ESTADÍSTICA II

PROGRAMA ADMINISTRACIÓN PÚBLICA TERRITORIAL

RAFAEL VARGAS BARRERA

ESCUELA SUPERIOR DE ADMINISTRACIÓN PÚBLICA

ESCUELA SUPERIOR DE ADMINISTRACIÓN PÚBLICA

Director
HONORIO MIGUEL HENRIQUEZ PINEDO

Subdirector académico
CARLOS ROBERTO CUBIDES OLARTE

Decano de pregrado
JAIME ANTONIO QUICENO GUERRERO

Coordinador Nacional de A.P.T
JOSE PLACIDO SILVA RUIZ

ESCUELA SUPERIOR DE ADMINISTRACIÓN PÚBLICA
RAFAEL VARGAS BARRERA

Bogotá D.C., Noviembre de 2008

INDICE DE CONTENIDOS

6.1 Intervalos de confianza.
6.2 Estimación para la media poblacional. Muestras grandes.
6.3 Estimación para la media poblacional. Muestras pequeñas.
6.4 Estimación de una proporción de poblacional. Muestras grandes.
6.5 Estimación de una proporción de la población Muestras pequeñas.
6.6 Análisis de sensibilidad en Excel para la estimación del intervalo.

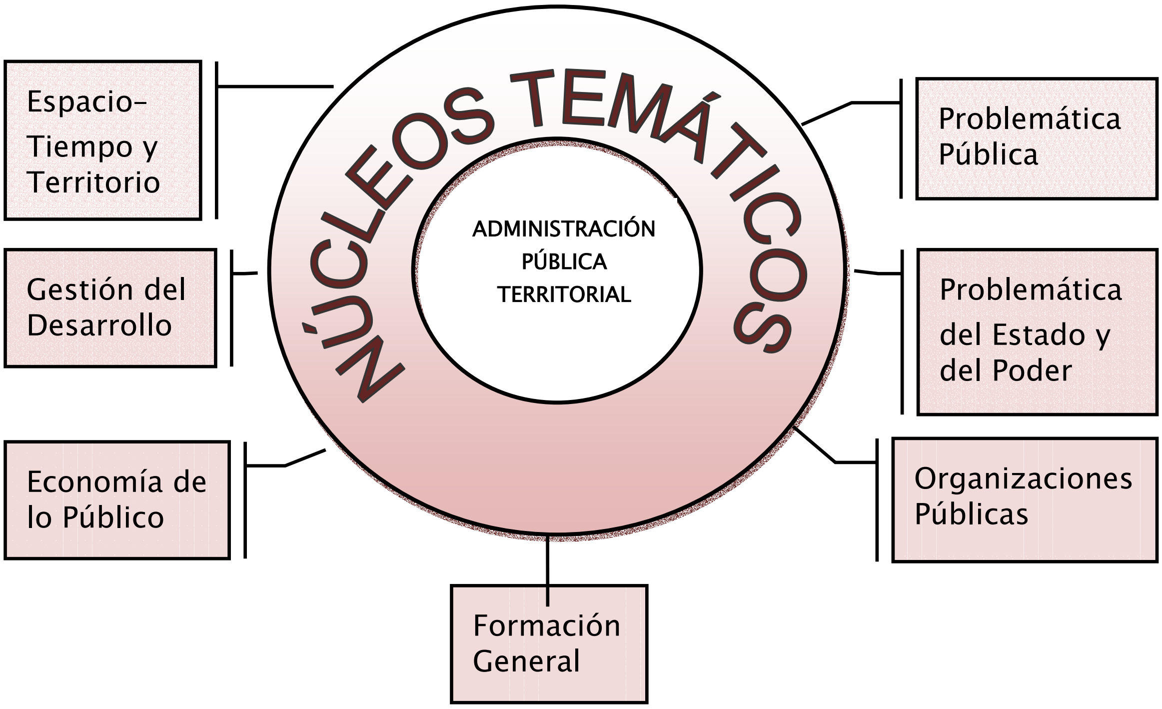
UNIDAD 7. PRUEBA DE HIPÓTESIS

7.1 Definición de hipótesis nula y alterna.
7.2 Tipos de errores.
7.3 Nivel de significancia
7.4 Pruebas de hipótesis unilaterales y bilaterales sobre la media
7.4.1 Muestras grandes
7.4.2 Muestras pequeñas
7.5 Prueba de hipótesis sobre la proporción de una población.
7.5.1 Muestras grandes
7.5.2 Muestras pequeñas
7.6 Calculo de la probabilidad para el error tipo II

UNIDAD 8. REGRESIÓN SIMPLE Y MÚLTIPLE.

8.1 Diagramas de dispersión.
8.2 Estimación del modelos de regresión lineal por el método de los mínimos cuadrados.
8.3 Coeficiente de correlación y de determinación.
8.4 Pronósticos por el modelo de regresión lineal.
8.5 Modelos de regresión no lineal.
8.6 Modelos de regresión múltiples.
8.7 Pronósticos

DE LOS NUCLEOS TEMÁTICOS Y PROBLEMÁTICOS



El plan de estudios del Programa de Administración Pública Territorial, modalidad a distancia, se encuentra estructurado en siete núcleos temáticos. Éstos, a su vez, se constituyen en los contenidos nucleares del plan de formación que, en la exposición didáctica del conocimiento, se acompañan de contenidos complementarios específicos.

Cada uno de los siete núcleos temáticos que componen el programa tiene una valoración relativa en número de créditos y, en consecuencia, varía también en el número de asignaturas que lo conjugan. El primer momento en cualquier proceso de formación ha de establecer las particularidades del programa, de ahí que sea necesario dar a conocer los núcleos temáticos con su respectiva valoración en número de créditos: Problemática pública, once (11) créditos; Problemática del estado y del poder, 23 créditos; Organizaciones públicas, 24 créditos; Espacio–tiempo y territorio, 22 créditos; Gestión del desarrollo, 16 créditos; Economía de lo público, 18 créditos; y Formación general, 21 créditos.

De igual manera, se debe reconocer que el plan de estudios se cimienta en el principio de la problematización. En otras palabras, la formación en Administración Pública Territorial parte del hecho de que la disciplina se encuentra en constante cambio teórico y práctico; lo cual genera, a su vez, problemas multifacéticos que implican la formación de profesionales con capacidad de comprender, explicar y resolver los distintos textos y contextos que conforman la administración pública.

EL TRABAJO DEL TUTOR

El tutor tendrá libertad de cátedra en cuanto a su posición teórica o ideológica frente a los contenidos del módulo, pero el desarrollo de los contenidos de los módulos son de obligatorio cumplimiento por parte de los tutores. Los Tutores podrán complementar los módulos con lecturas adicionales, pero lo obligatorio para el estudiante frente a la evaluación del aprendizaje son los contenidos de los módulos; es decir, la evaluación del aprendizaje deberá contemplar únicamente los contenidos de los módulos. Así mismo, la evaluación del Tutor deberá diseñarse para dar cuenta del cubrimiento de los contenidos del módulo.

El Tutor debe diseñar, planear y programar con suficiente anticipación las actividades de aprendizaje y los contenidos a desarrollar en cada sesión de tutoría (incluyendo la primera), y diseñar las actividades para todas las sesiones (una sesión es de cuatro horas tutoriales). También debe diseñar las estrategias de evaluación del trabajo estudiante que le permita hacer seguimiento del proceso de autoaprendizaje del estudiante. Los módulos (asignaturas) de APT son de dos créditos (16 horas de tutoría grupal presencial por crédito para un total de 32 horas), tres créditos (48 horas de tutoría grupal presencial) y de 4 créditos (64 horas de tutoría grupal presencial, distribuidas así:

MÓDULO DE ESTADÍSTICA II (3 créditos)						
No. Créditos	Horas por crédito	**Total horas Tutoría Grupal**	No. de sesiones	Horas por sesión	No. mínimo de encuentros tutoriales*	No. max. sesiones por encuentro
2	16	32	8	4	2	8
3	**16**	**48**	**12**	**4**	**3**	**12**
4	16	64	16	4	4	16

* El número de encuentros se programara de acuerdo con las distancias y costos de transporte de la Sede Territorial al CETAP, por ejemplo para los casos de los CETAP de Leticia, San Andrés, Mitú, Puerto Inírida y Puerto Carreño, se podrán programar un mínimo de dos encuentros para un módulo de 2 Créditos (16 horas por encuentro), tres encuentros para un módulo de 3 créditos y cuatro encuentros para un módulo de 4 créditos.
Encuentro: número de veces que se desplaza un Tutor a un CETAP para desarrollar un módulo.
Sesión: número de horas por cada actividad tutorial, por ejemplo: 8-12 a.m., 2-6 p.m., 6-10 p.m.

ESTADÍSTICA II

INTRODUCCION:
En el módulo anterior de estadística se vieron los conceptos y herramientas para recopilar datos, procesarlos, analizarlos y analizar los resultados. Generalmente estos datos corresponden a hechos cumplidos, pero a partir de ellos se debe mirar los que vendría hacia el futuro con el comportamiento de las mismas variables. Las organizaciones públicas y privadas toman decisiones permanentemente basándose en los resultados históricos; pero los resultados se verán en el futuro por lo que la toma de decisiones se vuelve incierta y llena de incertidumbre.
En este módulo ordenaremos nuestro conocimiento para darle valores a la posibilidad que ocurran esos hechos futuros y con otras herramientas estadísticas le daremos valor a la incertidumbre y la certeza que ocurran esos eventos.
La probabilidad y otros parámetros de medición hacia el futuro son elementos fundamentales en los ejecutivos privados y funcionarios públicos que tienen su hombro la responsabilidad de planear, dirigir y programar la ejecución de las actividades de su entorno, en beneficio de una sociedad, familia o empresa.

Objetivos
Formar el participante en la utilización de medias que analizan hacia el futuro los resultados de hechos. El asistente adquirirá los conceptos y herramientas que le permitirán ser un usuario casi permanente de las probabilidades, el muestreo, las estimaciones, las pruebas de hipótesis, los modelos de regresión lineal o lineal.

Objetivos específicos.

- Conceptualizar los principios teóricos básicos de la inferencia estadística.
- Aplicar las fases metodológicas de una investigación por muestreo.
- Estimar promedios poblacionales, totales, proporciones y varianzas
- Determinar tamaños de muestra.
- Revisar la fundamentación de los procedimientos de prueba de hipótesis.
- Estudiar la correlación de Pearson entre parejas de variables.
- Estudiar la fundamentación de los modelos de regresión como técnica de análisis econométrico.
- Analizar el comportamiento de series cronológicas económicas y sociales.
- Proyectar variables económicas utilizando los modelos de regresión lineal y no lineal.

Competencias que el alumno debe desarrollar
Comprender y aplicar la teoría de probabilidad como introducción a las técnicas estadísticas de inferencia y pronósticos.
Aplicar conceptos, métodos y técnicas para el manejo de la información numérica y no numérica relacionada con la administración de las organizaciones y el manejo de cifras económicas.
Utilizar hojas de cálculo para el procesamiento, análisis y presentación de grandes volúmenes de datos.

Inferir comportamiento de poblaciones (datos reales de un problema) a partir de muestras aleatorias.
Identificar la distribución de probabilidades, en las situaciones que se viven a diario en las empresas.
Utilizar correctamente un software estadístico, e interpretar acertadamente los resultados para la toma de decisión ante una situación real del mercadeo y la logística.

Mapa conceptual.

Metodología.

Evaluación

UNIDAD 1. ESTADÍSTICA DESCRIPTIVA.

Si uno ve el fútbol por televisión o escucha un noticiero por la radio o televisión, o lea algún periódico o revistas de negocios, se verá sometido a una gran cantidad de cifras a las que comúnmente se denomina estadísticas. Estas cifras pueden referirse a deportes, mercado de valores, desempleo, producción industrial o esperanza de vida.

A un dato numérico o valor aislado se le denomina dato, o valor, estadístico. El precio al cierre de acciones comunes de Ecopetrol es un dato estadístico. La utilidad de un negocio también es un valor estadístico. Las ventas totales al menudeo en un cierto mes, es asimismo un dato estadístico. A un conjunto de datos numéricos se le denomina estadística.

El estudio de las estadísticas tiene un significado mucho más amplio que la simple recopilación y publicación de hechos y datos numéricos. El estudio general de las estadísticas se define como la ciencia estadística o Estadística.

Estadística Ciencia que trata de la recopilación, organización, presentación, análisis e interpretación de datos numéricos (estadística) con el fin de realizar una toma de decisiones más efectiva.

Así como los abogados tienen "reglas de evidencia" y los contadores "prácticas de uso común", las personas que trabajan con datos numéricos siguen ciertos lineamientos estándares.

Las técnicas estadísticas se aplican de manera amplia en producción, en el almacén, en un estudio de mercados, en control de calidad, en la Bolsa de Valores Colombia, mercadotecnia, contabilidad, control de calidad, gestión pública y en otras actividades.

La estadística se divide en estadística descriptiva y estadística inferencial.

La estadística descriptiva se refiere a la organización, presentación y análisis de datos numéricos. Es un procedimientos empleados para organizar y resumir conjuntos de datos numéricos. Se dispone de técnicas estadísticas para organizar este tipo de datos en forma significativa. Algunos datos pueden organizarse en una distribución de frecuencias. Pueden utilizarse diversos tipos de gráficas para describir datos. Los promedios especializados, como la mediana, pueden calcularse para describir el valor central de un grupo de datos numéricos.

