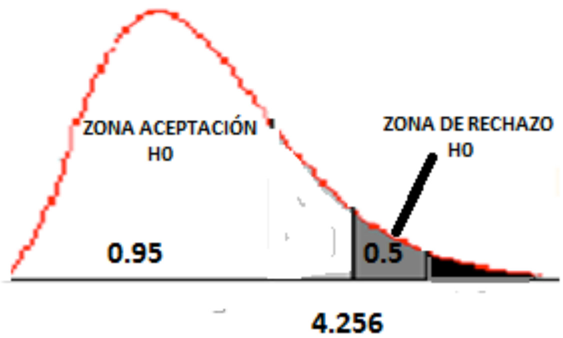
Posición al 5%

Buscando en la Tabla de Fisher, tenemos:

	1	2
1		
2		
. . .		
9		4.256

El valor en Tabla es: 4.256

5. REGLA DE DECISIÓN



CONDICIÓN

Si F $\leq$ VALOR TABLA, SE ACEPTA H_0

SI F $>$ VALOR TABLA, SE RECHAZA H_0

Como: 21.89 $>$ 4.256

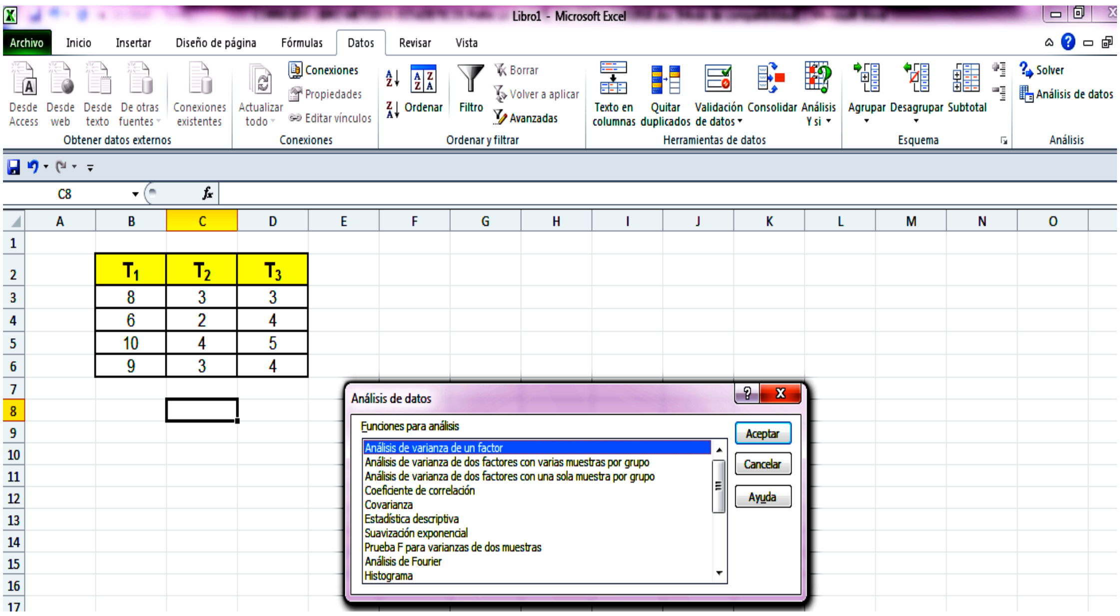
Se Rechaza H_0 y se Acepta H_1 al menos uma media es distinta

APLICACIÓN DEL SOFTWARE DEL PROGRAMA EXCEL

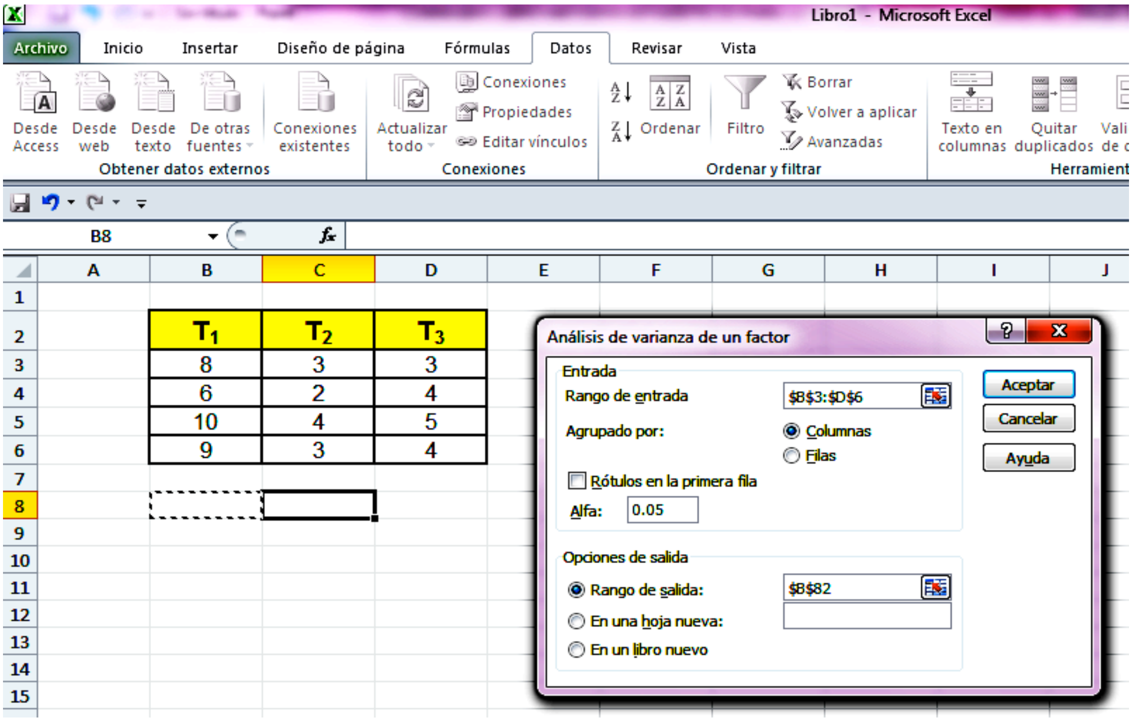
La siguiente información muestral sobre el tratamiento de acero, verifique la hipótesis de que las medias de tratamiento son iguales. Utilice un nivel de significancia del 0.05

T_1	T_2	T_3
8	3	3
6	2	4
10	4	5
9	3	4

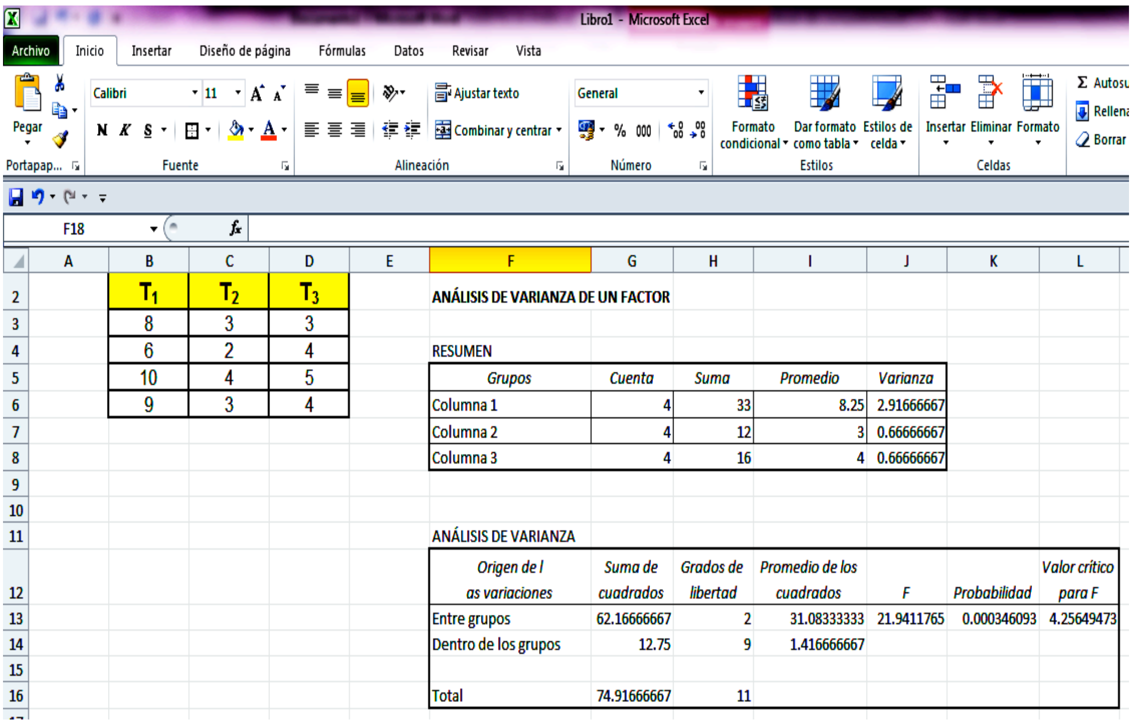
Ir a Datos – Analizar Datos – Análisis de Varianza de un Factor



Aceptar
Colocar rango de los datos y donde se mostrara los resultados


Libro1 - Microsoft Excel
Análisis de varianza de un factor
Entrada
Rango de entrada $B$3:$D$6
Agrupado por: Columnas
Filas
Rótulos en la primera fila
Alfa: 0.05
Opciones de salida
Rango de salida: $B$82
En una hoja nueva:
En un libro nuevo
Aceptar
Cancelar
Ayuda
T1 T2 T3
8 3 3
6 2 4
10 4 5
9 3 4

Aceptar


Libro1 - Microsoft Excel
T1 T2 T3
8 3 3
6 2 4
10 4 5
9 3 4
ANÁLISIS DE VARIANZA DE UN FACTOR
RESUMEN
Grupos Cuenta Suma Promedio Varianza
Columna 1 4 33 8.25 2.91666667
Columna 2 4 12 3 0.66666667
Columna 3 4 16 4 0.66666667
ANÁLISIS DE VARIANZA
Origen de las variaciones Suma de cuadrados Grados de libertad Promedio de los cuadrados F Probabilidad Valor crítico para F
Entre grupos 62.16666667 2 31.08333333 21.9411765 0.000346093 4.25649473
Dentro de los grupos 12.75 9 1.416666667
Total 74.91666667 11

CONCLUSIÓN

CONDICIÓN

Si F $\leq$ VALOR TABLA, SE ACEPTA H_0

SI F $>$ VALOR TABLA, SE RECHAZA H_0

Como: 21.89 $>$ 4.256

Se Rechaza H_0 y se Acepta H_1 al menos uma media es distinta

Ejemplo 2

Una Fábrica de rodamientos, tiene dos abastecedores de barras de acero para fabricar rodamientos. Un nuevo abastecedor indica que sus barras de acero son mejores en cuestión de flexibilidad pero con un precio mayor.
La fábrica para corroborar estas mejoras realiza unas pruebas, con sus tres abastecedores, obteniéndose los siguientes resultados.

	ABASTECEDOR		
	1	**2**	**3(NUEVO)**
Resultados obtenidos. Flexibilidad.	23	35	50
	28	36	43
	21	29	36
	27	40	34
	95	43	45
	41	49	52
	37	51	52
	30	28	43
	32	50	44
	36	52	34

a) Hipótesis planteada

H_0 : Las barras de acero del 1° y 2° abastecedor de la fábrica de rodajes son iguales que el tercer abastecedor.