Estadística inferencial o también denominada inferencia Estadística y Estadística inductiva. Lo más importante con respecto a la Estadística inferencial es determinar algo acerca de una población. Una población puede estar formada por personas como todos los estudiantes inscritos en una universidad, todos los alumnos de una clase de contabilidad, o todos los reclusos de una prisión. Una población también puede estar formada por objetos, como las llantas producidas durante una semana en una fábrica, o todas las truchas que habitan en una presa. Una población también puede estar formada por un grupo de medidas, como podrían ser los salarios de los empleados, o las estaturas de los alumnos de un curso.

1.1 Tablas de frecuencia en Excel

Es la técnica para la presentar en forma organizada los datos. Agrupamiento de datos en categorías que muestren el número de observaciones de cada categoría.
Para preparar una tabla de frecuencias se debe establecer un conjunto de agrupamientos que se denominan clases. Una clase puede ser el valor de una cualidad o un valor numérico o un intervalo.
Las variables pueden ser cualitativas, cuantitativas, discretas o continuas
En las variables continuas, cada categoría (clase) tiene dos límites, un límite inferior declarado y un límite superior declarado. En práctica es común hacer que el límite inferior de la primera clase sea igual a la más baja observación, y hacer que todas las clases tengan el mismo ancho o amplitud.

En lo posible los intervalos o clases de frecuencias deben ser iguales. Los intervalos de clase desiguales ofrecen problemas al representarse en forma gráfica. Sin embargo, en algunos casos pueden ser necesarios intervalos desiguales de clase para evitar un gran número de clases vacías, o casi vacías.

$$\text{Intervalo de clase} = \frac{X_{MAX} - X_{MIN}}{m} \quad \text{donde m es el número de clases}$$

Simbología Básica:
n_i Frecuencia absoluta
h_i Frecuencia relativa
N_i Frecuencia absoluta acumulada
H_i Frecuencia relativa acumulada

Ejemplo 1

VARIABLE Xi : NUMERO DE CLIENTES VISITADOS POR LOS VENDEDORES DE UNA EMPRESA.

VENDEDOR N°:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N° DE CLIENTES	6	4	6	6	6	5	4	5	6	5	3	4	6	4	3	4	6	6	7	4

Xi	ni	hi	Ni	Hi
3	2	###	2	0.10
4	6	###	8	0.40
5	3	###	11	0.55
6	8	###	19	0.95
7	1	###	20	1.00
	20			

FUNCION ESTADISTICA EN EXCEL
=FRECUENCIA(B4:U4;B7:B11)

2 vendedores visitaron 3 clientes	El 10% de los vendedores visitaron 3 cliente	2 vendedores visitaron máximo 3 clientes	El 10% de los vendedores visitaron clientes o menos
6 vendedores visitaron 4 clientes	El 30% de los vendedores visitaron 4 cliente	8 vendedores visitaron máximo 4 clientes	El 40% de los vendedores visitaron clientes o menos
3 vendedores visitaron 5 clientes	El 15% de los vendedores visitaron 5 cliente	11 vendedores visitaron máximo 5 clientes	El 55% de los vendedores visitaron clientes o menos
8 vendedores visitaron 6 clientes	El 40% de los vendedores visitaron 6 cliente	19 vendedores visitaron máximo 6 clientes	El 95% de los vendedores visitaron clientes o menos
1 vendedores visitaron 7 clientes	El 5% de los vendedores visitaron 7 clientes	20 vendedores visitaron máximo 7 clientes	El 100% de los vendedores visitaron clientes o menos

PARA ANALIZAR LOS DASTOS UTILICE LA FUNCION CONCATENAR

Fuente: archivo Frecuencias.xls; hoja: Ejemplo 1

Ejemplo 2

VARIABLE Xi: SALARIOS DE LOS EMPLEADOS DE UNA EMPRESA (SMLM)

11.6	7.5	5.1	3.4	4.0	15.5	2.2	8.0	13.0	15.8	6.0	8.0	6.4	10.9	7.8	10.8	13.4	10.3	16.0	12.1
5.9	1.9	6.8	14.9	10.2	8.2	15.2	5.1	6.2	2.8	6.5	11.4	10.3	4.3	2.3	3.3	11.3	3.4	4.2	7.7
15.5	8.9	11.0	5.7	1.0	7.8	4.4	8.5	9.8	4.9	6.4	7.2	15.6	10.7	7.4	9.0	4.9	11.0	3.4	2.2
15.1	4.8	8.1	14.8	10.8	5.2	11.2	10.7	2.5	6.7	13.3	5.6	12.8	8.1	11.3	5.2	12.9	5.2	2.3	9.1
8.6	4.4	5.3	13.7	12.6	1.1	6.5	9.5	8.9	3.4	9.0	7.4	13.6	6.6	12.7	11.8	14.0	14.1	5.3	14.6
2.2	2.3	5.1	2.0	6.5	12.0	13.1	9.2	15.1	7.6	15.1	13.1	11.7	1.3	11.3	7.8	2.2	12.4	14.3	10.8
11.1	11.0	12.2	11.3	4.5	9.7	2.0	3.8	11.7	1.4	4.2	13.1	11.9	4.5	13.1	5.5	10.5	11.5	12.7	6.4

Maximo	16.0
Minimo	1.0
Rango	15.0
Clases	5.0
Amplituc	3.0

X'i-1	X'i	Xi	ni	hi	Ni	Hi
1.0	4.0	2.5	23	16.43%	23	16.43%
4.0	7.0	5.5	33	23.57%	56	40.00%
7.0	10.0	8.5	25	17.86%	81	57.86%
10.0	13.0	11.5	36	25.71%	117	83.57%
13.0	16.0	14.5	23	16.43%	140	100.00%
			140	100.00%		

23 empleados devengan entre 1 y 4 salarios
33 empleados devengan entre 4 y 7 salarios
25 empleados devengan entre 7 y 10 salarios
36 empleados devengan entre 10 y 13 salarios
23 empleados devengan entre 13 y 16 salarios

23 empleados devengan máximo 4 salarios
56 empleados devengan máximo 7 salarios
81 empleados devengan máximo 10 salarios
117 empleados devengan máximo 13 salarios
140 empleados devengan máximo 16 salarios

El 16.43% de los empleados devengan entre 1 y 4 salarios
El 23.57% de los empleados devengan entre 4 y 7 salarios
El 17.86% de los empleados devengan entre 7 y 10 salarios
El 25.71% de los empleados devengan entre 10 y 13 salarios
El 16.43% de los empleados devengan entre 13 y 16 salarios

El 16.43% de los empleados devengan máximo 4 salarios
El 40% de los empleados devengan máximo 7 salarios
El 57.86% de los empleados devengan máximo 10 salarios
El 83.57% de los empleados devengan máximo 13 salarios
El 100% de los empleados devengan máximo 16 salarios

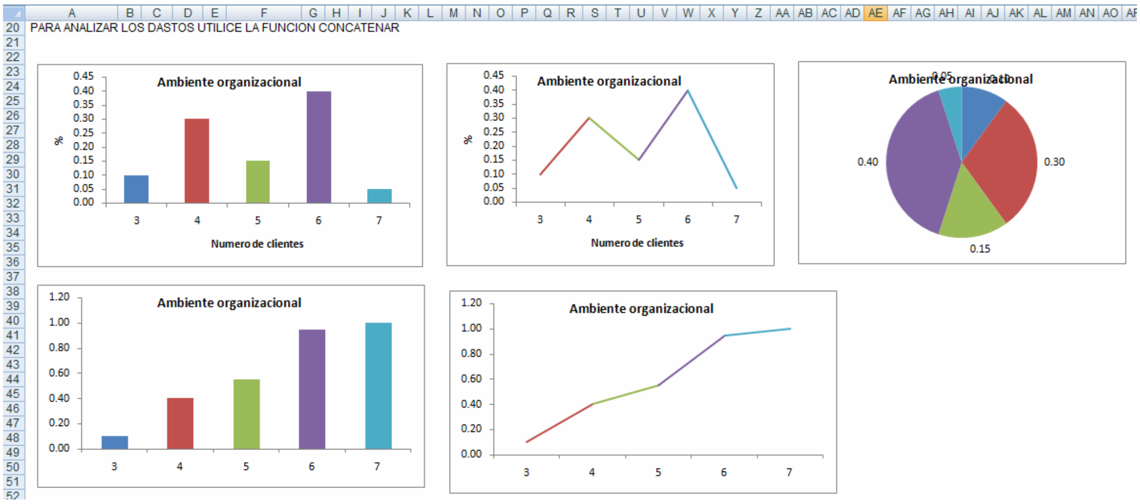
Ejemplo 1 | Ejemplo 2

Fuente: archivo Frecuencias.xls; hoja: Ejemplo 2

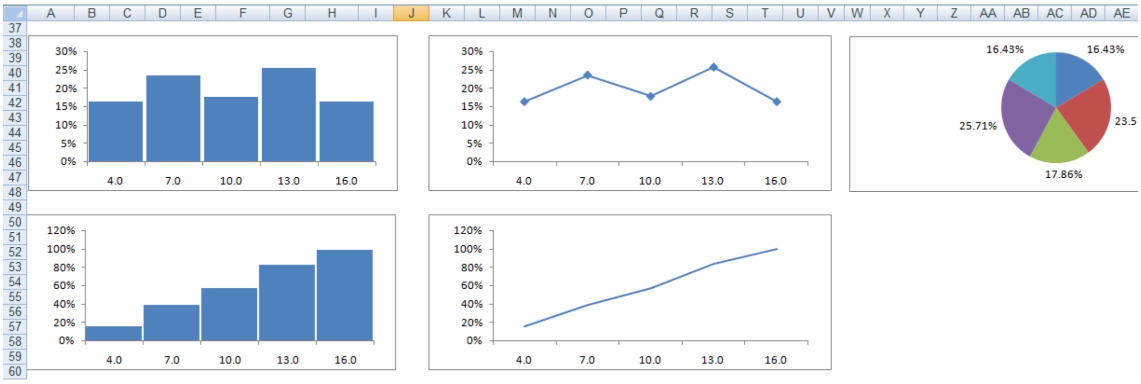
X'_{i-1}	Límite inferior del intervalo
X'_i	Límite superior del Intervalo
X_i	Marca de clase o valor medio del intervalo
n_i	Frecuencia absoluta
h_i	Frecuencia relativa
N_i	Frecuencia absoluta acumulada
H_i	Frecuencia relativa acumulada

1.2 Gráficas en Excel

Las frecuencias pueden mostrarse utilizando diagramas o gráficas. Tres diagramas que representan de manera adecuada, una distribución de frecuencias son el histograma, el polígono de frecuencias y diagrama de sectores.



Fuente: archivo Frecuencias.xls;   hoja: Ejemplo 1



Fuente: archivo Frecuencias.xls;   hoja: Ejemplo 2

1.3 Medidas tendencia central en Excel

Media aritmética. Es un valor tal que la suma de las desviaciones es igual a cero.

$$MEDIA\ ARITMETICA \quad ; \quad \bar{X} = \frac{X_i}{n} \quad ; \quad \bar{X} = \frac{\sum X_i * n_i}{n} = M[X]$$

Ejemplo 3

VARIABLE Xi : NUMERO DE CLIENTES VISITADOS POR LOS VENDEDORES DE UNA EMPRESA.

VENDEDOR N°:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N° DE CLIENTES	6	4	6	6	6	5	4	5	6	5	3	4	6	4	3	4	6	6	7	4

Media aritmética	5.00	=PROMEDIO(B4:U4)
Media aritmética	5.00	=D15/C15

Xi	ni	Xi*ni
3	2	6.00
4	6	24.00
5	3	15.00
6	8	48.00
7	1	7.00
	20	100

Fuente: archivo Medidas de posición.xls;   Hoja: Ejemplo 1

1.4 Medidas de dispersión en Excel.

Al aplicar una medida de dispersión es posible evaluar la confiabilidad del promedio que se está utilizando. Una dispersión pequeña indica que los datos se encuentran acumulados cercanamente, por ejemplo, alrededor de la media aritmética. Por tanto, la media se considera bastante representativa de los datos. Esto es, la media es un promedio confiable. Por el contrario, una dispersión grande indica que la media no es muy confiable, es decir, que no es muy representativa de los datos.

1.4.1 Varianza poblacional.

Media aritmética de las desviaciones cuadráticas con respecto a la media.

$$VARIANZA\ POBLACIONAL = \sigma_n{}^2 = \frac{\sum (X_i - \bar{X})^2}{n};\ = \frac{\sum (X_i - \bar{X})^2 * n_i}{n};\ \sigma_n{}^2 = \frac{\sum X_i{}^2}{n} - \left(\frac{\sum Xi}{n}\right)^2$$

$$VARIANZA\ MUESTRAL: \sigma_{n-1}^2 = S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1};\ \ \sigma_{n-1}^2 = S^2 = \frac{\sum (X_i - \bar{X})^2 * n_i}{n-1}$$

$$DESVIACION\ POBLACIONAL: \sigma_n = S = \sqrt{\sigma_n{}^2};\ DESVIACION\ MUESTRAL\ ;\ \sigma_{n-1} = S = \sqrt{\sigma_{n-1}{}^2}$$

Ejemplo 5:

VARIABLE Xi : NUMERO DE CLIENTES VISITADOS POR LOS VENDEDORES DE UNA EMPRESA.