H_1 : Las barras de acero del 1° y 2° abastecedor de la fábrica de rodajes no son iguales que el tercer abastecedor.

b) Nivel de significación

Alfa = 0.05

c) **Caculo Estadístico y Prueba**

	1	2	3
	23	35	50
	28	36	43
	21	29	36
	27	40	34
	95	43	45
	41	49	52
	37	51	52
	30	28	43
	32	50	44
	36	52	34
	Código 1	Código 2	Código 3
∑	370	413	433
	37	41.3	43.3
	40.5		
n	30		

1. Suma de Cuadrados Totales (SCT)

SCT= $(23-40.5)^2+(28-40)^2+(21-40.5)^2+(27-40.5)^2+(95-40.5)^2+(41-40.5)^2+(37-40.5)^2+(30-40.5)^2+(32-40.5)^2+(36-40.5)^2 = 4212.84$

SCT = $(35-40.5)^2+(36-40)^2+(29-40.5)^2+(40-40.5)^2+(43-40.5)^2+(49-40.5)^2+(51-40.5)^2+(28-40.5)^2+(50-40.5)^2+(52-40.5)^2=502.64$

SCT = $(50-40.5)^2+(43-40)^2+(36-40.5)^2+(34-40.5)^2+(45-40.5)^2+(52-40.5)^2+(52-40.5)^2+(43-40.5)^2+(44-40.5)^2+(34-40.5)^2 =4212.84$

SCT =5345.45

2. Suma de Cuadrado de Tratamiento (SCTR)

SCTR = $(37-40.5)^2+(41.340.5)^2+(43.5)^2$
SCTR= 207.57

3. Suma de Cuadrado Error (SCE)

$SCE = (23-37)^2+(28-37)^2+(21-73)^2+(27-73)^2+(95-37)^2+(41-37)^2+(37-37)^2+(32-37)^2+(30-37)^2+(36-37)^2 = 4088$

$SCE =(35-41.3)^2+(36-41.3)^2+(29-41.3)^2+(40-41.3)^2+(43-41.3)^2+(49-41.3)^2+(51-41.3)^2+(28-1.3)^2+(50-41.3)^2+(52-41.3)^2 = 744.1$

$SCE = (50-43.3)^2+(43-43.3)^2+(36-43.3)^2+(34-43.3)^2+(45-43.3)^2+(52-43.3)^2+(52-43.3)^2+(43-43.3)^2+(44-43.3)^2+(34-43.3)^2 = 426.1$

4. Cuadrado Medio Total (CMT)

CMT = 5345.45/29 =184.33

5. Cuadrado Medio del Error (CME)

SCE = 426.1/27 = 194.5

6. Cuadrado Medio de Tratamiento (CMT)

CMT =207.57/2 = 103.63

7. Prueba de F calculada (Fc)

Fc = 103.63 /194.5 = 0.53

Tabla de F para alfa 0.05
GL = (3-1)/30-3 =2/27

F tabulada (Ft) = 3.37

Por tanto: Fc <Ft; en consecuencia, se acepta la Ho y se rechaza la H_1

INTERPRETACION

No es cierto que el tercer abastecedor de barras de acero tengan mejor flexibilidad y más caros, y que eran mejor que las barras de acero del primero y segundo abastecedor.
En consecuencia, la fábrica de rodajes puede decidir si tomarlos o dejarlos

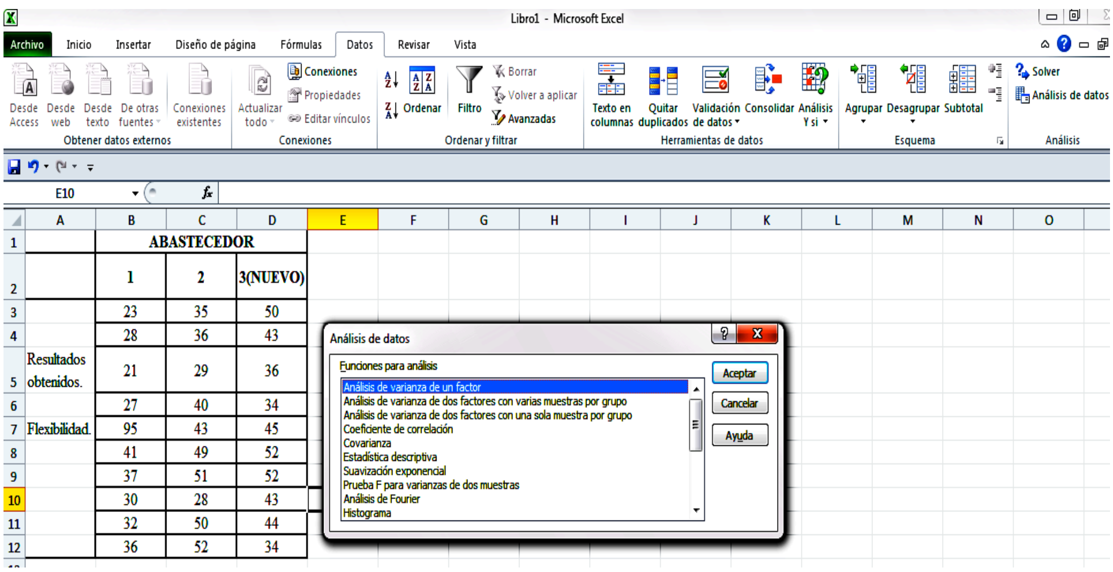
APLICACIÓN DEL SOFTWARE DEL PROGRAMA EXCEL

Ejemplo 2

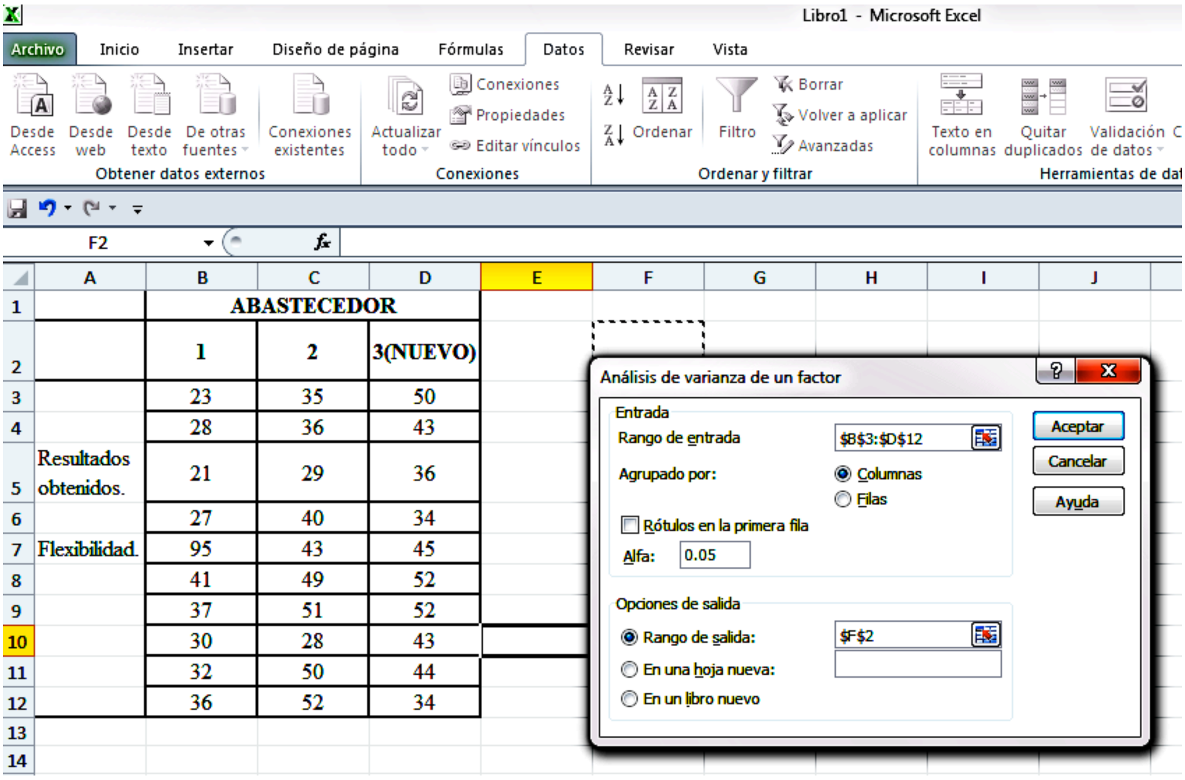
Una Fábrica de rodamientos, tiene dos abastecedores de barras de acero para fabricar rodamientos. Un nuevo abastecedor indica que sus barras de acero son mejores en cuestión de flexibilidad pero con un precio mayor.
La fábrica para corroborar estas mejoras realiza unas pruebas, con sus tres abastecedores, obteniéndose los siguientes resultados.

	ABASTECEDOR		
	1	**2**	**3(NUEVO)**
Resultados obtenidos. Flexibilidad.	23	35	50
	28	36	43
	21	29	36
	27	40	34
	95	43	45
	41	49	52
	37	51	52
	30	28	43
	32	50	44
	36	52	34

Ir a Datos – Analizar Datos – Análisis de Varianza de un Factor

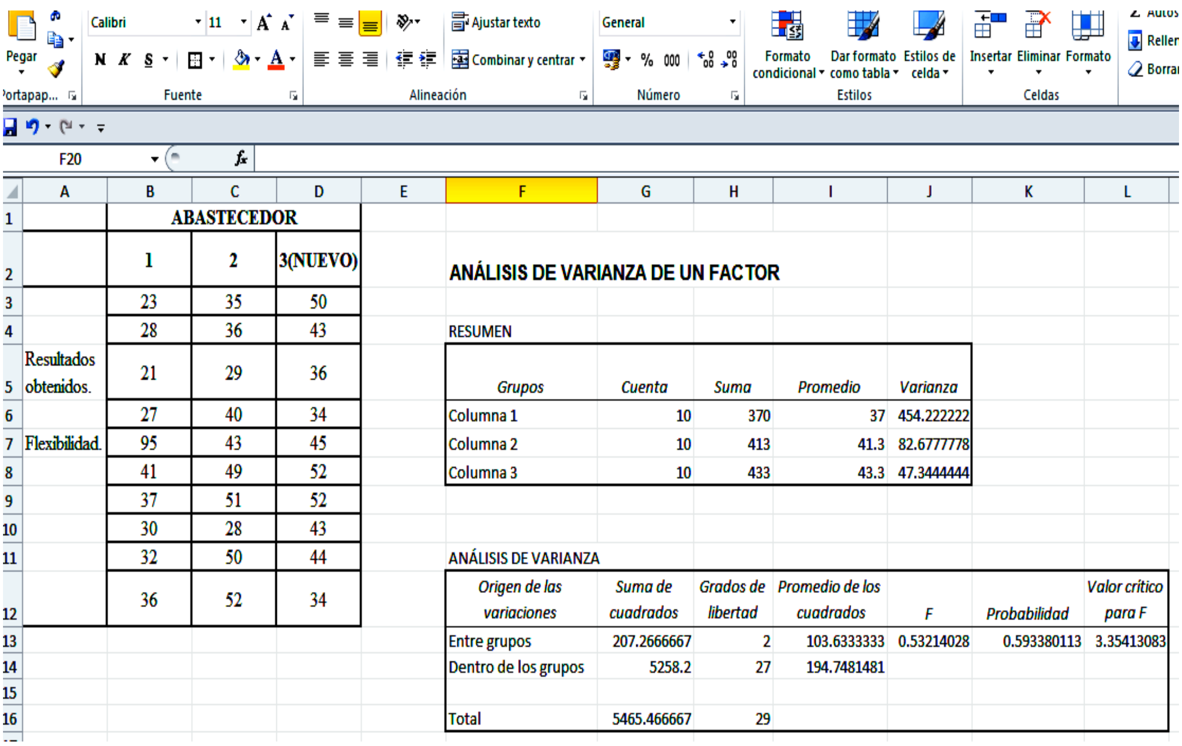


Aceptar
Colocar rango de los datos y donde se mostrara los resultados



	ABASTECEDOR		
	1	2	3(NUEVO)
	23	35	50
	28	36	43
Resultados obtenidos.	21	29	36
	27	40	34
Flexibilidad.	95	43	45
	41	49	52
	37	51	52
	30	28	43
	32	50	44
	36	52	34

Aceptar



ANÁLISIS DE VARIANZA DE UN FACTOR

RESUMEN

Grupos	Cuenta	Suma	Promedio	Varianza
Columna 1	10	370	37	454.222222
Columna 2	10	413	41.3	82.6777778
Columna 3	10	433	43.3	47.3444444

ANÁLISIS DE VARIANZA

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Entre grupos	207.2666667	2	103.6333333	0.53214028	0.593380113	3.35413083
Dentro de los grupos	5258.2	27	194.7481481			
Total	5465.466667	29				

CONCLUSIÓN

CONDICIÓN

Si F $\leq$ VALOR TABLA, SE ACEPTA H_0

SI F $>$ VALOR TABLA, SE RECHAZA H_0

Como: 0.53 < 3.35

Se Acepta H_0 y se Rechaza H_1

8.1 EXPERIMENTOS CON DOS FACTORES

Ejemplo.

Se requiere fabricar um reactor CSTR y se obtienen los siguientes resultados de corrosión en días de unas muestras de acero, atacadas con diferentes concentraciones de pH.

ACERO	pH 6	pH 5	pH 4
Acero al 75%	66	16	10
Acero al 75%	79	19	35
Acero al 75%	38	39	41
Acero al 75%	91	37	29
Acero al 85%	76	67	13
Acero al 85%	95	60	35
Acero al 85%	81	52	29
Acero al 85%	64	57	23
Acero al 95%	70	86	48
Acero al 95%	56	59	52
Acero al 95%	85	74	41
Acero al 95%	81	69	30

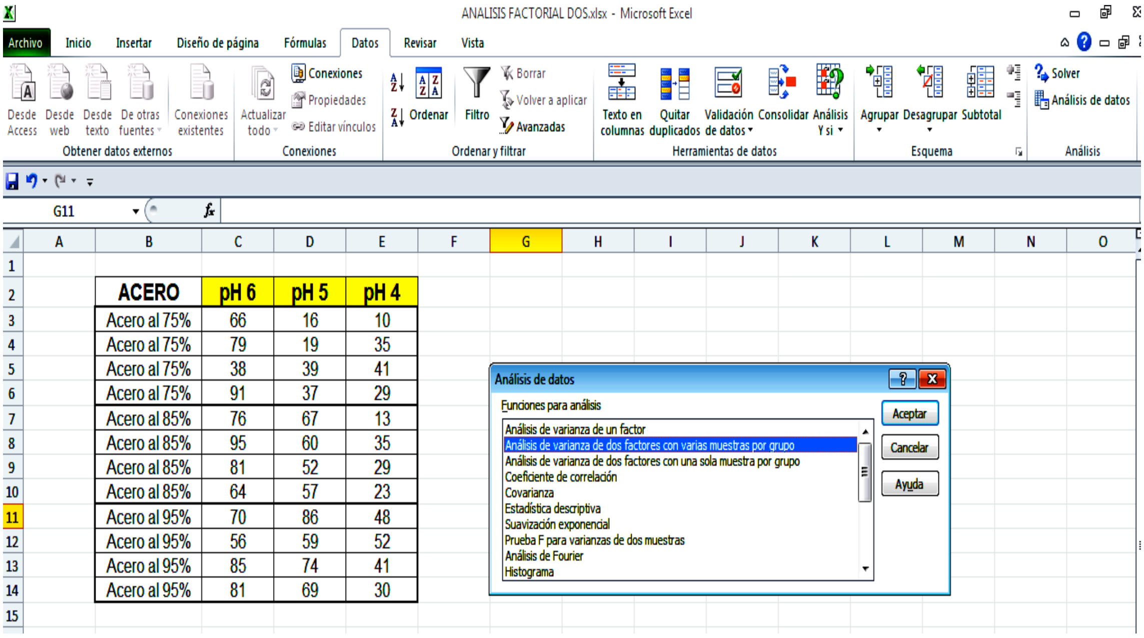
Calcular:

1. ¿Qué efectos tienen el tipo de acero y el pH en el reactor a fabricarse?

2. ¿Existe un acero que haga que la vida del reactor sea independiente del pH?

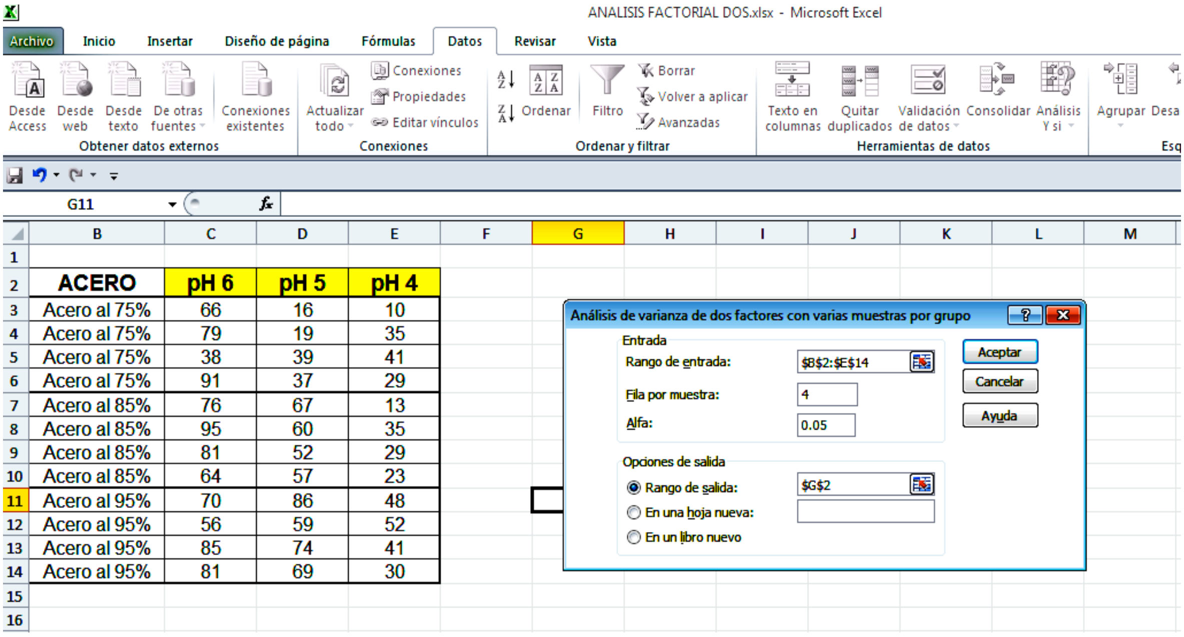
8.1.1 APLICACIÓN DEL SOFTWARE DEL PROGRAMA EXCEL

Ir a Datos
Analizar Datos
Análisis de Varianza de dos Factores con varias muestras por grupo



ACERO	pH 6	pH 5	pH 4
Acero al 75%	66	16	10
Acero al 75%	79	19	35
Acero al 75%	38	39	41
Acero al 75%	91	37	29
Acero al 85%	76	67	13
Acero al 85%	95	60	35
Acero al 85%	81	52	29
Acero al 85%	64	57	23
Acero al 95%	70	86	48
Acero al 95%	56	59	52
Acero al 95%	85	74	41
Acero al 95%	81	69	30

Aceptar
Llenar: Datos de rango de entrada, Fila por muestra 4, alfa 0.05 (95%) y Rango de Salida.


ANALISIS FACTORIAL DOS.xlsx - Microsoft Excel
Archivo Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista
ACERO pH 6 pH 5 pH 4
Acero al 75% 66 16 10
Acero al 75% 79 19 35
Acero al 75% 38 39 41
Acero al 75% 91 37 29
Acero al 85% 76 67 13
Acero al 85% 95 60 35
Acero al 85% 81 52 29
Acero al 85% 64 57 23
Acero al 95% 70 86 48
Acero al 95% 56 59 52
Acero al 95% 85 74 41
Acero al 95% 81 69 30
Análisis de varianza de dos factores con varias muestras por grupo
Entrada
Rango de entrada: $B$2:$E$14
Fila por muestra: 4
Alfa: 0.05
Opciones de salida
Rango de salida: $G$2
En una hoja nueva:
En un libro nuevo
Aceptar
Cancelar
Ayuda

Aceptar

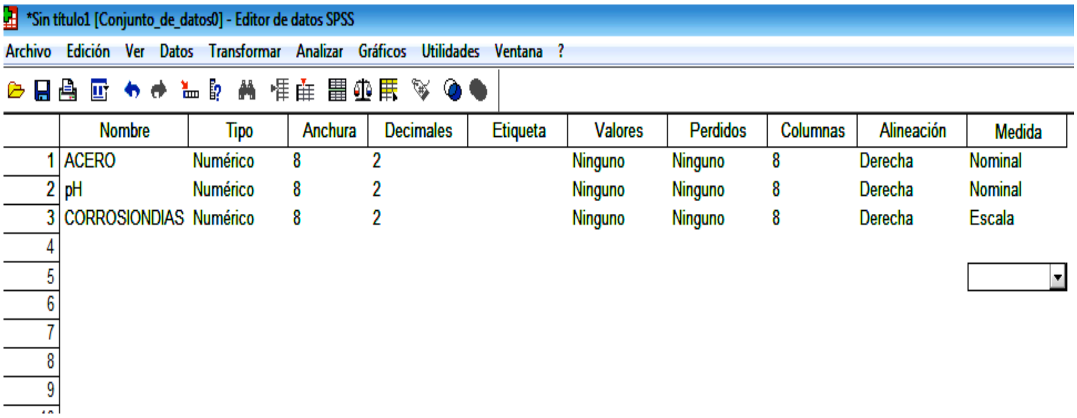
Análisis de varianza de dos factores con varias muestras por grupo				
RESUMEN	**pH 6**	**pH 5**	**pH 4**	**Total**
Acero al 75%				
Cuenta	4	4	4	12
Suma	274	111	115	500
Promedio	68.5	27.75	28.75	41.66667
Varianza	517.66667	142.25	180.25	622.0606
Acero al 85%				
Cuenta	4	4	4	12
Suma	316	236	100	652
Promedio	79	59	25	54.33333
Varianza	164.66667	39.333333	88	621.697
Acero al 95%				
Cuenta	4	4	4	12
Suma	292	288	171	751
Promedio	73	72	42.75	62.58333
Varianza	168.66667	126	92.91666667	320.447
Total				
Cuenta	12	12	12	
Suma	882	635	386	
Promedio	73.5	52.916667	32.16666667	
Varianza	252.27273	460.08333	162.1515152	

ANÁLISIS DE VARIANZA						
Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Muestra	2664.0556	2	1332.027778	7.888304	0.00200505	3.354130829
Columnas	10250.722	2	5125.361111	30.35253	1.2376E-07	3.354130829
Interacción	2396.2778	4	599.0694444	3.547705	0.018867987	2.727765306
Dentro del grupo	4559.25	27	168.8611111			
Total	19870.306	35				

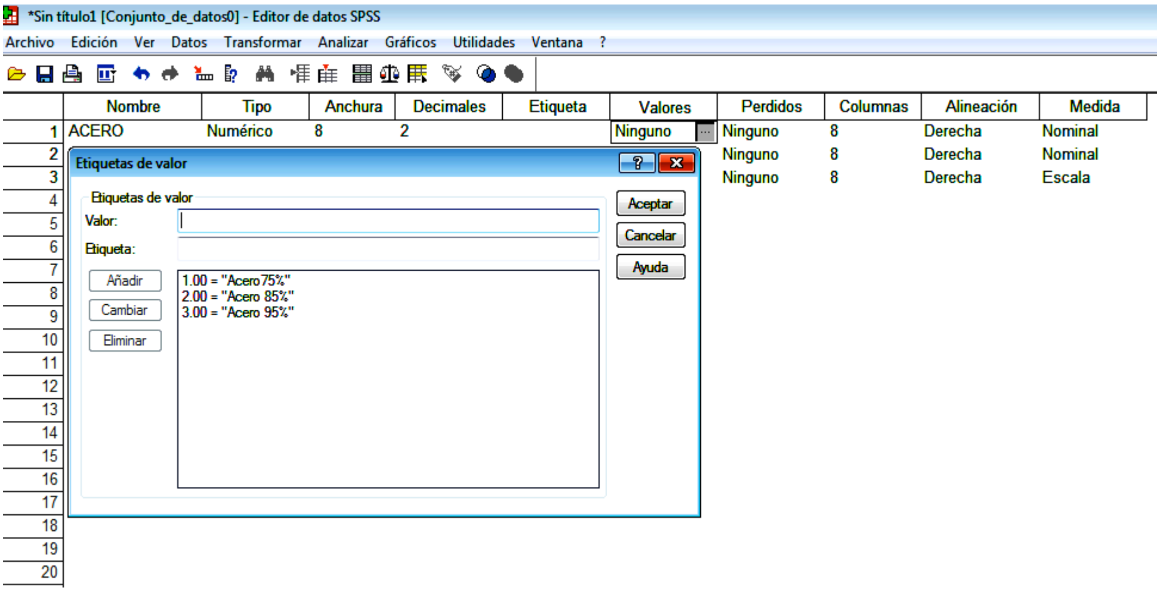
CONCLUSIÓN: El pH es el que más incide en la fabricación del reactor.