VENDEDOR N°:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N° DE CLIENTES	6	4	6	6	6	5	4	5	6	5	3	4	6	4	3	4	6	6	7	4

Media aritmética 5.00
Media aritmética 5.00

Xi	ni	Xi*ni	$(Xi-\bar{X})^2$*ni
3	2	6.00	8.00
4	6	24.00	6.00
5	3	15.00	0.00
6	8	48.00	8.00
7	1	7.00	4.00
	20	100	26.00

Varianza Poblacional	1.30	=VARP(B4:U4)
Varianza Poblacional	1.30	=E15/C15
Varianza muestral	1.37	=VAR(B4:U4)
Varianza muestral	1.37	=E15/(C15-1)

Fuente: archivo Medidas de posición.xls; Hoja: Ejemplo 3

1.4 Taller en Excel.

VARIABLE Xi: SALARIOS DE LOS EMPLEADOS DE UNA EMPRESA (SMLM)

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
2	11.6	7.5	5.1	3.4	4.0	15.5	2.2	8.0	13.0	15.8	6.0	8.0	6.4	10.9	7.8	10.8	13.4	10.3	16.0	12.1
3	5.9	1.9	6.8	14.9	10.2	8.2	15.2	5.1	6.2	2.8	6.5	11.4	10.3	4.3	2.3	3.3	11.3	3.4	4.2	7.7
4	15.5	8.9	11.0	5.7	1.0	7.8	4.4	8.5	9.8	4.9	6.4	7.2	15.6	10.7	7.4	9.0	4.9	11.0	3.4	2.2
5	15.1	4.8	8.1	14.8	10.8	5.2	11.2	10.7	2.5	6.7	13.3	5.6	12.8	8.1	11.3	5.2	12.9	5.2	2.3	9.1
6	8.6	4.4	5.3	13.7	12.6	1.1	6.5	9.5	8.9	3.4	9.0	7.4	13.6	6.6	12.7	11.8	14.0	14.1	5.3	14.6
7	2.2	2.3	5.1	2.0	6.5	12.0	13.1	9.2	15.1	7.6	15.1	13.1	11.7	1.3	11.3	7.8	2.2	12.4	14.3	10.8
8	11.1	11.0	12.2	11.3	4.5	9.7	2.0	3.8	11.7	1.4	4.2	13.1	11.9	4.5	13.1	5.5	10.5	11.5	12.7	6.4

Media aritmética	8.51
Media aritmética	8.56

Maxi	16.0
Minir	1.0
Rang	15.0
Clase	5.0
Ampl	3.0

X'i-1	X'i	Xi	ni	Xi*ni	Xi*ni	$(Xi-\bar{X})^2$*ni
1.0	4.0	2.5	23	0.164286	57.5	845.838
4.0	7.0	5.5	33	0.235714	181.5	309.865
7.0	10.0	8.5	25	0.178571	212.5	0.10332
10.0	13.0	12	36	0.257143	414	310.263
13.0	16.0	15	23	0.164286	333.5	810.352
			140	1	1199	2276.42

Varianza Poblacional	16.95	=VARP(B2:U8)
Varianza Poblacional	16.26	=H25/E25
Varianza muestral	16.94	=VAR(B5:U11)
Varianza muestral	16.38	=H25/(E25-1)

Fuente: Archivo: Medidas de posición y dispersion.xls; Hoja: ejemplo 4

Taller. Analice cada una de las variables de la encuesta que está en el archivo TALLER CAPITULO 1

UNIDAD 2. TEORÍA DE PROBABILIDADES.

2.1 Aspectos generales.

La probabilidad es una medida numérica entre cero y uno que mide la opción que en un futuro ocurra un evento específico como resultado de un experimento. Mide el grado de incertidumbre de la ocurrencia de un eventos o suceso.

Los encargados de tomar decisiones no saben con certeza lo que puede ocurrir en un futuro cercano o lejano. Sin embargo la decisión la tiene que tomar. Por ejemplo un fabricante ha desarrollado un nuevo producto basado en la determinación de las necesidades del mercado. Se desea saber si el mercado comprará o no el producto. Una forma de minimizar el riesgo de tomar una decisión incorrecta sería contratar a una empresa de encuestas para que tome una muestra de 100 o 1200 o 2000 elementos de la población y preguntarle a cada persona cómo reaccionaría ante el nuevo producto. Otro ejemplo podría darse cuando el director de un departamento de Planeación de una entidad oficial no sabe cuánto presupuestar para la reparación de vías. Debe tomar datos de las reparaciones anteriores y predecir aproximadamente el valor a presupuestar.

Debido a que existe una incertidumbre considerable al tomar decisiones, resulta importante que todos los riesgos implícitos conocidos se evalúen en forma científica minimizando el riesgo y la incertidumbre de la toma de decisiones.

La probabilidad es la posibilidad que ocurra un evento futuro cuyo valor está entre 0 y 1. La probabilidad que el año termine el 31 de diciembre es 1.0, pero la probabilidad que el año inicie en febrero es 0. En estos casos hay certeza de los que puede ocurrir con respecto a esos dos eventos.

Pero, la probabilidad que el año termine bien es una medida cuyo valor dependerá de varios factores. Una persona que se acaba de ganar el Baloto dirá que es muy cercana a uno, pero una víctima de una pirámide financiera dirá que está cercana a 0.

2.2 Experimento aleatorio

Es definido como un proceso o actividad que la ejecutarse puede uno o varios posibles resultados. Si el azar es el que define el resultado, se dice que el experimento es aleatorio. En cada una de las repeticiones del experimento, habrá uno y sólo uno de los posibles resultados experimentales.
Todos los posibles resultados se conocen como espacio muestral.

Ejemplo:

Experimento Aleatorio	Posibles resultados	Espacio Muestral
Jugar un partido de futbol	Ganar, empatar, perder	S= {ganar, empatar, perder}

Lanzar un Dado	1, 2, 3, 4, 5, 6	S= {1, 2, 3, 4, 5, 6}
Lanzar una moneda	Cara, Sello	S= {cara, sello}

Otros ejemplos de experimentos. Preguntar a un grupo de estudiantes universitarios que probaron tres computadoras personales, cuál prefieren. Medir el niel del agua de un río. Contar el número de empleados de una empresa que tienen más de 60 años de edad. Hacer girar la llave de un mecanismo de encendido de una máquina para determinar si el motor arrancará o no.
Un experimento puede tener uno o más resultados posibles, a los que se denomina eventos.
Si una empresa tiene sólo cinco regiones de ventas y el nombre o número de cada zona se escribe en un trozo de papel y éstos se colocan en una urna, la probabilidad de seleccionar una de las cinco regiones es 1/5.

Si todos los papeles tienen el mismo nombre, “ACEROS LTDA”, la probabilidad que al seleccionar un papel alzar diga “ACEROS LTDA” es 1. De esta forma, la probabilidad 1, representa algo que seguramente va a suceder, y la probabilidad 0 corresponde a algo que no puede suceder.

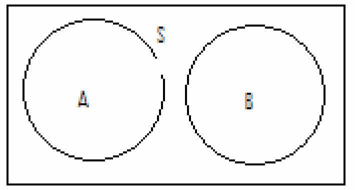
Cuanto más se acerca una probabilidad a 1, es más improbable que suceda el evento al que se asocia, mientras que cuanto más se acerca la probabilidad a 0, más seguros estamos de que no sucederá.

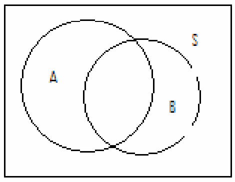
Las probabilidades se analizan desde dos puntos de vista. Probabilidad clásica y probabilidad empírica.

El enfoque clásico o a priori de la probabilidad se basa en la consideración de que los resultados de un experimento son igualmente posibles. Empleando el punto de vista clásico, la probabilidad de que suceda un evento se calcula dividiendo el número de resultados favorables, entre el número total de posibles, como por ejemplo, la probabilidad que la lanzar una moneda salga cara, la probabilidad que al lanzar un dado salga un número mayor de 4, la probabilidad que al lanzar dos dados salga pares, la probabilidad que al lanzar una moneda tres veces, ganen los sellos.
En la probabilidad empírica, la probabilidad de que un evento ocurra a largo plazo se determina observando en que fracción de tiempo sucedieron eventos semejantes en el pasado.

2.4 Regla básicas de probabilidad.

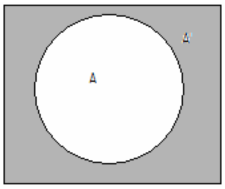
Existe una marcada relación entre la Teoría de Conjuntos y la Teoría de las Probabilidades, derivándose algunas expresiones que representan las operaciones entre conjuntos.





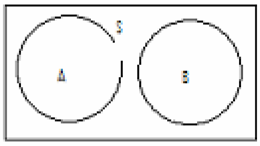
Eventos disjuntos o excluyentes Eventos no disjuntos

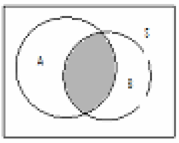
Complemento.



A’ se conoce como el complemento de un evento, donde el complemento donde A’ = {x ∈Ω / x∉ A}

Intersección

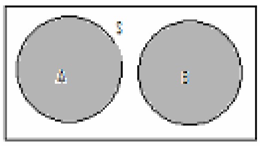


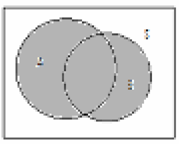


A ∩ B = ∅ **A ∩ B**

Si A ocurre y B ocurre, se representa por A ∩ B. Esta probabilidad se simboliza como P(A ∩ B) = P(A y B).

Unión de eventos.





A U B A U B

Si A ocurre o B ocurre, se representa por A U B y significa que A o B ocurre, o al menos uno de los dos eventos ocurren. Esta probabilidad se simboliza como P(A U B) = P(A o B).

P(AUB)=P(A) + P(B). Eventos excluyentes.
P(AUB)=P(A) + P(B) – P(A∩B). Eventos no excluyentes.
P(AUBUC)=P(A) + P(B) + P(C). Eventos excluyentes.
P(AUBUC)=P(A) + P(B) + P(C) - P(A∩B) - P(A∩C) - P(B∩C) + P(A∩B∩C). Eventos no excluyentes.

2.5 Reglas de conteo.

2.5.1 **Se denomina permutación** a las diferentes ordenaciones que se pueden hacer con un conjunto de eventos. (a_1, a_2, a_3, …, a_n). El número de permutaciones que se pueden hace con n elementos es n!, donde

$$n! = 1 \times 2 \times 3 \times \ldots\ldots \times n$$

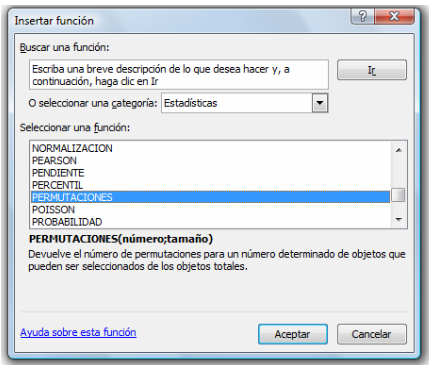
El número de permutaciones de r elementos que se pueden tomar de un conjunto de n elementos es

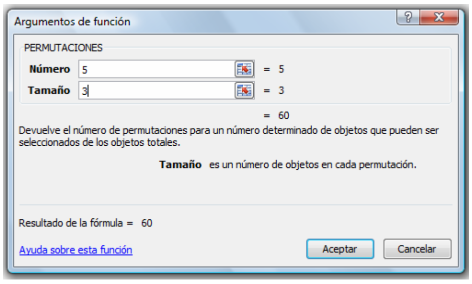
$${}_nP_r = n \,x\, (n-1)\, x\, (n-2)\, x\ldots\ldots x\, (n-r+1) = \frac{n!}{(n-r)!}$$

Ejemplo. Una junta está compuesta por 5 vocales de los cuales se debe elegir el presidente, el secretario y el fiscal. El número de forma como se puede formar esa junta es

$${}_5P_3 = 5 \times 4 \times 3 = 60 = \frac{5!}{(5-3)!} = \frac{120}{2} = 60$$

En Excel =PERMUTACIONES(5;3)





2.5.2 **Se denomina combinación** a los subconjuntos no ordenados de un conjunto, es decir, sin tener en cuenta el orden en la colocación de los objetos.

El número de combinaciones de r elementos tomados de un conjunto de n elemento es:

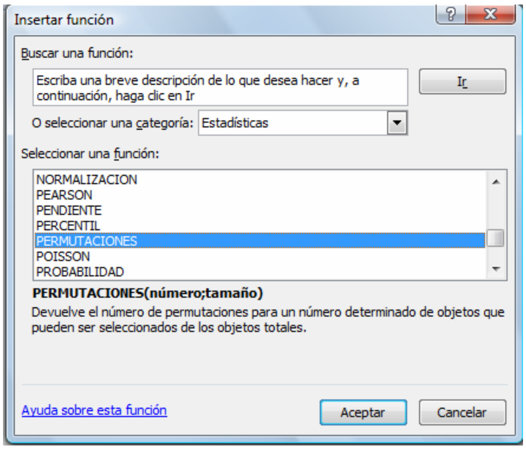
$$\binom{n}{r} = {}_nC_r = \frac{n!}{(n-r)!*r!}$$

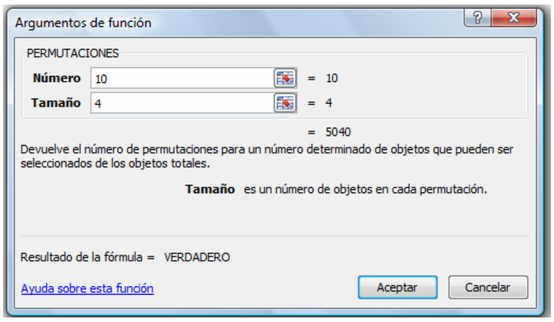
Ejemplo. De junta compuesta por 10 miembros se va a sacar la comisión de empalme compuesta por 4 personas. ¿De cuántas formas se puede sacar esa comisión?

$$\binom{n}{r} = {}_nC_r = \frac{n!}{(n-r)!*r!}$$

$$\binom{10}{4} = {}_{10}C_4 = \frac{10!}{(10-4)!*4!} = 5040$$

En Excel =PERMUTACIONES(10;4)





UNIDAD 3. DISTRIBUCIONES DE PROBABILIDAD EN EXCEL

3.1 **La variable aleatoria** está asociada con los resultados de los experimentos aleatorios.
Ejemplo:
Experimento: Lanzar tres monedas.
Variable aleatoria: Número de caras que resulten

X: Número de caras obtenidas en el lanzamiento.
S = { ccc, ccs, csc, css, scc, scs, ssc, sss } $X = \{ 0, 1, 2, 3\}$

La variable aleatoria toma diferente valores dependiendo del resultado del experimento aleatorio.