APLICACIÓN DEL SOFTWARE ESTADÍSTICO SPSS

1. Abrir software SPSS y ingresar a vista variables, digitar variables

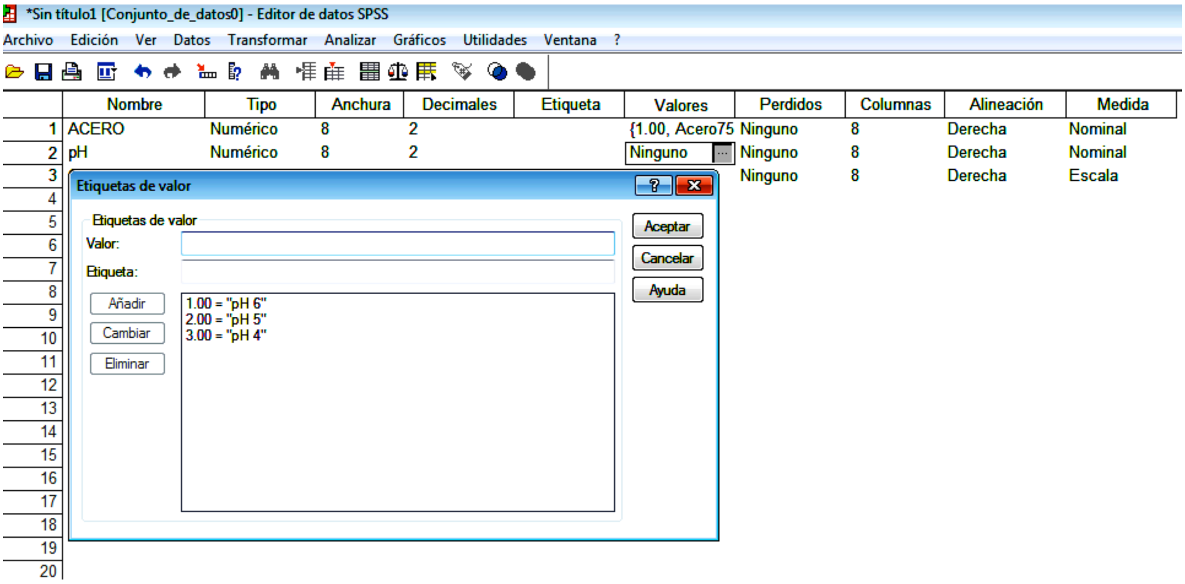


2. Ir a Valores de Acero y digitar
   1 para Acero 75%
   2 para Acero 85%
   3 para Acero 95%

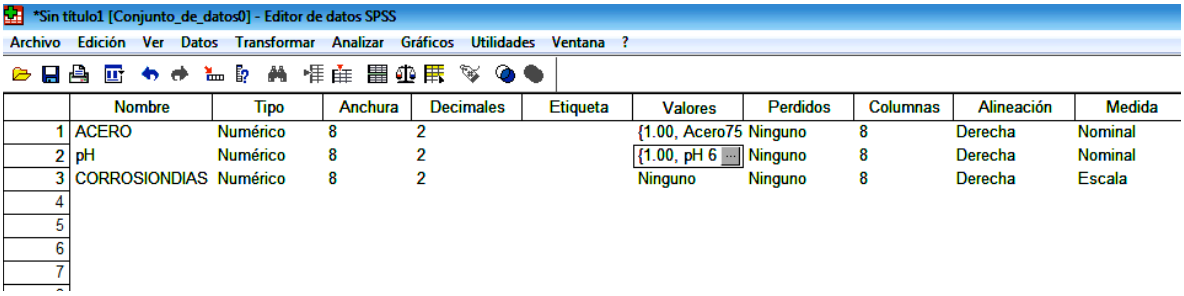


Aceptar

3. Ir a Valores de pH y digitar
   1 para pH 6
   2 para pH 5
   3 para pH 4



Aceptar

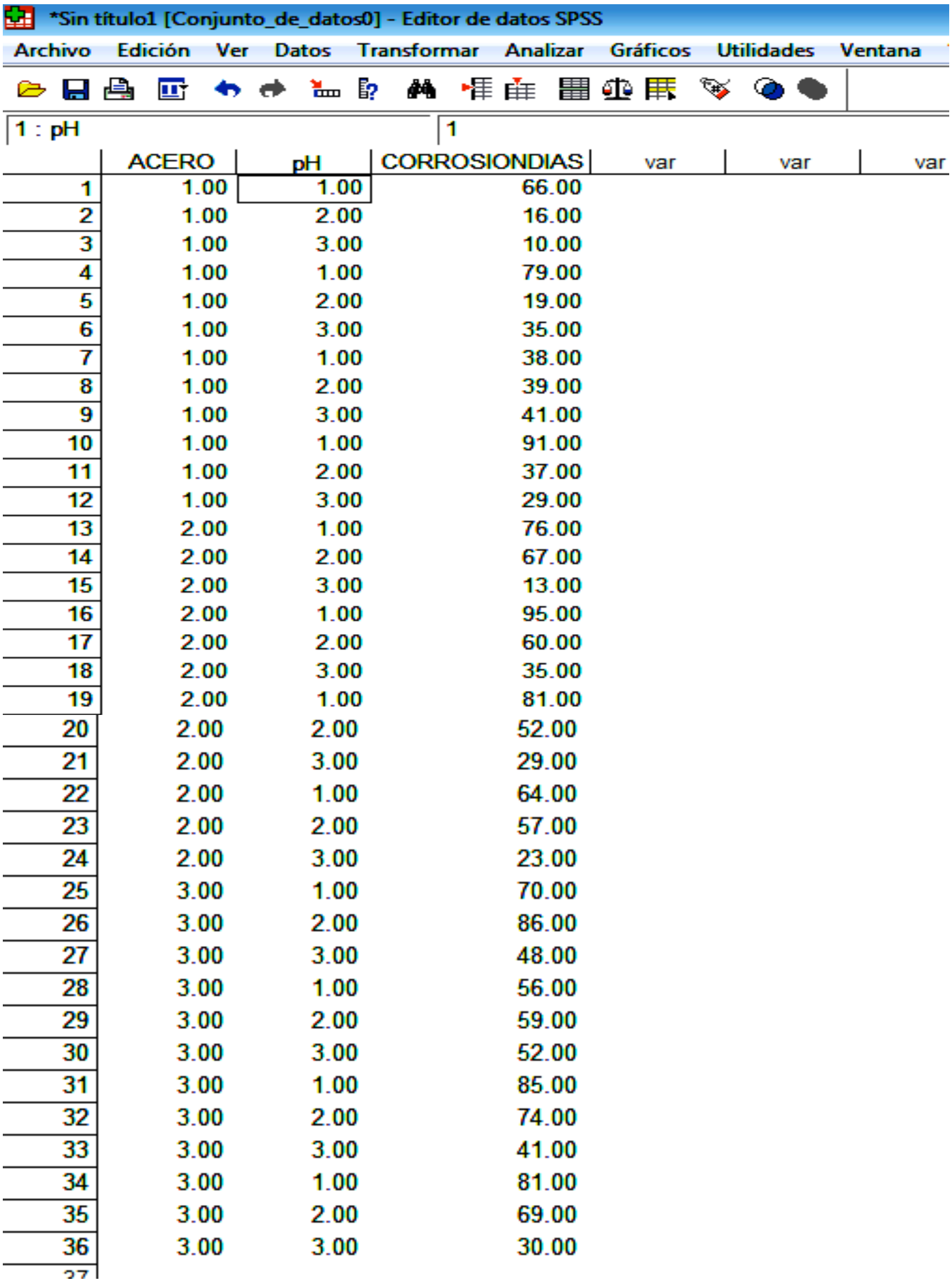


Ir a vistas de Datos y llenar la información del problema

- En la Columna acero colocar 1 (Acero 75%) . Repertilo 12 veces.
  Luego colocar 2 (Acero 85%). Repetirlo 12 veces
  Luego Colocar 3 (Acero 95%). Repetirlo 12 veces

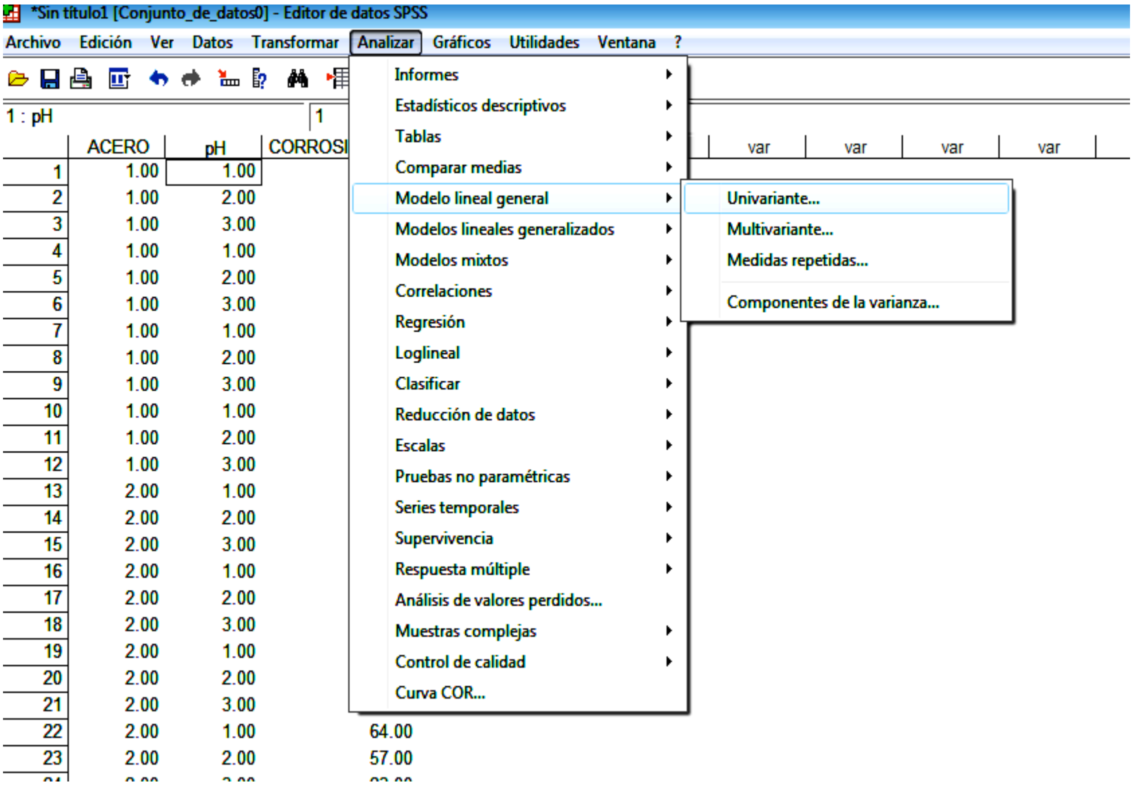
- En la Columna pH colocar 1 (pH 6), 2 (pH 5) y 3 (pH 4) . Repetirlo 4 veces (Acero 75%, Acero 85% y Acero 95%)