Según el tipo de valor que toma la variable discretas, reales y continuas. Las variables discretas son aquellas cuyos valores son números enteros. Las variables reales son las que toman valores decimales. La variables continuas son las que toman valores en un intervalo.

3.2 Distribución de probabilidades
En el experimento de lanzar tres monedas la distribución para el valor de la variable es :

Xi	ni	P(X = Xi)	P(X ≤ Xi)
0	1	0.125	0.125
1	3	0.375	0.500
2	3	0.375	0.875
3	1	0.125	1.000
	8	1.000	

Otro ejemplo de distribución de probabilidades empírica es el caso con la variable numero de accidentes.

VARIABLE Xi : Número de accidentes presentadas en un fin de semana en Bogotá

Numero de accidents por día	Numero de dias en que ocurrieron esos accidentes	Probabilidad simple $P(X = X_i)$	Probabilidad acumulada $P(X \leq X_i)$
4	2	0.050	0.050
5	3	0.075	0.125
6	10	0.250	0.375
7	14	0.350	0.725
8	8	0.200	0.925
9	3	0.075	1.000
	40	**1.000**	

En ella observamos que esas considerados todos los posibles valores de la variable y que las suma de estas probabilidad es 1.

Esta tabla de probabilidades se le conoce como **Distribución de probabilidades**. La distribución de probabilidades es un modelo teórico que describe la forma en que varían los resultados de un experimento aleatorio.
La función de probabilidad

La distribución de probabilidad acumulada es la que acumula las probabilidades:

$P(X\ \ X_i) = P(X_1) + P(X_2) + P(X_3) + \ldots\ldots P(X_m) = P(X \leq X_i) = \sum_{m}^{i=1} [P(X_i)]$

3.2.1 Valor esperado o esperanza matemática

Es el valor promedio probabilístico asociado con el valor de una variable.

$$\mu = E(X) = \sum_{m}^{i=1} P(X_i) * X_i ; con \quad \sum_{m}^{i=1} P(X_i) = 1$$

Ejemplo. Valor esperado del lanzamiento de monedas.

Xi	P(X = Xi)	P(X ≤ Xi)	Xi * P(Xi)
0	0.125	0.125	0.000
1	0.375	0.500	0.375
2	0.375	0.875	0.750
3	0.125	1.000	0.375
	1.000		1.500

μ = 1.500

En un juego largo el promedio de caras es e 1.5

Ejemplo. Valor esperado de accidentes en cada fin de semana.

VARIABLE Xi : Número de accidentes presentadas en un fin de semana en Bogotá

Xi	P(X = Xi)	P(X ≤ Xi)	Xi * P(Xi)
4	0.050	0.050	0.200
5	0.075	0.125	0.375
6	0.250	0.375	1.500
7	0.350	0.725	2.450
8	0.200	0.925	1.600
9	0.075	1.000	0.675
	1.000		6.800

Se espera que en promedio ocurran 6.8 accidentes en cada fin de semana

3.2.2 Varianza esperada

La **varianza esperada** es el promedio esperado de los cuadrados de las dispersiones con respecto al valor esperado.

$$\sigma^2 = \sum_{m}^{i=1} \left((X_i - \mu)^2 * P(X_i) \right)$$

Ejemplo. Varianza para el lanzamiento de las monedas

Xi	P(X = Xi)	P(X ≤ Xi)	Xi * P(Xi)	$(Xi - \mu)^2$ * P(Xi)
0	0.125	0.125	0.000	0.28125
1	0.375	0.500	0.375	0.09375
2	0.375	0.875	0.750	0.09375
3	0.125	1.000	0.375	0.28125
	1.000		1.500	0.75000

$\mu =$ 1.500

$\sigma^2 =$ 0.7500

Ejemplo. Varianza para los accidentes de los fines de semana

VARIABLE Xi : Número de accidentes presentadas en un fin de semana en Bogotá

Xi	$P(X = X_i)$	$P(X \le X_i)$	Xi * P(Xi)	$(Xi - \mu)^2$ * P(Xi)
4	0.050	0.050	0.200	0.392
5	0.075	0.125	0.375	0.243
6	0.250	0.375	1.500	0.16
7	0.350	0.725	2.450	0.014
8	0.200	0.925	1.600	0.288
9	0.075	1.000	0.675	0.363
	1.000		6.800	1.460

$\mu =$ 6.800

$\sigma^2 =$ 1.4600

3.5 Distribución de probabilidad variables discretas

3.3.1 Distribución Binomial. Es una distribución de probabilidad discreta, mide el número de éxitos en una secuencia de n ensayos independientes de Bernoulli , con una probabilidad fija *p* de ocurrencia del éxito entre los ensayos.

La variable binomial es una variable aleatoria discreta, sólo puede tomar los valores 0, 1, 2, 3, 4, ..., n suponiendo que se han realizado n pruebas. Como hay que considerar todas las maneras posibles de obtener x-éxitos y (n-x) fracasos debemos calcular éstas por combinaciones (número combinatorio n sobre x).

La función de probabilidad binomial está dada por

$$P(X = x) = \binom{n}{x} p^{x} + q^{n-x}; \quad con \quad x = 1,2,3,.....$$

Cada experimento tiene dos resultados posibles: ÉXITO y FRACASO
La probabilidad de éxito es *p*
La probabilidad de fracaso es *q = 1 – p*
El resultado obtenido en cada prueba es independiente de los resultados obtenidos anteriormente.
La probabilidad de éxito permanece contante entre un experimento y otro. (Población infinita o ensayos con reemplazamiento).
n es el número de ensayos o el número de veces que se aplica el experimento

Todo experimento que tenga esas características se le puede aplicar el modelo de la **distribución Binomial**. A la variable x que expresa el número de éxitos obtenidos en cada prueba del experimento se llamará **variable aleatoria binomial**.

Ejemplo:
En una universidad el 20% de los alumnos son empleados oficiales. Se selecciona una muestra de 20 alumnos. Cuál es la probabilidad que en la muestra haya 0, 1, 2, 3, etc., empleados oficiales.

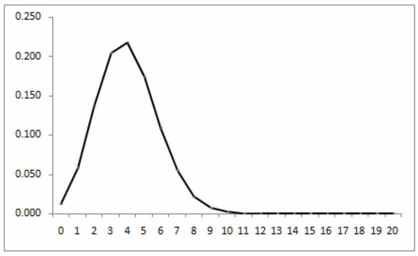
La distribución de probabilidades es:

En una universidad el 20% de los alumnos son empleados oficiales. Se selecciona una muestra de 20 alumnos. Cuál es la probabilidad que en la muestra haya 0, 1, 2, 3, etc., empleados oficiales.

Tamaño de la población.		Infinita	
Éxito	Que el lumnos sea empleado oficial	p =	0.2
Fracaso	Que el lumnos no sea empleado oficial	q =	0.8
n =	20		

x	P(X =x)		P(X ≤ x)	Xi *P(X = xi)	$(Xi-\mu)^2$*P(X = x)
0	0.01152921505	=DISTR.BINOM(A9;B6;E4;0)	0.01152921505	0	0.1844674407
1	0.05764607523	=DISTR.BINOM(A10;B6;E4;0)	0.06917529028	0.0576460752	0.5188146771
2	0.13690942867	=DISTR.BINOM(A11;B6;E4;0)	0.20608471895	0.2738188573	0.5476377147
3	0.20536414301	=DISTR.BINOM(A12;B6;E4;0)	0.41144886196	0.6160924290	0.2053641430
4	0.21819940195	=DISTR.BINOM(A13;B6;E4;0)	0.62964826390	0.8727976078	0.0000000000
5	0.17455952156	=DISTR.BINOM(A14;B6;E4;0)	0.80420778546	0.8727976078	0.1745595216
6	0.10909970097	=DISTR.BINOM(A15;B6;E4;0)	0.91330748643	0.6545982058	0.4363988039
7	0.05454985049	=DISTR.BINOM(A16;B6;E4;0)	0.96785733692	0.3818489534	0.4909486544
8	0.02216087676	=DISTR.BINOM(A17;B6;E4;0)	0.99001821368	0.1772870141	0.3545740282
9	0.00738695892	=DISTR.BINOM(A18;B6;E4;0)	0.99740517260	0.0664826303	0.1846739730
10	0.00203141370	=DISTR.BINOM(A19;B6;E4;0)	0.99943658630	0.0203141370	0.0731308933
11	0.00046168493	=DISTR.BINOM(A20;B6;E4;0)	0.99989827123	0.0050785343	0.0226225617
12	0.00008656592	=DISTR.BINOM(A21;B6;E4;0)	0.99998483716	0.0010387911	0.0055402192
13	0.00001331783	=DISTR.BINOM(A22;B6;E4;0)	0.99999815499	0.0001731318	0.0010787446
14	0.00000166473	=DISTR.BINOM(A23;B6;E4;0)	0.99999981972	0.0000233062	0.0001664729
15	0.00000016647	=DISTR.BINOM(A24;B6;E4;0)	0.99999998620	0.0000024971	0.0000201432
16	0.00000001301	=DISTR.BINOM(A25;B6;E4;0)	0.99999999920	0.0000002081	0.0000018728
17	0.00000000077	=DISTR.BINOM(A26;B6;E4;0)	0.99999999997	0.0000000130	0.0000001293
18	0.00000000003	=DISTR.BINOM(A27;B6;E4;0)	1.00000000000	0.0000000006	0.0000000062
19	0.00000000000	=DISTR.BINOM(A28;B6;E4;0)	1.00000000000	0.0000000000	0.0000000002
20	0.00000000000	=DISTR.BINOM(A29;B6;E4;0)	1.00000000000	0.0000000000	0.0000000000
Suma	1.00000000000	=SUMA(B9:B29)		4.0000000000	3.2000000000

$E(x) = \mu = 4.0000$
$V[x] = \sigma^2 = 3.2000$



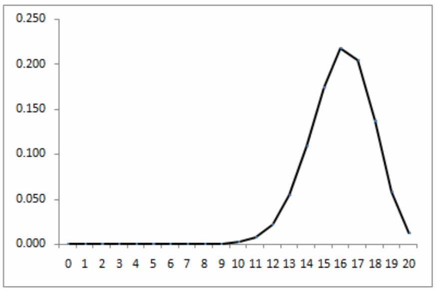
Repita el ejercicio con una probabilidad de *p* = 0.8

En una universidad el 80% de los alumnos son empleados oficiales. Se selecciona una muestra de 20 alumnos. Cuál es la probabilidad que en la muestra haya 0, 1, 2, 3, etc., empleados oficiales.

Tamaño de la población.		Infinita	
Éxito	Que el lumnos sea em	p =	0.8
Fracaso	Que el lumnos no sea	q =	0.2
n =	20		

x	P(X =x)	P(X ≤ x)	Xi *P(X = x)	$(Xi-\mu)^2$*P(X = x)
0	0.00000000000	0.00000000000	0	0.0000000000
1	0.00000000000	0.00000000000	0.0000000000	0.0000000002
2	0.00000000003	0.00000000003	0.0000000001	0.0000000062
3	0.00000000077	0.00000000080	0.0000000023	0.0000001293
4	0.00000001301	0.00000001380	0.0000000520	0.0000018728
5	0.00000016647	0.00000018028	0.0000008324	0.0000201432
6	0.00000166473	0.00000184501	0.0000099884	0.0001664729
7	0.00001331783	0.00001516284	0.0000932248	0.0010787446
8	0.00008656592	0.00010172877	0.0006925274	0.0055402192
9	0.00046168493	0.00056341370	0.0041551644	0.0226225617
10	0.00203141370	0.00259482740	0.0203141370	0.0731308933
11	0.00738695892	0.00998178632	0.0812565481	0.1846739730
12	0.02216087676	0.03214266308	0.2659305211	0.3545740282
13	0.05454985049	0.08669251357	0.7091480563	0.4909486544
14	0.10909970097	0.19579221454	1.5273958136	0.4363988039
15	0.17455952156	0.37035173610	2.6183928234	0.1745595216
16	0.21819940195	0.58855113804	3.4911904311	0.0000000000
17	0.20536414301	0.79391528105	3.4911904311	0.2053641430
18	0.13690942867	0.93082470972	2.4643697161	0.5476377147
19	0.05764607523	0.98847078495	1.0952754294	0.5188146771
20	0.01152921505	1.00000000000	0.2305843009	0.1844674407
Suma	1.00000000000		16.0000000000	3.2000000000

$E(x) = \mu =$ 16.000
$V[x] = \sigma^2 =$ 3.200



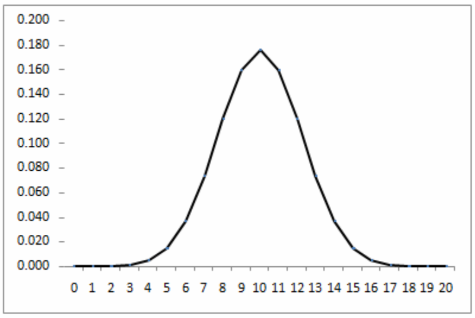
Repita el ejercicio con una probabilidad de p = 0.5