- En la Columna Corrosión días car los valores

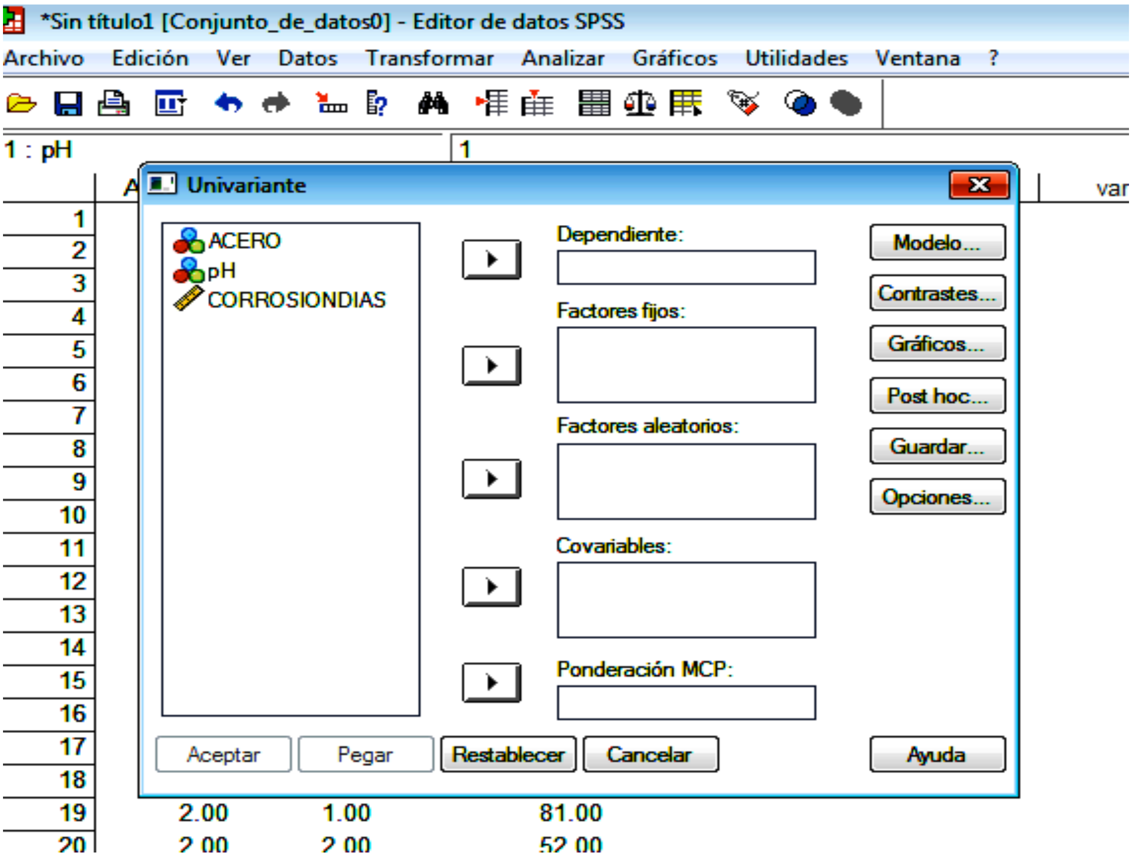


	ACERO	pH	CORROSIONDIAS	var	var	var
1	1.00	1.00	66.00			
2	1.00	2.00	16.00			
3	1.00	3.00	10.00			
4	1.00	1.00	79.00			
5	1.00	2.00	19.00			
6	1.00	3.00	35.00			
7	1.00	1.00	38.00			
8	1.00	2.00	39.00			
9	1.00	3.00	41.00			
10	1.00	1.00	91.00			
11	1.00	2.00	37.00			
12	1.00	3.00	29.00			
13	2.00	1.00	76.00			
14	2.00	2.00	67.00			
15	2.00	3.00	13.00			
16	2.00	1.00	95.00			
17	2.00	2.00	60.00			
18	2.00	3.00	35.00			
19	2.00	1.00	81.00			
20	2.00	2.00	52.00			
21	2.00	3.00	29.00			
22	2.00	1.00	64.00			
23	2.00	2.00	57.00			
24	2.00	3.00	23.00			
25	3.00	1.00	70.00			
26	3.00	2.00	86.00			
27	3.00	3.00	48.00			
28	3.00	1.00	56.00			
29	3.00	2.00	59.00			
30	3.00	3.00	52.00			
31	3.00	1.00	85.00			
32	3.00	2.00	74.00			
33	3.00	3.00	41.00			
34	3.00	1.00	81.00			
35	3.00	2.00	69.00			
36	3.00	3.00	30.00			

4. Ir a Analizar - Modelo lineal General - Univariante

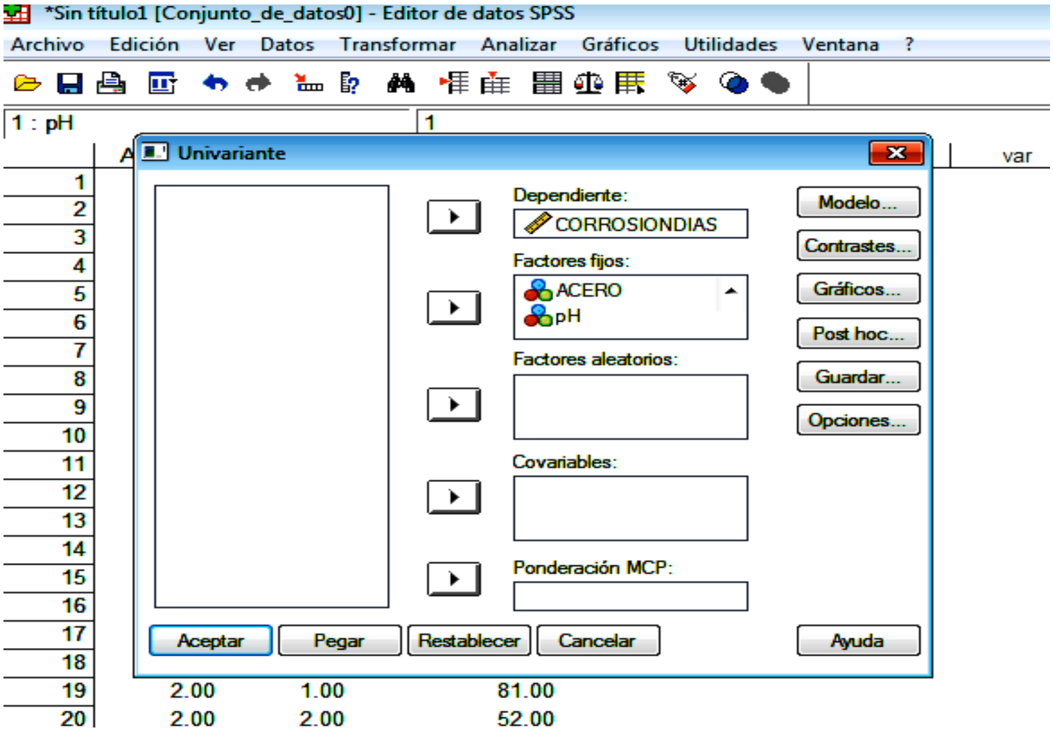


Click

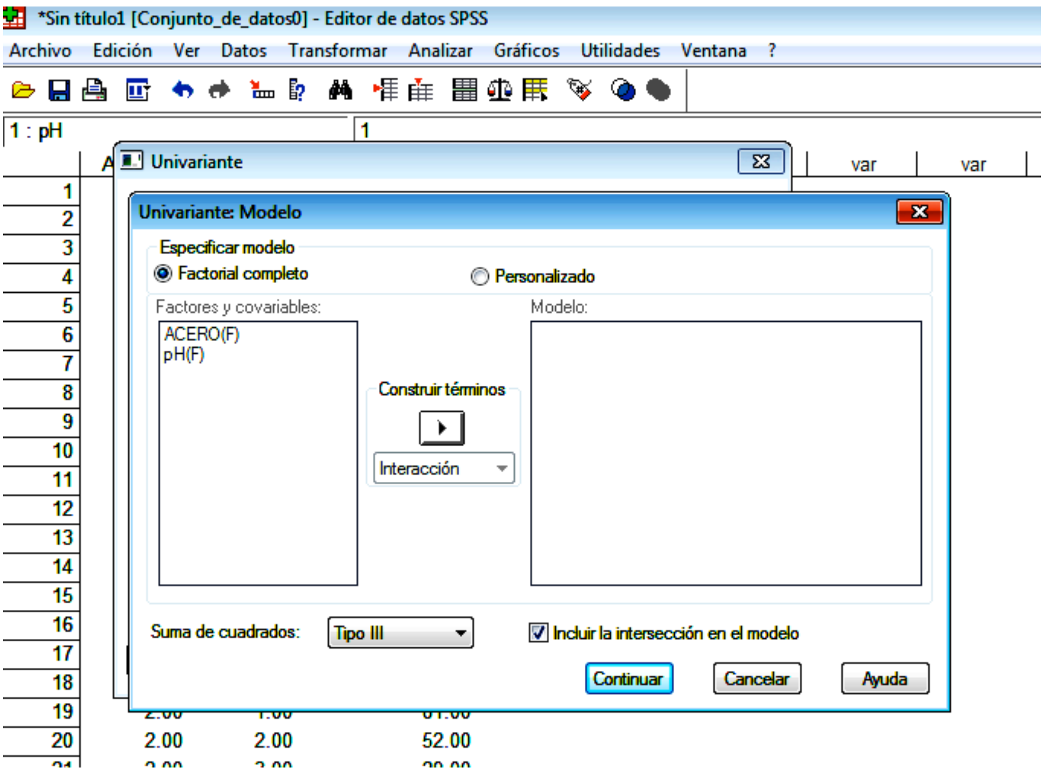


5. Pasar Corrosión días a Variable dependiente

Pasar Acero y pH valores fijos

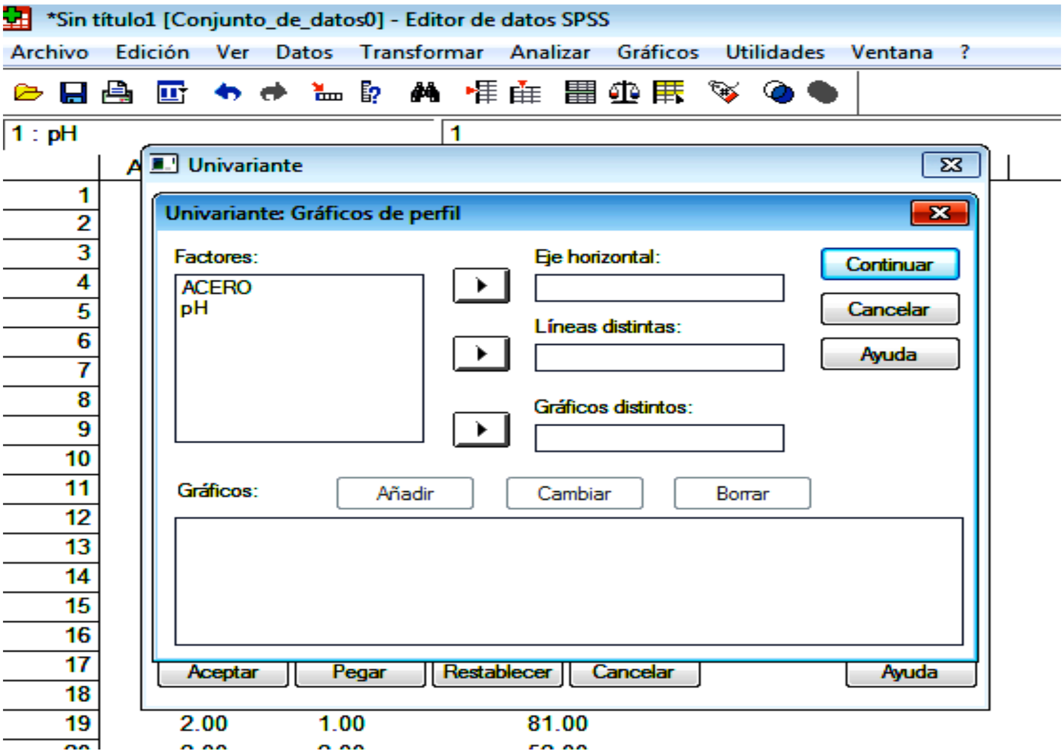


6. Ir a Modelo

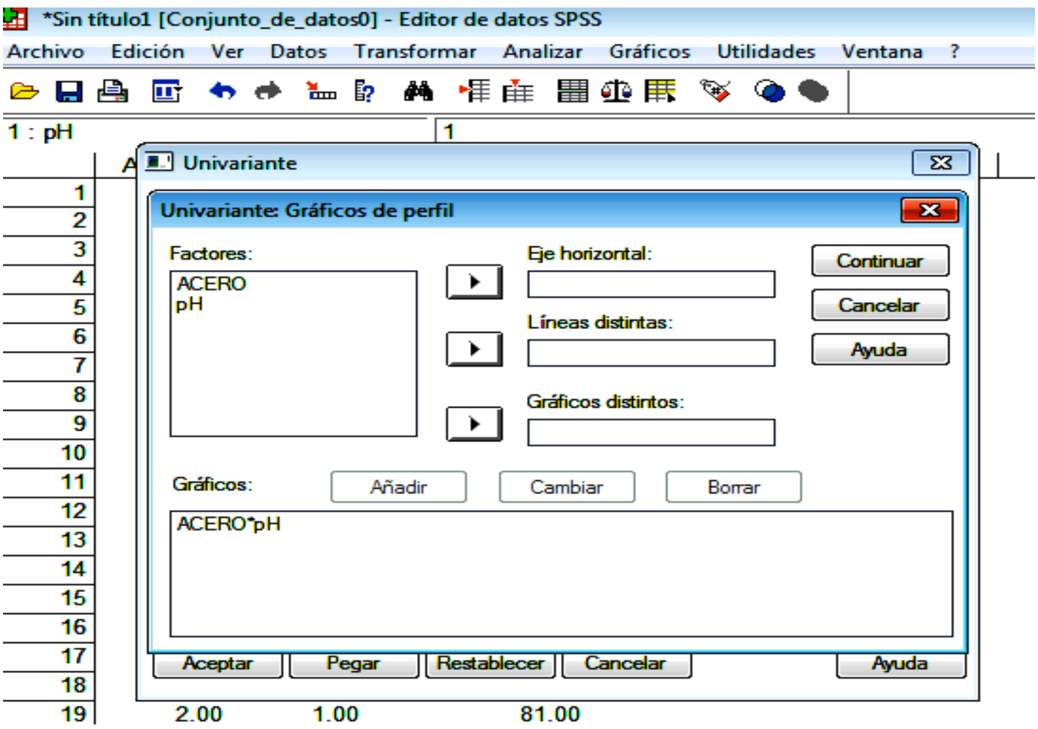


Continuar

7. Ir a gráficos



Pasa Factor Acero a Eje Horizontal
Pasar Factor pH a Lineas distintas
Luego añadir



Continuar y Aceptar

Análisis de varianza univariante

[Conjunto_de_datos0]

Factores inter-sujetos

		Etiqueta del valor	N
ACERO	1.00	Acero75%	12
	2.00	Acero 85%	12
	3.00	Acero 95%	12
pH	1.00	pH 6	12
	2.00	pH 5	12
	3.00	pH 4	12

Pruebas de los efectos inter-sujetos

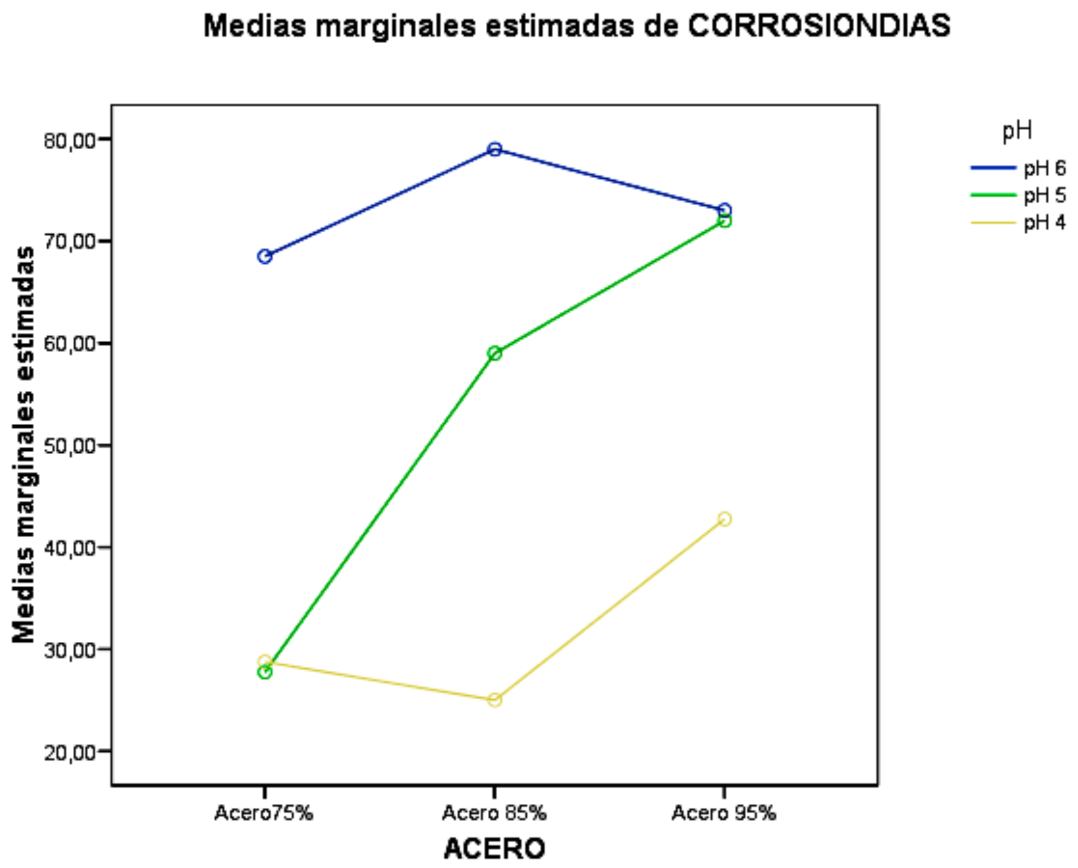
Variable dependiente: CORROSIONDIAS

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación
Modelo corregido	15311.056[a]	8	1913.882	11.334	.000
Intersección	100594.694	1	100594.694	595.724	.000
ACERO	2664.056	2	1332.028	7.888	.002
pH	10250.722	2	5125.361	30.353	.000
ACERO * pH	2396.278	4	599.069	3.548	.019
Error	4559.250	27	168.861		
Total	120465.000	36			
Total corregida	19870.306	35			

a. R cuadrado = .771 (R cuadrado corregida = .703)

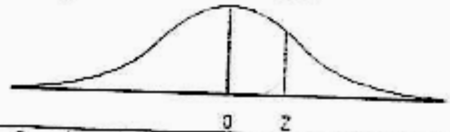
CONCLUSIÓN: El pH es el que más incide en la fabricación del reactor.

Gráficos de perfil



TABLAS

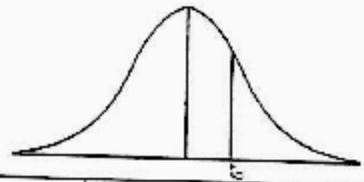
Arias bajo la Curva Normal Tipificada de 0 a Z



Z	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0754
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2258	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2996	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4223	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	04992	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4995	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4996	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4997	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4998	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Percentiles (t_p) de la distribución t de Student con v grados de libertad

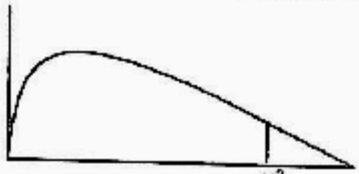
$1 - \frac{\alpha}{2} = \frac{1-0{,}05}{2} =$



V	$t_{0.995}$	$t_{0.99}$	$t_{0.975}$	$t_{0.95}$	$T_{0.90}$	$t_{0.80}$	$t_{0.75}$	$t_{0.70}$	$t_{0.60}$	$t_{0.55}$
1	63.66	31.82	12.71	6.31	3.08	1.376	1.000	0.727	0.325	0.158
2	9.92	6.96	4.30	2.92	1.89	1.061	0.816	0.617	0.289	0.142
3	5.84	4.54	3.18	2.35	1.64	0.978	0.765	0.584	0.277	0.137
4	4.60	3.75	2.78	2.13	1.53	0.941	0.741	0.569	0.271	0.134
5	4.03	3.36	2.57	2.02	1.48	0.920	0.727	0.559	0.267	0.132
6	3.71	3.14	2.45	1.94	1.44	0.906	0.718	0.553	0.265	0.131
7	3.50	3.00	2.36	1.90	1.42	0.896	0.711	0.549	0.263	0.130
8	3.36	2.90	2.31	1.86	1.40	0.889	0.706	0.546	0.262	0.130
9	3.25	2.82	2.26	1.83	1.38	0.883	0.703	0.543	0.261	0.129
10	3.17	2.76	2.23	1.81	1.37	0.879	0.700	0.542	0.260	0.129
11	3.11	2.72	2.20	1.80	1.36	0.876	0.697	0.540	0.260	0.129
12	3.06	2.68	2.18	1.78	1.36	0.873	0.695	0.539	0.259	0.128
13	3.01	2.65	2.16	1.77	1.35	0.870	0.694	0.538	0.259	0.128
14	2.98	2.62	2.14	1.76	1.34	0.868	0.692	0.537	0.258	0.128
15	2.95	2.60	2.13	1.75	1.34	0.866	0.691	0.536	0.258	0.128
16	3.92	2.58	2.12	1.75	1.34	0.865	0.690	0.535	0.258	0.128
17	2.90	2.57	2.11	1.74	1.33	0.863	0.689	0.534	0.257	0.128
18	2.88	2.55	2.10	1.73	1.33	0.862	0.688	0.534	0.257	0.127
19	2.86	2.54	2.09	1.73	1.33	0.861	0.688	0.533	0.257	0.127
20	2.84	2.53	2.09	1.72	1.32	0.860	0.687	0.533	0.257	0.127
21	2.83	2.52	2.08	1.72	1.32	0.859	0.686	0.532	0.257	0.127
22	2.82	2.51	2.07	1.72	1.32	0.858	0.686	0.532	0.256	0.127
23	2.81	2.50	2.07	1.71	1.32	0.858	0.685	0.532	0.256	0.127
24	2.80	2.49	2.06	1.71	1.32	0.857	0.685	0.531	0.256	0.127
25	2.79	2.48	2.06	1.71	1.32	0.856	0.684	0.531	0.256	0.127
26	2.78	2.48	2.06	1.71	1.32	0.856	0.684	0.531	0.256	0.127
27	2.77	2.47	2.05	1.70	1.31	0.855	0.684	0.531	0.256	0.127
28	2.76	2.47	2.05	1.70	1.31	0.855	0.683	0.530	0.256	0.127
29	2.76	2.46	2.04	1.70	1.31	0.854	0.683	0.530	0.256	0.127
30	2.75	2.46	2.04	1.70	1.31	0.854	0.683	0.530	0.256	0.127
40	2.70	2.42	2.02	1.68	1.30	0.851	0.681	0.529	0.255	0.126
60	2.66	2.39	2.00	1.67	1.30	0.848	0.679	0.527	0.254	0.126
120	2.62	2.36	1.98	1.66	1.29	0.845	0.677	0.526	0.254	0.126
∞	2.58	2.33	1.96	1.645	1.28	0.842	0.674	0.524	0.253	0.126