En una universidad el 50% de los alumnos son empleados oficiales. Se selecciona una muestra de 20 alumnos. Cuál es la probabilidad que en la muestra haya 0, 1, 2, 3, etc., empleados

Tamaño de la población.		Infinita	
Éxito	Que el lumnos s	p =	0.5
Fracaso	Que el lumnos n	q =	0.5
n =	20		

x	P(X =x)	P(X ≤ x)	Xi *P(X = x)	$(Xi-\mu)^2$*P(X = x)
0	0.00000095367	0.00000095367	0	0.0000953674
1	0.00001907349	0.00002002716	0.0000190735	0.0015449524
2	0.00018119812	0.00020122528	0.0003623962	0.0115966797
3	0.00108718872	0.00128841400	0.0032615662	0.0532722473
4	0.00462055206	0.00590896606	0.0184822083	0.1663398743
5	0.01478576660	0.02069473267	0.0739288330	0.3696441650
6	0.03696441650	0.05765914917	0.2217864990	0.5914306641
7	0.07392883301	0.13158798218	0.5175018311	0.6653594971
8	0.12013435364	0.25172233582	0.9610748291	0.4805374146
9	0.16017913818	0.41190147400	1.4416122437	0.1601791382
10	0.17619705200	0.58809852600	1.7619705200	0.0000000000
11	0.16017913818	0.74827766418	1.7619705200	0.1601791382
12	0.12013435364	0.86841201782	1.4416122437	0.4805374146
13	0.07392883301	0.94234085083	0.9610748291	0.6653594971
14	0.03696441650	0.97930526733	0.5175018311	0.5914306641
15	0.01478576660	0.99409103394	0.2217864990	0.3696441650
16	0.00462055206	0.99871158600	0.0739288330	0.1663398743
17	0.00108718872	0.99979877472	0.0184822083	0.0532722473
18	0.00018119812	0.99997997284	0.0032615662	0.0115966797
19	0.00001907349	0.99999904633	0.0003623962	0.0015449524
20	0.00000095367	1.00000000000	0.0000190735	0.0000953674
Suma	1.00000000000		10.0000000000	5.0000000000

$E(x) = \mu =$ 10.000
$V[x] = \sigma^2 =$ 5.000



3.3.2 Distribución Hipergeométrica.

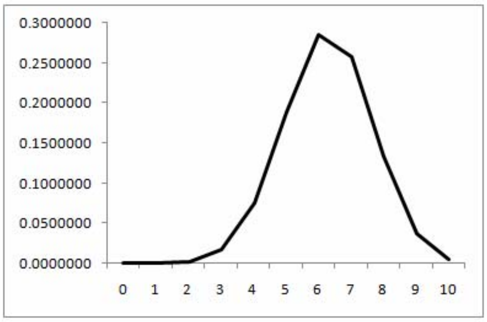
El curso de estadística está conformado por 25 mujeres y 15 hombres. Para formar parte del comité de acreditación se tomé una muestra de 10 alumnos. Construya la distribución de probabilidad hipergeométrica

El curso de estadística eastá compuesto por 25 mujeres y 15 hombres. Para formar parte del comité de acreditación se tomé una muestra de 10 alumnos.

Tamaño de la población.			40
Éxito	Ser mujer	N1	25
Fracaso	Ser hombre	N2	15
n =	10		

=DISTR.HIPERGEOM(x,n,N1,N)

x	P(X =x)	P(X ≤ x)	Xi *P(X = x)	$(Xi-\mu)^2$*P(X = x)
0	0.0000035	0.0000035	0.0000000	0.00013839
1	0.0001476	0.0001512	0.0001476	0.00406856
2	0.0022774	0.0024286	0.0045549	0.04113635
3	0.0174604	0.0198890	0.0523812	0.18442558
4	0.0746918	0.0945808	0.2987670	0.37812699
5	0.1882232	0.2828040	0.9411161	0.29409877
6	0.2851867	0.5679907	1.7111201	0.01782417
7	0.2580261	0.8260167	1.8061824	0.14513965
8	0.1339751	0.9599918	1.0718005	0.41029864
9	0.0361520	0.9961438	0.3253680	0.27339951
10	0.0038562	1.0000000	0.0385621	0.05422800
	1.0000000		6.2500000	1.80288462



$E(x) = \mu = 6.2500$

$V[x] = \sigma^2 = 1.8029$

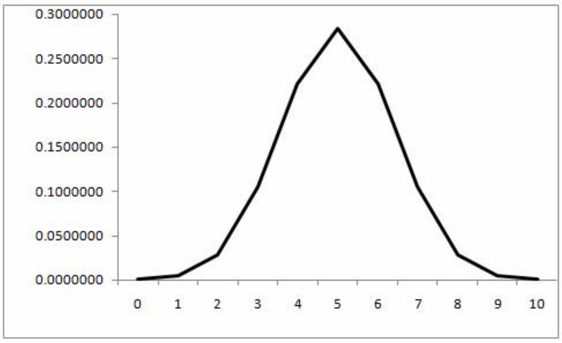
Repetir el ejercicio suponiendo que el número de empleados oficiales es 20 y no oficiales, es 20

El curso de estadística eastá compuesto por 20 mujeres y 20 hombres. Para formar parte del comité de acreditación se tomé una muestra de 10 alumnos.

Tamaño de la población.			40
Éxito	Ser mujer	N1	20
Fracaso	Ser hombre	N2	20
n =	10		

=DISTR.HIPERGEOM(x,n,N1,N)

x	P(X =x)	P(X ≤ x)	Xi *P(X = x)	$(Xi-\mu)^2$*P(X = x)
0	0.0002180	0.0002180	0.0000000	0.00544900
1	0.0039629	0.0041809	0.0039629	0.06340652
2	0.0282357	0.0324166	0.0564714	0.25412142
3	0.1042549	0.1366715	0.3127648	0.41701977
4	0.2215418	0.3582133	0.8861670	0.22154175
5	0.2835734	0.6417867	1.4178672	0.00000000
6	0.2215418	0.8633285	1.3292505	0.22154175
7	0.1042549	0.9675834	0.7297846	0.41701977
8	0.0282357	0.9958191	0.2258857	0.25412142
9	0.0039629	0.9997820	0.0356662	0.06340652
10	0.0002180	1.0000000	0.0021796	0.00544900
	1.0000000		5.0000000	1.92307692



$E(x) = \mu = 5.0000$

$V[x] = \sigma^2 = 1.9231$

Repetir el ejercicio suponiendo que el número de empleados oficiales es 15 y no oficiales es 25

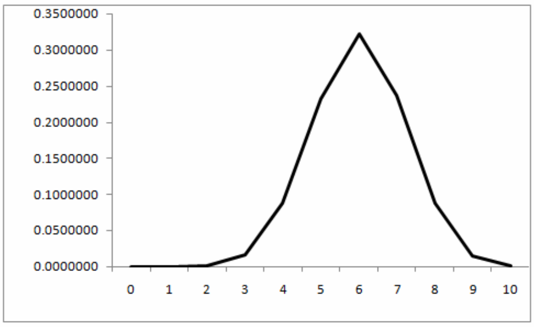
El curso de estadística eastá compuesto por 15 mujeres y 25 hombres. Para formar parte del comité de acreditación se tomé una muestra de 10 alumnos.

Tamaño de la población.			40
Éxito	Ser mujer	N1	15
Fracaso	Ser hombre	N2	25
n =	10		

=DISTR.HIPERGEOM(x,n,N1,N)

x	P(X =x)	P(X ≤ x)	Xi *P(X = x)	(Xi-µ)²*P(X = x)
0	0.0000003	0.0000003	0.0000000	0.00001101
1	0.0000459	0.0000462	0.0000459	0.00114722
2	0.0014455	0.0014917	0.0028910	0.02312804
3	0.0167036	0.0181953	0.0501107	0.15033224
4	0.0876938	0.1058891	0.3507752	0.35077522
5	0.2315116	0.3374007	1.1575582	0.23151164
6	0.3215439	0.6589447	1.9292637	0.00000000
7	0.2362364	0.8951810	1.6536546	0.23623637
8	0.0885886	0.9837697	0.7087091	0.35435456
9	0.0153116	0.9990813	0.1378045	0.13780455
10	0.0009187	1.0000000	0.0091870	0.01469915
	1.0000000		6.0000000	1.50000000

$E(x) = \mu = 6.0000$
$V[x] = \sigma^2 = 1.5000$



3.3.3 Distribución de Poisson

La llamada distribución de probabilidad de Poisson describe la cantidad de veces que ocurre un evento en un intervalo de tiempo determinado o de espacio o de volumen. El valor de la probabilidad está dada por la siguiente función.

$$P(X = x) = \frac{e^{-\lambda} * \lambda^{x}}{x!}; \quad para \;\; x = 0,1,2,3,........$$

La variable a la cual se le aplicará es distribución de probabilidad debe cumplir los siguientes aspectos.

- x es el número de éxitos en el intervalo.
- λ es el promedio de éxitos en el intervalo.
- La ocurrencia del evento en un intervalo no puede afectar la ocurrencia en otro intervalo, es decir la ocurrencia el evento en cada intervalo es independiente de los otros intervalos.
- Se utiliza para medir la gestión de servicio al cliente o para analizar la calidad del servicio o de cumplimiento de los agentes productivos.

Ejemplo. El director de un Empresa de Servicio Público domiciliario que está interesado en mejorar la atención a los clientes, colocó un encuestador en la puerta de ingreso y le pidió que en intervalos de 15 minutos contara los clientes que llegaban. El encuestador entregó los siguientes datos, (clientes que llegaron en intervalos de 15 minutos.

12	15	10	12	15	18	15	12
11	15	16	14	13	16	17	13
14	18	12	14	16	15	14	12

Según esos datos la media de ellos es 14.125 clientes. Es decir en promedio cada 15 minutos llegan 14.125 clientes, es decir λ = 14.125 clientes cada 15 minutos.

Con estos datos se pueden resolver preguntas como: ¿Cuál es la probabilidad que en un intervalo de 15 minutos lleguen exactamente 10 clientes?

$$P(X = 10) = \frac{e^{-14.125} * 14.125^{10}}{10!} = 0.06393107$$

En Excel =POISSON(10;14.125;0)

La distribución de probabilidad, en Excel, para esta variable sería.

	A	B	C	D	E
1	Variable Xi: Númerode clientes que llegan al banco en intervalos de 15 minutos				
2		12	15	10	12
3		11	15	16	14
4		14	18	12	14
5					
6	λ =	=PROMEDIO(B2:I4)		Clientes cada 15 minutos	
7	x =	10			
8	P(X = 10) =	=POISSON(B7;B6;0)			
9					
10		x	P(X = x)	P(X ≤ x)	P(X > x)
11		0	=POISSON(B11;B6;0)	=POISSON(B11;B6;1)	=1-D11
12		1	=POISSON(B12;B6;0)	=POISSON(B12;B6;1)	=1-D12
13		2	=POISSON(B13;B6;0)	=POISSON(B13;B6;1)	=1-D13
14		3	=POISSON(B14;B6;0)	=POISSON(B14;B6;1)	=1-D14
15		4	=POISSON(B15;B6;0)	=POISSON(B15;B6;1)	=1-D15
16		5	=POISSON(B16;B6;0)	=POISSON(B16;B6;1)	=1-D16
17		6	=POISSON(B17;B6;0)	=POISSON(B17;B6;1)	=1-D17
18		7	=POISSON(B18;B6;0)	=POISSON(B18;B6;1)	=1-D18
19		8	=POISSON(B19;B6;0)	=POISSON(B19;B6;1)	=1-D19
20		9	=POISSON(B20;B6;0)	=POISSON(B20;B6;1)	=1-D20
21		10	=POISSON(B21;B6;0)	=POISSON(B21;B6;1)	=1-D21
22		11	=POISSON(B22;B6;0)	=POISSON(B22;B6;1)	=1-D22
23		12	=POISSON(B23;B6;0)	=POISSON(B23;B6;1)	=1-D23
24		13	=POISSON(B24;B6;0)	=POISSON(B24;B6;1)	=1-D24
25		14	=POISSON(B25;B6;0)	=POISSON(B25;B6;1)	=1-D25
26		15	=POISSON(B26;B6;0)	=POISSON(B26;B6;1)	=1-D26

Variable Xi: Númerode clientes que llegan al banco en intervalos de 15 minutos

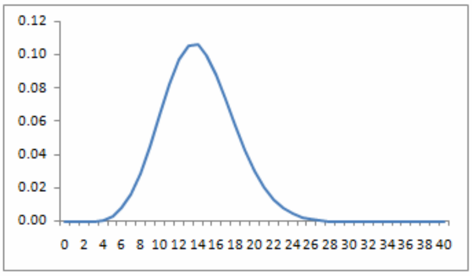
12	15	10	12	15	18	15	12
11	15	16	14	13	16	17	13
14	18	12	14	16	15	14	12

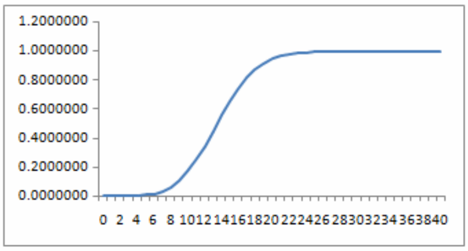
λ = 14.125 Clientes cada 15 minutos

x = 10

P(X = 10) = 0.0639

x	P(X = x)	P(X ≤ x)	P(X > x)
0	0.0000007	0.0000007	0.9999993
1	0.0000104	0.0000111	0.9999889
2	0.0000732	0.0000843	0.9999157
3	0.0003447	0.0004290	0.9995710
4	0.0012171	0.0016461	0.9983539
5	0.0034384	0.0050845	0.9949155
6	0.0080945	0.0131789	0.9868211
7	0.0163335	0.0295124	0.9704876
8	0.0288388	0.0583512	0.9416488
9	0.0452609	0.1036122	0.8963878
10	0.0639311	0.1675433	0.8324567
11	0.0820933	0.2496366	0.7503634
12	0.0966307	0.3462672	0.6537328
13	0.1049929	0.4512602	0.5487398
14	0.1059304	0.5571905	0.4428095
15	0.0997511	0.6569416	0.3430584
16	0.0880615	0.7450031	0.2549969
17	0.0731688	0.8181719	0.1818281
18	0.0574172	0.8755890	0.1244110
19	0.0426851	0.9182742	0.0817258
20	0.0301464	0.9484205	0.0515795
21	0.0202770	0.9686975	0.0313025
22	0.0130188	0.9817163	0.0182837
23	0.0079952	0.9897115	0.0102885
24	0.0047055	0.9944171	0.0055829
25	0.0026586	0.9970757	0.0029243
26	0.0014443	0.9985200	0.0014800





3.4 Taller en Excel.

UNIDAD 4. DISTRIBUCIÓN DE PROBABILIDAD NORMAL

4.1 Variable aleatoria continua.

Se ha dicho que una variable es discreta cuando los resultados del experimento son valores enteros. Cuando los resultados toman valores con decimales, se dirá que la variable es real. Esos valores podría estar en un intervalo, como en el caso que los salarios de una empresa oscilan entre 1 SMLM y 8 SMLM , entendiéndose que una persona puede tener como salario 5.75 SMLM .E

En el caso de la variable número de personas que van a pagar a una oficina de recaudos, es un valor enero, sin embargo, es común afirma que el valor de la variable toma valores en un rango. Por ejemplo en el día van entre 100 y 200 personas a pagar a EPS. En estos casos la variable la asociaremos como continua.

En el caso de variable continua la distribución de probabilidad es la integral de la función de densidad, por lo que tenemos entonces que:

El valor esperado y la varianza esperada se calculará de la misma forma como se calculó para la variable discreta.

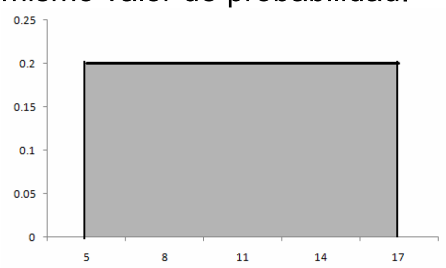
Variable: Número de clientes en un almacén por día

X'_{i-1}	X'_i	X_i	$P(X_i)$	$X_i * P(X_i)$	$(Xi-\mu)^2*P(Xi)$
5	8	6.5	0.164	1.066	6.059
8	11	9.5	0.236	2.242	2.237
11	14	12.5	0.179	2.238	0.001
14	17	15.5	0.258	3.999	2.202
17	20	18.5	0.164	3.034	5.751
			1.000	12.579	16.250

Se esperan que lleguen en promedio 12.579 clientes por día
La varianza es 16.250
La desviación estándar es 4.03110486

4.2 Distribución de probabilidad Uniforme.

Una variable es uniforme cuando cualquier resultado de la variable tiene el mismo valor de probabilidad.

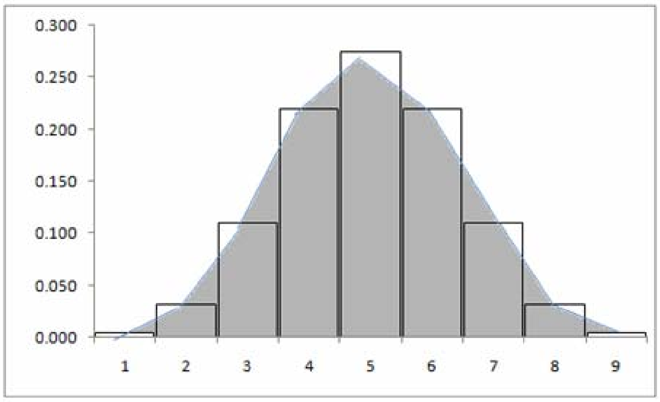


4.3 Distribución normal.

Para entender el comportamiento de la distribución normal, tomemos la siguiente distribución de Probabilidad.

Variable: Tiempo de viaje entre dos ciudades en horas

X'_{i-1}	X'_i	X_i	$P(X_i)$	$X_i * P(X_i)$	$(Xi-\mu)^2*P(Xi)$
2.00	2.25	2.125	0.004	0.009	0.039
2.25	2.50	2.375	0.031	0.074	0.297
2.50	2.75	2.625	0.109	0.286	0.991
2.75	3.00	2.875	0.219	0.630	1.849
3.00	3.25	3.125	0.274	0.856	2.227
3.25	3.50	3.375	0.219	0.739	1.849
3.50	3.75	3.625	0.109	0.395	0.991
3.75	4.00	3.875	0.031	0.120	0.297
4.00	4.25	4.125	0.004	0.017	0.039
				3.125	8.580



La distribución de probabilidad es simétrica y la grafica tiende a tener una forma de campana simétrica.

Distribuciones con estas características se le conocerá como distribuciones de probabilidad normal o también llamada **distribución de Gauss** o **distribución gaussiana**.

De las distribuciones de probabilidad para variable contínua es la mas importante pues describe en forma aproximada muchos fenómenos que La función de densidad está dada por:

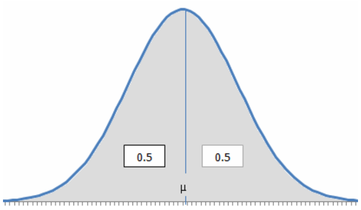
$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Si f(x) =Esta función cumple con

$$si\ f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}};\ entonces\ \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1;\ \int_{-\infty}^{0} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 0.5;\ \int_{0}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 0.5$$

Donde μ es la media y σ es la desviación estándar y σ^2 es la varianza.

Como la gran dificultad está el cálculo de las integrales en función de la media μ, y la desviación estándar σ, se puede medir las dispersiones en desviaciones estándar de la siguiente forma:



$Xi - \overline{X}$ es la dispersión de la variable Xi con respecto a la media aritmética.

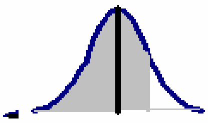
$\dfrac{Xi - \overline{X}}{\sigma} = Z$ Es el número de dispersiones de la variable medida en desviaciones estándar. Esta medida se conoce como Z y el procedimiento se le conoce como estandarización.

El valor de Z puede ir desde -∝ a +∝

$$si\ \ f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2}; entonces \quad \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dx = 1; \quad \int_{-\infty}^{0} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dx = 0.5; \quad \int_{0}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dx = 0.5$$

Esto permite que se hayan construido tablas que permitan rápidamente aproximar el valor de la probabilidad en función dl valor de Z.

TABLA DE DISTRIBUCION NORMAL

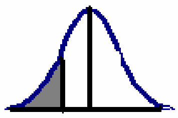


$$Z = \frac{X_i - \overline{X}}{\sigma}$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	0.5000000	0.5039894	0.5079783	0.5119665	0.5159534	0.5199388	0.5239222	0.5279032	0.5318814	0.5358564	0.0
0.1	0.5398278	0.5437953	0.5477584	0.5517168	0.5556700	0.5596177	0.5635595	0.5674949	0.5714237	0.5753454	0.1
0.2	0.5792597	0.5831662	0.5870644	0.5909541	0.5948349	0.5987063	0.6025681	0.6064199	0.6102612	0.6140919	0.2
0.3	0.6179114	0.6217195	0.6255158	0.6293000	0.6330717	0.6368307	0.6405764	0.6443088	0.6480273	0.6517317	0.3
0.4	0.6554217	0.6590970	0.6627573	0.6664022	0.6700314	0.6736448	0.6772419	0.6808225	0.6843863	0.6879331	0.4
0.5	0.6914625	0.6949743	0.6984682	0.7019440	0.7054015	0.7088403	0.7122603	0.7156612	0.7190427	0.7224047	0.5
0.6	0.7257469	0.7290691	0.7323711	0.7356527	0.7389137	0.7421539	0.7453731	0.7485711	0.7517478	0.7549029	0.6
0.7	0.7580363	0.7611479	0.7642375	0.7673049	0.7703500	0.7733726	0.7763727	0.7793501	0.7823046	0.7852361	0.7
0.8	0.7881446	0.7910299	0.7938919	0.7967306	0.7995458	0.8023375	0.8051055	0.8078498	0.8105703	0.8132671	0.8
0.9	0.8159399	0.8185887	0.8212136	0.8238145	0.8263912	0.8289439	0.8314724	0.8339768	0.8364569	0.8389129	0.9
1.0	0.8413447	0.8437524	0.8461358	0.8484950	0.8508300	0.8531409	0.8554277	0.8576903	0.8599289	0.8621434	1.0

Tabla N° 1

TABLA DE DISTRIBUCION NORMAL

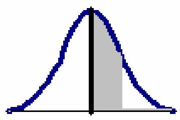


$$Z = \frac{X_i - \overline{X}}{\sigma}$$

Z	0.00	-0.01	-0.02	-0.03	-0.04	-0.05	-0.06	-0.07	-0.08	-0.09	Z
0.0	0.5000000	0.4960106	0.4920217	0.4880335	0.4840466	0.4800612	0.4760778	0.4720968	0.4681186	0.4641436	0.00
-0.1	0.4601722	0.4562047	0.4522416	0.4482832	0.4443300	0.4403823	0.4364405	0.4325051	0.4285763	0.4246546	-0.10
-0.2	0.4207403	0.4168338	0.4129356	0.4090459	0.4051651	0.4012937	0.3974319	0.3935801	0.3897388	0.3859081	-0.20
-0.3	0.3820886	0.3782805	0.3744842	0.3707000	0.3669283	0.3631693	0.3594236	0.3556912	0.3519727	0.3482683	-0.30
-0.4	0.3445783	0.3409030	0.3372427	0.3335978	0.3299686	0.3263552	0.3227581	0.3191775	0.3156137	0.3120669	-0.40
-0.5	0.3085375	0.3050257	0.3015318	0.2980560	0.2945985	0.2911597	0.2877397	0.2843388	0.2809573	0.2775953	-0.50
-0.6	0.2742531	0.2709309	0.2676289	0.2643473	0.2610863	0.2578461	0.2546269	0.2514289	0.2482522	0.2450971	-0.60
-0.7	0.2419637	0.2388521	0.2357625	0.2326951	0.2296500	0.2266274	0.2236273	0.2206499	0.2176954	0.2147639	-0.70
-0.8	0.2118554	0.2089701	0.2061081	0.2032694	0.2004542	0.1976625	0.1948945	0.1921502	0.1894297	0.1867329	-0.80
-0.9	0.1840601	0.1814113	0.1787864	0.1761855	0.1736088	0.1710561	0.1685276	0.1660232	0.1635431	0.1610871	-0.90
-1.0	0.1586553	0.1562476	0.1538642	0.1515050	0.1491700	0.1468591	0.1445723	0.1423097	0.1400711	0.1378566	-1.00

Tabla N° 2

TABLA DE DISTRIBUCION NORMAL



$$Z = \frac{X_i - \overline{X}}{\sigma}$$

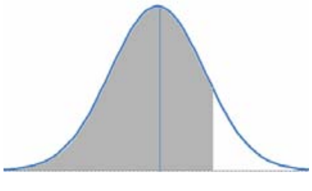
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	0.0000000	0.0039894	0.0079783	0.0119665	0.0159534	0.0199388	0.0239222	0.0279032	0.0318814	0.0358564	0.0
0.1	0.0398278	0.0437953	0.0477584	0.0517168	0.0556700	0.0596177	0.0635595	0.0674949	0.0714237	0.0753454	0.1
0.2	0.0792597	0.0831662	0.0870644	0.0909541	0.0948349	0.0987063	0.1025681	0.1064199	0.1102612	0.1140919	0.2
0.3	0.1179114	0.1217195	0.1255158	0.1293000	0.1330717	0.1368307	0.1405764	0.1443088	0.1480273	0.1517317	0.3
0.4	0.1554217	0.1590970	0.1627573	0.1664022	0.1700314	0.1736448	0.1772419	0.1808225	0.1843863	0.1879331	0.4
0.5	0.1914625	0.1949743	0.1984682	0.2019440	0.2054015	0.2088403	0.2122603	0.2156612	0.2190427	0.2224047	0.5
0.6	0.2257469	0.2290691	0.2323711	0.2356527	0.2389137	0.2421539	0.2453731	0.2485711	0.2517478	0.2549029	0.6
0.7	0.2580363	0.2611479	0.2642375	0.2673049	0.2703500	0.2733726	0.2763727	0.2793501	0.2823046	0.2852361	0.7
0.8	0.2881446	0.2910299	0.2938919	0.2967306	0.2995458	0.3023375	0.3051055	0.3078498	0.3105703	0.3132671	0.8
0.9	0.3159399	0.3185887	0.3212136	0.3238145	0.3263912	0.3289439	0.3314724	0.3339768	0.3364569	0.3389129	0.9
1.0	0.3413447	0.3437524	0.3461358	0.3484950	0.3508300	0.3531409	0.3554277	0.3576903	0.3599289	0.3621434	1.0

Tabla N° 3

Aplicaciones de la distribución normal

Ejemplo N° 1. Las ventas diarias de una empresa tiene media igual a 3.5 millones con una desviación estándar de 1.1 millones.

¿Cuál es la probabilidad que en un día se vendan 4 o menos millones de pesos?