Percentiles (x^2_p) de la distribución Chi Cuadrada con v grados de libertad

DE LA
DISTRIBUCIÓN CHI-CHUADRADO
CON v GRADOS DE LIBERTAD
(AREA COMPREADA = P)



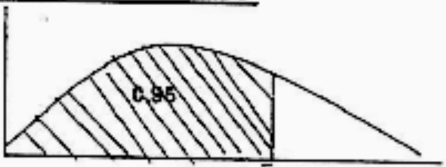
χ^2_p

v	$\chi^2_{0.995}$	$\chi^2_{0.99}$	$\chi^2_{0.975}$	$\chi^2_{0.95}$	$\chi^2_{0.90}$	$\chi^2_{0.75}$	$\chi^2_{0.50}$	$\chi^2_{0.25}$	$\chi^2_{0.10}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$	$\chi^2_{0.01}$	$\chi^2_{0.005}$
1	7.88	6.63	5.02	3.84	2.71	1.32	0.455	0.102	0.0158	0.0039	0.0010	0.0002	0.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	0.575	0.211	0.103	0.0506	0.0201	0.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	0.584	0.352	0.216	0.115	0.072
4	14.9	13.3	11.1	9.49	7.78	5.39	3.36	1.92	1.06	0.711	0.484	0.297	0.207
5	16.7	15.1	12.8	11.1	9.24	6.63	4.35	2.67	1.61	1.15	0.831	0.554	0.412
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	0.872	0.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.17	1.69	1.24	0.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	10.3	7.58	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.3	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.3	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.1	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.8	30.6	27.5	25.0	22.3	18.2	14.3	11.0	8.55	7.26	6.26	5.23	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	7.96	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.7	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.4	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.9	35.5	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.03
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89
25	46.9	44.3	40.6	37.7	34.4	29.3	24.3	19.9	16.5	14.6	13.1	11.5	10.5
26	48.3	45.6	41.9	38.9	35.6	30.4	25.3	20.8	17.3	15.4	13.8	12.2	11.2
27	49.6	47.0	43.2	40.1	36.7	31.5	26.3	21.7	18.1	16.2	14.6	12.9	11.8
28	51.0	48.3	44.5	41.3	37.9	32.6	27.3	22.7	18.9	16.9	15.3	13.6	12.5
29	52.3	49.6	45.7	42.6	39.1	33.7	28.3	33.6	19.8	17.7	16.0	14.3	13.1
30	53.7	50.9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66.8	63.7	59.3	55.8	51.8	45.6	39.3	33.7	29.1	26.5	24.4	22.2	20.7
50	79.5	76.2	71.4	67.5	63.2	56.3	49.3	42.9	37.7	34.8	32.4	29.7	28.0
60	92.0	88.4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104.2	100.4	95.0	90.5	85.5	77.6	69.3	61.7	55.3	51.7	48.8	45.4	43.3
80	166.3	112.3	106.6	101.9	96.6	88.1	79.3	71.1	64.3	60.4	57.2	53.5	51.2
90	128.3	124.1	118.1	113.1	107.6	98.6	89.3	80.6	73.3	69.1	65.6	61.6	59.2
100	140.2	135.8	129.6	124.3	118.5	109.1	99.3	90.1	82.4	77.9	74.2	70.1	67.3

Bandas $f_{1-\alpha;m1;m2}$ de la distribución F

Para una probabilidad de seguridad 1 - α

$P(F \leq f_{1-\alpha;m1;m2}) = F_f(f_{1-\alpha;m1;m2}) = 1 - \alpha$



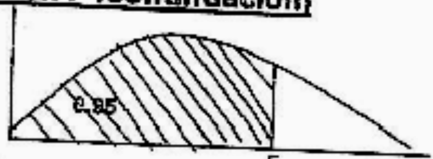
$F_{m1,m2;0.95}$

m2 \ m1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	162	200	216	225	230	234	237	239	241	242	243	244	245	245	246
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22	2.20	2.18
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20	2.17	2.15
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.18	2.15	2.13
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.15	2.13	2.11
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14	2.11	2.09
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.12	2.09	2.07
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13	2.10	2.08	2.06
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	2.09	2.06	2.04
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.08	2.05	2.03
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97	1.95	1.92
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.92	1.89	1.87
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.89	1.86	1.84
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.93	1.89	1.86	1.84	1.81
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.91	1.88	1.84	1.82	1.79
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.90	1.86	1.83	1.80	1.78
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.82	1.79	1.77
200	3.88	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80	1.77	1.74	1.72
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.77	1.74	1.71	1.69
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.72	1.69	1.67

Bandas $f_{1-\alpha;m1;m2}$ de la distribución F (continuación)

Para una probabilidad de seguridad $1 - \alpha$

$P(F \le f_{1-\alpha;m1;m2}) = F_f\,(f_{1-\alpha;m1;m2}) = 1 - \alpha$

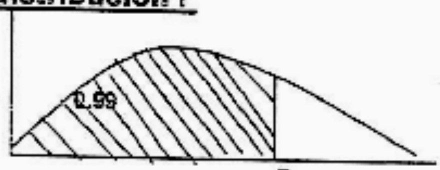


m2 \ m1	16	17	18	19	20	30	40	50	60	70	80	100	200	500	∞
1	246	247	247	248	248	250	251	252	252	252	253	253	254	254	254
2	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5
3	8.69	8.68	8.67	8.67	8.66	8.62	8.60	8.58	8.57	8.57	8.56	8.55	8.54	8.53	8.53
4	5.84	5.83	5.82	5.81	5.80	5.75	5.72	5.70	5.69	5.66	5.67	5.66	5.65	5.64	5.63
5	4.60	4.59	4.58	4.57	4.56	4.50	4.46	4.44	4.43	4.42	4.41	4.41	4.39	4.37	4.37
6	3.92	3.91	3.90	3.88	3.87	3.81	3.77	3.75	3.74	3.73	3.72	3.71	3.69	3.68	3.67
7	3.49	3.48	3.47	3.46	3.44	3.38	3.34	3.32	3.30	3.29	3.29	3.27	3.25	3.24	3.23
8	3.20	3.19	3.17	3.16	3.15	3.08	3.04	3.02	3.01	2.89	2.99	2.97	2.95	2.94	2.93
9	2.99	2.97	2.96	2.95	2.94	2.86	2.83	2.80	2.79	2.78	2.77	2.76	2.73	2.72	2.71
10	2.83	2.81	2.80	2.79	2.77	2.70	2.66	2.64	2.62	2.61	2.60	2.59	2.56	2.55	2.54
11	2.70	2.68	2.67	2.66	2.65	2.57	2.53	2.51	2.49	2.48	2.47	2.46	2.43	2.42	2.40
12	2.60	2.58	2.57	2.56	2.54	2.47	2.43	2.40	2.38	2.37	2.36	2.35	2.32	2.31	2.30
13	2.51	2.50	2.48	2.47	2.46	2.38	2.34	2.31	2.30	2.28	2.27	2.26	2.23	2.22	2.21
14	2.44	2.43	2.41	2.40	2.39	2.31	2.27	2.24	2.22	2.21	2.20	2.19	2.16	2.14	2.13
15	2.38	2.37	2.35	2.34	2.33	2.25	2.20	2.18	2.16	2.15	2.14	2.12	2.10	2.08	2.07
16	2.33	2.32	2.30	2.29	2.28	2.18	2.15	2.12	2.11	2.09	2.08	2.07	2.04	2.02	2.01
17	2.29	2.27	2.26	2.24	2.23	2.15	2.10	2.08	2.06	2.04	2.03	2.02	1.99	1.97	1.96
18	2.25	2.23	2.22	2.20	2.19	2.11	2.06	2.04	2.02	2.00	1.99	1.98	1.95	1.93	1.92
19	2.21	2.20	2.18	2.17	2.16	2.07	2.03	2.00	1.98	1.97	1.96	1.94	1.91	1.89	1.88
20	2.18	2.17	2.15	2.14	2.12	2.04	1.99	1.97	1.95	1.93	1.92	1.91	1.88	1.86	1.84
21	2.16	2.14	2.12	2.11	2.10	2.01	1.96	1.94	1.92	1.90	1.89	1.88	1.84	1.82	1.81
22	2.13	2.11	2.10	2.08	2.07	1.98	1.94	1.91	1.89	1.86	1.86	1.85	1.82	1.80	1.78
23	2.11	2.09	2.08	2.06	2.05	1.96	1.91	1.88	1.86	1.85	1.84	1.82	1.79	1.77	1.76
24	2.09	2.07	2.05	2.04	2.03	1.94	1.89	1.86	1.84	1.83	1.82	1.80	1.77	1.75	1.73
25	2.07	2.05	2.04	2.02	2.01	1.92	1.87	1.84	1.82	1.81	1.80	1.78	1.75	1.73	1.71
26	2.05	2.03	2.02	2.00	1.99	1.90	1.85	1.82	1.80	1.79	1.78	1.76	1.73	1.71	1.69
27	2.04	2.02	2.00	1.99	1.97	1.88	1.84	1.81	1.79	1.77	1.76	1.74	1.71	1.69	1.67
28	2.02	2.00	1.99	1.97	1.96	1.87	1.82	1.79	1.77	1.75	1.74	1.73	1.69	1.57	1.65
29	2.01	1.99	1.97	1.96	1.94	1.85	1.81	1.77	1.75	1.74	1.73	1.71	1.67	1.65	1.64
30	1.99	1.98	1.96	1.95	1.93	1.84	1.79	1.76	1.74	1.72	1.71	1.70	1.66	1.64	1.62
40	1.90	1.89	1.87	1.85	1.84	1.74	1.69	1.66	1.64	1.62	1.61	1.59	1.55	1.53	1.51
50	1.85	1.83	1.81	1.80	1.78	1.69	1.63	1.60	1.58	1.56	1.54	1.52	1.46	1.46	1.44
60	1.82	1.80	1.78	1.76	1.75	1.65	1.59	1.56	1.53	1.52	1.50	1.48	1.44	1.41	1.39
70	1.79	1.77	1.75	1.74	1.72	1.62	1.57	1.53	1.50	1.49	1.47	1.45	1.40	1.37	1.35
80	1.77	1.75	1.73	1.72	1.70	1.60	1.54	1.51	1.48	1.46	1.45	1.43	1.38	1.35	1.32
90	1.76	1.74	1.72	1.70	1.69	1.59	1.53	1.49	1.46	1.44	1.43	1.41	1.36	1.33	1.30
100	1.75	1.73	1.71	1.69	1.68	1.57	1.52	1.48	1.45	1.43	1.41	1.39	1.34	1.31	1.28
200	1.69	1.67	1.66	1.64	1.62	1.52	1.46	1.41	1.39	1.36	1.35	1.32	1.26	1.22	1.19
500	1.66	1.64	1.62	1.61	1.59	1.48	1.42	1.38	1.34	1.32	1.30	1.28	1.21	1.16	1.11
∞	1.64	1.62	1.60	1.59	1.57	1.46	1.39	1.35	1.32	1.29	1.27	1.24	1.17	1.11	1.00