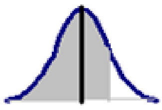
$$\overline{X} = 3.5; \quad \sigma = 1.1 \quad X_i = 4; \quad Z = \frac{4 - 3.5}{1.1} = 0.45$$

El valor de Z se busca en la tabla N° 1 Ubicando la fila 0.4 y la columna 0.05.

La probabilidad que la venta sea menor o igual a 4 millones es 0.67364, es decir

P(X ≤ 4) = P(Z ≤ 0.45)= 0.67364

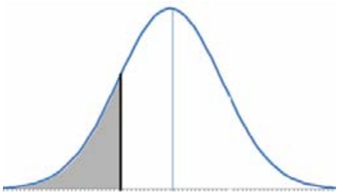
TABLA DE DISTRIBUCION NORMAL



$$Z = \frac{X_i - \bar{X}}{\sigma}$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586	0.0
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535	0.1
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409	0.2
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173	0.3
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793	0.4
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240	0.5

¿Cuál es la probabilidad que en un día se vendan 2.8 o menos millones de pesos?

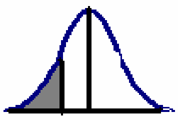


$$\bar{X} = 3.5; \quad \sigma = 1.1 \quad X_i = 2.8; \quad Z = \frac{2.8 - 3.5}{1.1} = -0.67$$

En la tabla N° 2 se encuentra el valor de Z

El valor de Z se busca en la tabla N° 2 Ubicando la fila -0.6 y la columna -0.07. La probabilidad que la venta sea menor o igual a 2.8 millones es 0.2514289, es

TABLA DE DISTRIBUCION NORMAL

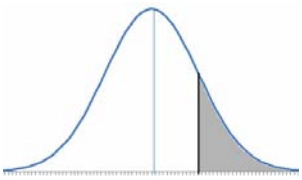


$$Z = \frac{X_i - \bar{X}}{\sigma}$$

Z	0.00	-0.01	-0.02	-0.03	-0.04	-0.05	-0.06	-0.07	-0.08	-0.09	Z
0.0	0.5000000	0.4960106	0.4920217	0.4880335	0.4840466	0.4800612	0.4760778	0.4720968	0.4681186	0.4641436	0.00
-0.1	0.4601722	0.4562047	0.4522416	0.4482832	0.4443300	0.4403823	0.4364405	0.4325051	0.4285763	0.4246546	-0.10
-0.2	0.4207403	0.4168338	0.4129356	0.4090459	0.4051651	0.4012937	0.3974319	0.3935801	0.3897388	0.3859081	-0.20
-0.3	0.3820886	0.3782805	0.3744842	0.3707000	0.3669283	0.3631693	0.3594236	0.3556912	0.3519727	0.3482683	-0.30
-0.4	0.3445783	0.3409030	0.3372427	0.3335978	0.3299686	0.3263552	0.3227581	0.3191775	0.3156137	0.3120669	-0.40
-0.5	0.3085375	0.3050257	0.3015318	0.2980560	0.2945985	0.2911597	0.2877397	0.2843388	0.2809573	0.2775953	-0.50
-0.6	0.2742531	0.2709309	0.2676289	0.2643473	0.2610863	0.2578461	0.2546269	0.2514289	0.2482522	0.2450971	-0.60
-0.7	0.2419637	0.2388521	0.2357625	0.2326951	0.2296500	0.2266274	0.2236273	0.2206499	0.2176954	0.2147639	-0.70

decir, P(X ≤ 2.8) = P(Z ≤ -0.67)= 0.2514289

¿Cuál es la probabilidad que en un día se vendan 3.8 o más millones de pesos?



Hay que entender que P(X ≥ x) = 1 - P(X ≤ x)

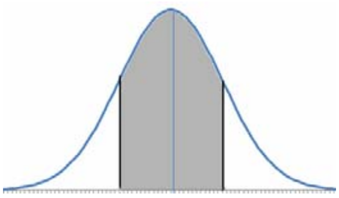
$$\bar{X} = 3.5; \quad \sigma = 1.1 \quad X_i = 3.8; \quad Z = \frac{3.8 - 3.5}{1.1} = 0.27$$

El valor de Z se busca en la tabla N° 1.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	0.5000000	0.5039894	0.5079783	0.5119665	0.5159534	0.5199388	0.5239222	0.5279032	0.5318814	0.5358564	0.0
0.1	0.5398278	0.5437953	0.5477584	0.5517168	0.5556700	0.5596177	0.5635595	0.5674949	0.5714237	0.5753454	0.1
0.2	0.5792597	0.5831662	0.5870644	0.5909541	0.5948349	0.5987063	0.6025681	0.6064199	0.6102612	0.6140919	0.2
0.3	0.6179114	0.6217195	0.6255158	0.6293000	0.6330717	0.6368307	0.6405764	0.6443088	0.6480273	0.6517317	0.3

P(Xi≥3.8)=1-P(X≤3.8) = P(Z ≥ 0.27) =1–P(Z < 0.27) = 0.6064199 = 0.3935801

¿Cuál es la probabilidad que en un día se vendan entre 3 y 4 millones de pesos?



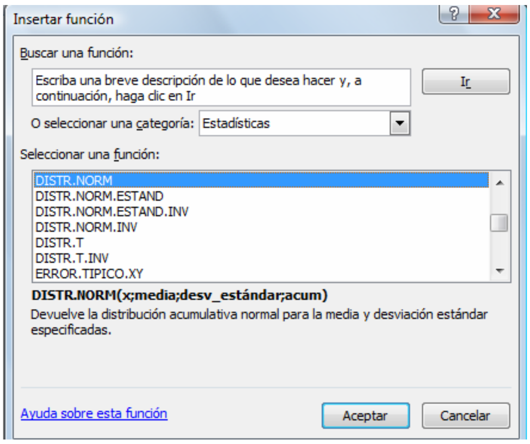
P(a ≤ X ≤ b) = P(X ≤ b) - P(X ≤ a)

$$\overline{X} = 3.5; \quad \sigma = 1.1 \quad X_i = 30; \quad Z = \frac{3.0 - 3.5}{1.1} = -0.45$$

$$\overline{X} = 3.5; \quad \sigma = 1.1 \quad X_i = 4; \quad Z = \frac{4.0 - 3.5}{1.1} = 0.45$$

P(3 ≤ X ≤ 4) = P(X ≤ 4) - P(X ≤ 3) = P(X ≤ 0.45) - P(X ≤ 0.45) = = 0.6736448 - 0.3263552 = 0.3472896
P(X ≤ 0.45) - P(X ≤ 0.45) = = 0.6736448 - 0.3263552 = 0.3472896

Estas mismas funciones e pueden aplicar en Excel utilizando las siguientes funciones estadísticas.



=DISTR.NORM(4;3.5;1.1;1) = 0.675281858

=DISTR.NORM(3.5;3.5;1.1;1) = 0.262269718

Realizar el siguiente taller en Excel utilizando las funciones para distribución normal

	A	B	C	D	E	F
1	Dada una distribución normal, encuentre el área bajo la curva que cae:					
2	a) a la izquierda de z = 1.52		$P(Z \le Z_1)$	$P(Z \le 1.52) =$		=DISTR.NORM.ESTAND(1.52)
3	b) a la derecha de z = -0.9		$P(Z > Z_1)$	$P(Z > -0.9) =$		=1-DISTR.NORM.ESTAND(-0.9)
4	c) a la derecha de z = 1.28		$P(Z > Z_1)$	$P(Z > 1.28) =$		=1-DISTR.NORM.ESTAND(1.28)
5	e) entre 1.8 y 2.7		$P(Z_1 \le Z \le Z_2)$	$P(1.8 \le Z \le 2.7) =$	$P(Z \le 2.7) - P(Z \le 1.8)$	=DISTR.NORM.ESTAND(2.7)-DISTR.NORM.ESTAND(1.8)
6	f) a la izquierda de z = -1.93		$P(Z \le Z_1)$	$P(Z \le -1.93) =$		=DISTR.NORM.ESTAND(-1.93)
7	c) entre -0.5 y 1.84		$P(Z_1 \le Z \le Z_2)$	$P(-0.5 \le Z \le 1.84) =$	$P(Z \le 1.84) - P(Z \le -0.5)$	=DISTR.NORM.ESTAND(1.84)-DISTR.NORM.ESTAND(-0.5)
8						
9						
10	Encuentre el valor de z si el área bajo una curva estándar:					
11	a) a la izquierda es 0.3510		$P(Z \le Z_1) = 0.3510$	=DISTR.NORM.ESTAND.INV(0.351)		
12	b) a la izquierda es 0.1234		$P(Z \le Z_1) = 0.1234$	=DISTR.NORM.ESTAND.INV(0.1234)		
13	c) a la izquierda es 0.7245		$P(Z \le Z_1) = 0.7245$	=DISTR.NORM.ESTAND.INV(0.7245)		
14	c) a la izquierda es 0.979587		$P(Z \le Z_1) = 0.979587$	=DISTR.NORM.ESTAND.INV(0.979587)		
15	c) a la derecha es 0.2523		$P(Z > Z_1) = 0.2523$	=DISTR.NORM.ESTAND.INV(1-0.2523)		
16	c) a la derecha es 0.6508		$P(Z > Z_1) = 0.6508$	=DISTR.NORM.ESTAND.INV(1-0.6508)		
17	c) a la derecha es 0.8586		$P(Z > Z_1) = 0.8586$	=DISTR.NORM.ESTAND.INV(1-0.8586)		
18	c) a la derecha es 0.979591					
19						
20	4. Sea X una variable que tiene distribución normal y los siguientes parámetros. Media 100 y desviación 25.					
21						
22	Halle las siguientes probabilidadesprobabilidades					
23	a) $P[X \le 92.5]$					
24	b) $P[X \ge 76]$					
25	c) $P[X \le 107.5]$					
26	d) $P[X \le 124]$					
27	e) $P[77.5 \le X \le 100]$					
28	f) $P[91 \le X \le 127]$					
29						
30	Media	100				
31	Desviación	25				
32						
33	a) $P(X \le 92.5) =$			=DISTR.NORM(92.5;B30;B31;1)		
34	b) $P(X \ge 76) =$			=1-DISTR.NORM(76;B30;B31;1)		
35	c) $P(X \le 107.5) =$			=DISTR.NORM(107.5;B30;B31;1)		
36	d) $P(X \le 124) =$			=DISTR.NORM(24;B30;B31;1)		
37	e) $P[77.5 \le X \le 100]$	$P[X \le 100]-P[X<=77.5]$		=DISTR.NORM(100;B30;B31;1)-DISTR.NORM(77.5;B30;B31;1)		
38	f) $P[91 \le X \le 127]$	$P[X \le 127]-P[X<=91]$		=DISTR.NORM(127;B30;B31;1)-DISTR.NORM(91;B30;B31;1)		

UNIDAD 5. MUESTREO Y DISTRIBUCIONES MUESTRALES.

5.1 Muestreo

El análisis de la información que resultan de los procesos de las organizaciones es importante para la toma de decisiones y controlar la gestión.

En la mayoría de los casos resulta casi imposible tomar toda la información de esos procesos, o bien por incapacidad física o por falta de tiempo o por aspectos económicos, etc.

Cuando no se puede hacer un censo para el análisis global de la población, se recurre al proceso estadístico por medio del cual se toma una parte de la población, bajo ciertos parámetros, se toma la información requerida y a partir de ahí se sacan conclusiones del comportamiento del universo o población. Esta técnica se conoce como muestreo.

El muestreo lo aplican en la investigación científica, en los estudios de mercados en los análisis sociales. Se puede decir que la función básica es determinar que parte de una realidad en estudio (población o universo) debe examinarse con la finalidad de hacer inferencias sobre dicha población.

Cuando se aplica esta técnica se está expuesto a que los valores de la muestra no correspondan con los parámetros de población, es decir se está expuesto a lo que se conoce como el error del muestreo. Con la muestra adecuada SE trata de conocer los rasgos de la los parámetros poblacionales.

5.1.1 Metodología para hacer muestreo.

Para realizar un muestreo se debe realizar un proceso que va desde la planeación y termina con la inferencia de los resultados. En eso proceso metodológico para realizar el muestreo se debe tener en cuenta ls siguientes características.
Tamaño de la Población: Finita o infinita.
Objetivos de la investigación
Preguntas a realizar
Tipos de de preguntas y respuestas
Tipos de muestreo
Forma de recolectar la información
Recurso humano.
Recursos económicos
Recursos físicos.
Recursos tecnológicos.
Cronograma.

5.1.2 Tipos de muestreo

Según la forma como se seleccione los elementos a encuestas el muestreo se puede clasificar en:

- Muestreo No probabilístico
- Muestreo Probabilístico o aleatorio.
    - Muestreo aleatorio simple.

- o Muestreo aleatorio sistemático
- o Muestreo aleatorio estratificado
- o Muestreo aleatorio por conglomerados

En el Método de Muestreo no probabilísticos, se seleccionan a los sujetos siguiendo determinados criterios personales procurando que la muestra sea representativa.

Los métodos de **muestreo probabilísticos** son aquellos que se basan en el principio del azar, es decir todo los elementos tienen la misma probabilidad de ser elegidos para formar parte de una muestra. Los métodos de muestreo probabilísticos nos aseguran la representatividad de la muestra extraída y son, por tanto, los más recomendables.

En el Muestreo Aleatorio Simple se asigna un número a cada elemento de la población y a través de algún medio mecánico (papeletas, números aleatorios generados por la calculadora o e l Excel, etc) se eligen tantos sujetos como sea necesario para completar el tamaño de muestra requerido.