Bandas $f_{1-\alpha;m1;m2}$ de la distribución F

Para una probabilidad de seguridad 1 - α

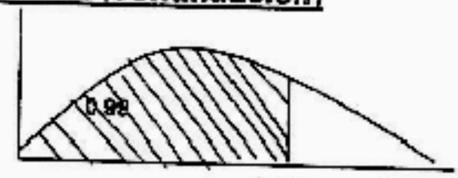
$P(F \leq f_{1-\alpha;m1;m2}) = F_f(f_{1-\alpha;m1;m2}) = 1 - \alpha$



m2 \ m1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	4052	4999	5403	5625	5764	5859	5928	5981	6023	6056	6083	6106	6126	6143	6157
2	98.5	99.0	99.2	99.3	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4
3	34.1	30.8	29.4	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	27.1	27.0	26.9	26.9
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.5	14.4	14.3	14.2	14.2
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89	9.82	9.77	9.72
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.66	7.60	7.56
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.41	6.36	6.31
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.61	5.56	5.52
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	5.05	5.01	4.96
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.65	4.60	4.56
11	9.64	7.20	6.21	5.67	5.31	5.07	4.88	4.74	4.63	4.54	4.46	4.39	3.34	4.29	4.25
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.10	4.05	4.01
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.90	3.86	3.82
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.75	3.70	3.66
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.61	3.56	3.52
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.50	3.45	3.41
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46	3.40	3.35	3.31
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.32	3.27	3.23
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.24	3.19	3.15
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.18	3.13	3.09
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.24	3.17	3.12	3.07	3.03
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.07	3.02	2.98
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	3.02	2.97	2.93
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03	2.98	2.93	2.89
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	3.06	2.99	2.94	2.89	2.85
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96	2.90	2.86	2.81
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.99	2.93	2.87	2.82	2.78
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90	2.84	2.79	2.75
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.93	2.87	2.81	2.77	2.73
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.79	2.74	2.70
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.61	2.56	2.52
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.62	2.56	2.51	2.46	2.42
60	7.07	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.44	2.39	2.35
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.51	2.45	2.40	2.35	2.31
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.48	2.42	2.36	2.31	2.27
90	6.92	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52	2.45	2.39	2.33	2.29	2.24
100	6.89	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.43	2.37	2.31	2.27	2.22
200	6.75	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27	2.22	2.17	2.13
500	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36	2.28	2.22	2.17	2.12	2.07
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18	2.13	2.08	2.04

Bandas $f_{1-\alpha;m1;m2}$ de la distribución F(continuación)

Para una probabilidad de seguridad $1 - \alpha$
$P(F \leq f_{1-\alpha;m1;m2}) = F_f(f_{1-\alpha;m1;m2}) = 1 - \alpha$



m2 \ m1	16	17	18	19	20	30	40	50	60	70	80	100	200	500	∞
1	6169	6182	6192	6201	6209	6261	6287	6303	6313	6320	6326	6335	6350	6361	6366
2	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
3	26.8	26.8	26.8	26.7	26.7	26.5	26.4	26.4	26.3	26.3	26.3	26.2	26.2	26.1	26.1
4	14.2	14.1	14.1	14.0	14.0	13.8	13.7	13.7	13.7	13.6	13.6	13.6	13.5	13.5	13.5
5	9.68	9.64	9.61	9.58	9.55	9.38	9.29	9.24	9.20	9.18	9.16	9.13	9.08	9.04	9.02
6	7.52	7.48	7.45	7.42	7.40	7.23	7.14	7.09	7.06	7.03	7.01	6.99	6.93	6.90	6.88
7	6.28	6.24	6.21	6.18	6.16	5.99	5.91	5.86	5.82	5.80	5.78	5.75	5.70	5.67	5.65
8	5.48	5.44	5.41	5.38	5.36	5.20	5.12	5.07	5.03	5.01	4.99	4.96	4.91	4.88	4.86
9	4.92	4.89	4.86	4.83	4.81	4.65	4.57	4.52	4.48	4.46	4.44	4.41	4.36	4.33	4.31
10	4.52	4.49	4.46	4.43	4.41	4.25	4.17	4.12	4.08	4.06	4.04	4.01	3.96	3.93	3.91
11	4.21	4.18	4.15	4.12	4.10	3.94	3.86	3.81	3.77	3.75	3.73	3.70	3.65	3.62	3.60
12	3.97	3.94	3.91	3.88	3.86	3.70	3.62	3.57	3.54	3.51	3.49	3.47	3.41	3.38	3.36
13	3.78	3.74	3.72	3.69	3.66	3.51	3.42	3.37	3.34	3.32	3.30	3.27	3.22	3.19	3.17
14	3.62	3.59	3.56	3.53	3.51	3.35	3.27	3.22	3.18	3.16	3.14	3.11	3.06	3.03	3.00
15	3.49	3.45	3.42	3.40	3.37	3.21	3.13	3.08	3.05	3.02	3.00	2.98	2.92	2.89	2.87
16	3.37	3.34	3.31	3.28	3.26	3.10	3.02	2.97	2.93	2.91	2.89	2.86	2.81	2.78	2.75
17	3.27	3.24	3.21	3.19	3.16	3.00	2.92	2.87	2.83	2.81	2.79	2.76	2.71	2.68	2.65
18	3.19	3.16	3.13	3.10	3.08	2.92	2.84	2.78	2.75	2.72	2.70	2.68	2.62	2.59	2.57
19	3.12	3.08	3.05	3.03	3.00	2.84	2.76	2.71	2.67	2.65	2.63	2.60	2.55	2.51	2.49
20	3.05	3.02	2.99	2.96	2.94	2.78	2.69	2.64	2.61	2.58	2.56	2.54	2.48	2.44	2.42
21	2.99	2.96	2.93	2.90	2.88	2.72	2.64	2.58	2.55	2.52	2.50	2.48	2.42	2.38	2.36
22	2.94	2.91	2.88	2.85	2.83	2.67	2.58	2.53	2.50	2.47	2.45	2.42	2.36	2.33	2.31
23	2.89	2.86	2.83	2.80	2.78	2.62	2.54	2.48	2.45	2.42	2.40	2.37	2.32	2.28	2.26
24	2.85	2.82	2.79	2.76	2.74	2.58	2.49	2.44	2.40	2.38	2.36	2.33	2.27	2.24	2.21
25	2.81	2.78	2.75	2.72	2.70	2.54	2.45	2.40	2.36	2.34	2.32	2.29	2.23	2.19	2.17
26	2.78	2.75	2.72	2.69	2.66	2.50	2.42	2.36	2.33	2.30	2.28	2.25	2.19	2.15	2.13
27	2.75	2.71	2.68	2.66	2.63	2.47	2.38	2.33	2.29	2.27	2.25	2.22	2.16	2.12	2.10
28	2.72	2.68	2.65	2.63	2.60	2.44	2.35	2.30	2.26	2.24	2.22	2.19	2.13	2.09	2.06
29	2.69	2.66	2.63	2.60	2.57	2.41	2.33	2.27	2.23	2.21	2.19	2.16	2.10	2.06	2.03
30	2.66	2.63	2.60	2.57	2.55	2.39	2.30	2.25	2.21	2.18	2.16	2.13	2.07	2.03	2.01
40	2.48	2.45	2.42	2.39	2.37	2.20	2.11	2.06	2.02	1.99	1.97	1.94	1.87	1.83	1.80
50	2.38	2.35	2.32	2.29	2.27	2.10	2.01	1.95	1.91	1.88	1.86	1.82	1.76	1.71	1.68
60	2.31	2.28	2.25	2.22	2.20	2.03	1.94	1.88	1.84	1.81	1.78	1.75	1.66	1.63	1.60
70	2.27	2.23	2.20	2.18	2.15	1.98	1.89	1.83	1.78	1.75	1.73	1.70	1.62	1.57	1.54
80	2.23	2.20	2.17	2.14	2.12	1.94	1.85	1.79	1.75	1.71	1.69	1.65	1.58	[illegible]	1.49
90	2.21	2.17	2.14	2.11	2.09	1.92	1.82	1.76	1.72	1.68	1.66	1.62	1.55	1.50	1.46
100	2.19	2.15	2.12	2.09	2.07	1.89	1.80	1.74	1.69	1.66	1.63	1.60	1.52	1.47	1.43
200	2.09	2.06	2.03	2.00	1.97	1.79	1.69	1.63	1.58	1.55	1.52	1.48	1.39	1.33	1.28
500	2.04	2.00	1.97	1.94	1.92	1.74	1.63	1.56	1.52	1.46	1.45	1.41	1.31	1.23	1.16
∞	2.00	1.97	1.93	1.90	1.88	1.70	1.59	1.52	1.47	1.43	1.40	1.36	1.25	1.15	1.00

BIBLIOGRAFIA

BIBLIOGRAFIA

1. **CANAVOS, C George.** Probabilidad y estadística para ingenieros. México: Mc Graw Hill. 1.992. 651 p.
2. **CHAO, Lincoln L.** Estadística para ciencias administrativas. México: Mc Graw Hill. 1.995. 454 p.
3. **GRAY, A William.** Probabilidad y estadística elementales. México: Continental. 1.976. 340 p
4. **KAZMIER, Leonard J.** Teoría y problemas de estadística aplicada a la administración y a la economía. México: Mc Graw Hill. 1978. 446 p.
5. **LARSON, Harold J.** Introducción a la teoría de las probabilidades e inferencia estadística. México: D.F Limusa. 1.978. 466 p.
6. **LIPSCHUTZ, Seymour.** Teoría y problemas de probabilidad. México: Mc Graw Hill. 1.978. 151 p.
7. **MARTINEZ, B Ciro.** Estadística comercial. Bogotá: Norma. 1981. 208 p.
8. **MILLER, Irwin- FREUND, John E.** Probabilidad y estadística para ingenieros. México: Prentice- Hall hispanoamericana s.a. 1.985. 574 p.
9. **ORDOÑEZ, P Hermelinda,** Inferencia estadística, regresión y muestreo. Manizales: Universidad Nacional de Colombia, sede Manizales. 2.000. 162 p.
10. **OSPINA, B David.** Muestreo básico y aplicaciones. Bogotá: Sección de publicaciones, Facultad de ciencias. Universidad Nacional de Colombia. 1.995. 127 p
11. **SANCHEZ, Javier I.** Estadística básica aplicada. Bucaramanga: Editora de libros técnicos ltda. 1976. 497p.
12. **SHAO; Stephen P.** Estadística para economistas y administradores de empresas. México: Herrero hermanos, sucesores s.a. 1.973. 786 p.
13. **WALPOLE, Ronald E.** Probabilidad y estadística para ingenieros. México: Interamericana s.a. 1987. 783 p.
14. **ARNALDOS, F. DÍAZ, M.T., FAURA, U., MOLERA, L., PARRA, I. -** (2003): "Estadística Descriptiva para economía y Administración de empresas". Thomson.
15. **CANAVOS, G.C.** (1988): "Probabilidad y estadística. Aplicaciones y métodos". Mc Graw Hill.
16. **MENDENHALL, W. - SINCICH, T.** (1997): "Probabilidad y estadística para ingeniería y ciencias". Prentice Hall.
17. **PEÑA, D.** (1991): "Estadística. Modelos y métodos". Alianza Universidad.
18. **SNEDECOR, G. Y COCHRAN, W.** (1974) Métodos Estadísticos CECSA México.
19. **PÉREZ, C.** (2002): "Estadística aplicada a través de Excel". Prentice Hall.

20. **SCHEAFFER - MCCLAVE.** (1993): "Probabilidad y estadística para ingeniería". Grupo Editorial Iberoamérica.
21. **PATROCINIO FLORES JUÁREZ GRUPO UNIVERSITARIO.** (2008): "SPSS V16. Grupo Editorial Universitaria.

Dr. Alejandro Néstor Salas Begazo

Ingeniero Químico, Doctor en Administración, Magister Scientiae en Administración y Finanzas, Segunda Especialidad en Ingeniería y Gestión Ambiental, Egresado de Maestría en Investigación y Docencia en Educación Superior.
Especialista en Medio Ambiente, Administración y Gestión de Empresas, Estadística, Investigación de Operaciones Unitarias y Procesos, Asesor de Proyectos de Investigación.
Docente Universitario Pre-Grado, Postgrado en Maestría y Doctorado, en la Universidad Andina Néstor Cáceres Velásquez - UANCV Filial Arequipa.
Docente Universitario Pre-Grado, Postgrado en Maestría y Doctorado, en la Universidad Nacional de San Agustín de Arequipa - UNSA en la Facultad de Ingeniería de Procesos - Escuela Profesional de Ingeniería Química

ISBN: 978-9942-33-046-8
9 789942 330468

compÁs