En el Muestreo Aleatorio Sistemático todos los elementos de la población, deben estar ordenados en una lista en la cual se toman grupos para seleccionarlos elementos al azar haciendo desplazamientos sistemáticos. La manera de la selección depende del número de elementos incluidos en la población y el tamaño de la muestra. El número de elementos en la población es, primero, dividido por el número deseado en la muestra. El cociente indicará si cada décimo, cada onceavo, o cada centésimo elemento en la población va a ser seleccionado. El primer elemento de la muestra es seleccionado al azar.
La muestra sistemática puede dar la misma precisión de estimación acerca de la población, que una muestra aleatoria simple cuando los elementos en la población están ordenados al azar.

En el muestreo aleatorio estratificado se divide la población en grupos homogéneos al interior, de donde se extraen al azar de cada grupo un cierto número de elementos. Lo que se pretende con este tipo de muestreo es asegurarse de que todos los estratos de interés estarán representados adecuadamente en la muestra. Cada estrato funciona independientemente, pudiendo aplicarse dentro de ellos el muestreo aleatorio simple o el estratificado para elegir los elementos concretos que formarán parte de la muestra. Las estimaciones de la población, basadas en la muestra estratificada, usualmente tienen mayor precisión (o menor error muestral) que si la población entera muestreada mediante muestreo aleatorio simple. El número de elementos seleccionado de cada estrato puede ser proporcional o desproporcional al tamaño del estrato en relación con la población.

En el Muestreo Aleatorio por Conglomerados el universo se divide en grupos que son muy heterogéneos al interior y homogéneos al exterior. Luego se selecciona y luego se toman todos los elementos del grupo o parte de ellos.

Una muestra de conglomerados, usualmente produce un mayor error muestral que una muestra aleatoria simple del mismo tamaño. En una localidad pueden vivir gente pobre como también pueden vivir puede vivir gente acomodada.

Muestreo con reposición y sin reposición.

Cuando un elemento se toma para ser parte de la muestra, con el se puede tomar la decisión de no regresarlo a la población y tener la opción de volverlo a seleccionar o excluirlo definitivamente de la población. En el primer caso se habla de muestreo con reposición y en el segundo muestreo con reposición. En el primer caso la probabilidad entre u ensayo y otro se mantiene constante, es decir equivale a trabajar con una población infinita.

5.2 Distribuciones de medias muestrales.

Una distribución de medias muestrales consiste en tomar todas las posibles muestra de cierto tamaño de una población y con ellas hacer el cálculo de las medias para encontrar la distribución de probabilidad.

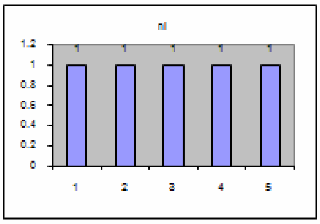
Como ejemplo podemos suponer que una población está compuesta por cinco elementos cuyos valores son S = {1,2,3,4,5}.

La media de esa población es $\mu = 3.0$ con una varianza de $\sigma^2 = 2.0$

Se van a tomar todas las muestras de tamaño 2 con reemplazamiento lo que equivale a trabajar con una de tamaño infinito, pues la probabilidad entre un ensayo y otro no cambia.

Los cálculos se encuentran en el archivo de Distri muestal-xls

	A	B	C	D	E	F	G
1	Valor de la variable : 1 - 2 - 3 - 4 - 5						
2							
3	Xi	ni					
4	1	1					
5	2	1					
6	3	1					
7	4	1					
8	5	1					
9		5					
10							
11	Media = μ =		3.0	=PROMEDIO(A4:A8)			
12	Varianza = σ^2 =		2.0	=VARP(A4:A8)			
13	Desviación = σ		1.414	=raiz(C12)			



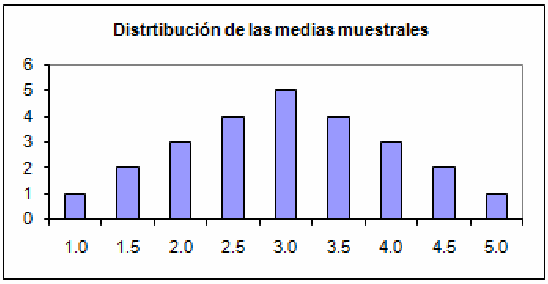
MUESTREO DE TAMAÑO 2 CON REEMPLAZAMIENTO (POBLACION INFINITA)

X1	X2	MEDIA
1	1	1.0
1	2	1.5
1	3	2.0
1	4	2.5
1	5	3.0
2	1	1.5
2	2	2.0
2	3	2.5
2	4	3.0
2	5	3.5
3	1	2.0
3	2	2.5
3	3	3.0
3	4	3.5
3	5	4.0
4	1	2.5
4	2	3.0
4	3	3.5
4	4	4.0
4	5	4.5
5	1	3.0
5	2	3.5
5	3	4.0
5	4	4.5
5	5	5.0

Medias muestrales	Frecuencia relativa	Xi*ni	$(Xi - \mu)^2$ *ni
1.0	1	1	4.0
1.5	2	3	4.5
2.0	3	6	3.0
2.5	4	10	1.0
3.0	5	15	0.0
3.5	4	14	1.0
4.0	3	12	3.0
4.5	2	9	4.5
5.0	1	5	4.0
	25	75	25

Media muestral = μ =	3.00	3.00
Varianza muestral = σ_x^2 =	1.00	1.00
Desviación muestral = σ	1.00	1.00

$$\mu = \mu_{\bar{x}} \qquad \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}; \quad \sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$



Se observa que la media Poblacional (μ) es igual a la media de todas la media de tamaño 2 ($(\mu_{\bar{x}} = 3.0)$ = 3.0).

No ocurre lo mismo con la varianza, pues la varianza poblacional es 2.0; σ^2 = 2.0, mientras que la varianza de las medias muestrales es 1.0

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{2}{2} = 1$$

$$Si\ \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\ entonces\ \sigma = \frac{\sigma}{\sqrt{n}},$$

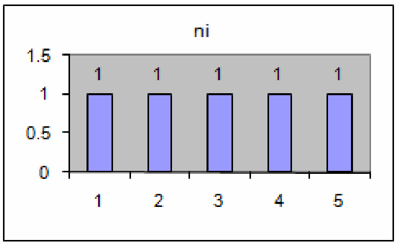
$$donde\ \frac{\sigma}{\sqrt{n}} \cdots es\ el\ error\ es\tan dar\ del\ muestreo\ para\ poblaciones \inf initas.$$

Si ahora tomamos todas las muestras posibles de tamaño 2 pero con reemplazamiento, loo que equivale a decir que se tiene una población finita, pues la probabilidad entre un ensayo y otro cambia.

Los resultados se pueden ver en el archivo XXXXX

Valor de la variable : 1 - 2 - 3 - 4 - 5

Xi	ni
1	1
2	1
3	1
4	1
5	1
	5



Media = μ =	3.0
Varianza = σ^2 =	2.0
Desviación = σ	1.414

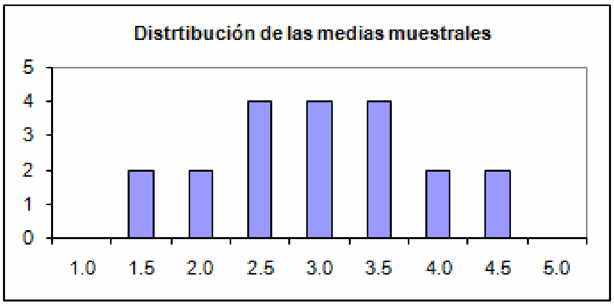
MUESTREO DE TAMAÑO 2 SIN REEMPLAZAMIENTO (POBLACION FINITA)

X1	X2	MEDIA
1	2	1.5
1	3	2.0
1	4	2.5
1	5	3.0
2	1	1.5
2	3	2.5
2	4	3.0
2	5	3.5
3	1	2.0
3	2	2.5
3	4	3.5
3	5	4.0
4	1	2.5
4	2	3.0
4	3	3.5
4	5	4.5
5	1	3.0
5	2	3.5
5	3	4.0
5	4	4.5

Medias muestrales	Frecuencia relativa	Xi*ni	$(Xi - \mu)^2$ *ni
1.0	0	0	0.0
1.5	2	3	4.5
2.0	2	4	2.0
2.5	4	10	1.0
3.0	4	12	0.0
3.5	4	14	1.0
4.0	2	8	2.0
4.5	2	9	4.5
5.0	0	0	0.0
	20	60	15

Media muestral = μ =	3.00	3.00
Varianza muestral = σ_x^2 =	0.75	0.75
Desviación muestral = σ	0.87	0.87

$$\mu = \mu_{\bar{X}} \qquad \sigma^2_{\bar{X}} = \frac{\sigma^2}{n} * \frac{N-n}{N-1}$$



Se observa que la media muestral es igual a la media poblacional $\mu_{\bar{X}} = \mu$

Pero la relación entre la varianza de las medias muestrales y la varianza poblacional es

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n} * \frac{N-n}{N-1} ; donde \ \frac{N-n}{N-1} ; se\ le\ conoce\ como\ factor\ de\ correción\ para\ población\ finita$$

El mismo procedimiento se puede aplicar la misma población pero tomando muestras de tamaño 3.

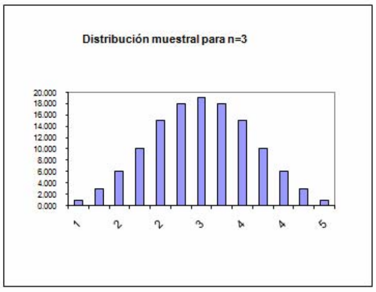
MUESTREO DE TAMAÑO 3 CON REEMPLAZAMIENTO (POBLACION INFINITA)

X1	X2	X3	MEDIA
1	1	1	1.000
1	1	2	1.333
1	1	3	1.667
1	1	4	2.000
1	1	5	2.333
1	2	1	1.333
1	2	2	1.667
1	2	3	2.000
1	2	4	2.333
1	2	5	2.667
1	3	1	1.667
1	3	2	2.000
1	3	3	2.333
1	3	4	2.667
1	3	5	3.000
1	4	1	2.000
1	4	2	2.333
1	4	3	2.667
1	4	4	3.000
1	4	5	3.333
1	5	1	2.333
1	5	2	2.667
1	5	3	3.000
1	5	4	3.333
1	5	5	3.667
2	1	1	1.333

Medias muestrales	Frecuencia relativa	Xi*ni	$(Xi - \mu)^2$ *ni
1.000	1.000	1.000	4.000
1.333	3.000	4.000	8.333
1.667	6.000	10.000	10.667
2.000	10.000	20.000	10.000
2.333	15.000	35.000	6.667
2.667	18.000	48.000	2.000
3.000	19.000	57.000	0.000
3.333	18.000	60.000	2.000
3.667	15.000	55.000	6.667
4.000	10.000	40.000	10.000
4.333	6.000	26.000	10.667
4.667	3.000	14.000	8.333
5.000	1.000	5.000	4.000
	125.000	375.000	83.333

Media muestral = μ =	3.000	3.000
Varianza muestral = σ_x^2 =	0.667	0.667
Desviación muestral = σ	0.816	0.816

$$\mu = \mu_{\overline{X}} \qquad \sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}; \quad \sigma_{\overline{X}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$



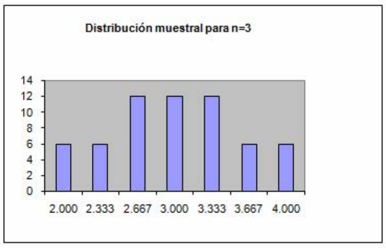
MUESTREO DE TAMAÑO 3 CON REEMPLAZAMIENTO (POBLACION FINITA)

1	2	3	2.000
1	2	4	2.333
1	2	5	2.667
1	3	2	2.000
1	3	4	2.667
1	3	5	3.000
1	4	2	2.333
1	4	3	2.667
1	4	5	3.333
1	5	2	2.667
1	5	3	3.000
1	5	4	3.333
2	1	3	2.000
2	1	4	2.333
2	1	5	2.667
2	3	1	2.000
2	3	4	3.000
2	3	5	3.333
2	4	1	2.333
2	4	3	3.000
2	4	5	3.667
2	5	1	2.667
2	5	3	3.333

Medias muestrales	Frecuencia relativa	Xi*ni	$(Xi - \mu)^2$ *ni
2.000	6	12.000	6.000
2.333	6	14.000	2.667
2.667	12	32.000	1.333
3.000	12	36.000	0.000
3.333	12	40.000	1.333
3.667	6	22.000	2.667
4.000	6	24.000	6.000
	60	180	20

Media muestral = μ =	3.000	3.000
Varianza muestral = σ_x^2 =	0.333	0.333
Desviación muestral = σ	0.577	0.577

$$\mu = \mu_{\overline{X}} \qquad \sigma_{\overline{X}}^2 = \frac{\sigma^2}{n} * \frac{N-n}{N-1}$$



Un resultado importante de este análisis es que a pesar que la distribución de probabilidades de la población original es uniforme, las distribuciones de probabilidad de las medias muestrales tienen forma normal y entre mayor sea el tamaño de la muestras, mayor es la aproximación.

5.2 Teorema del Límite central

De los anteriores resultados se obtiene una conclusión importante para determinar el tamaño de la muestra.

E! teorema de límite central establece que para muestras aleatorias grandes, la distribución de probabilidad para las medias muéstrales, se acerca a una distribución del tipo normal. La aproximación es más exacta en la media que la muestra es más grande que para pequeñas. Esta es una de las conclusiones más útiles en Estadística. Se puede razonar acerca de la distribución de las medias muéstrales sin contar con alguna información respecto de la forma de la distribución original de la cual se toma la muestra. En otras palabras, el teorema de límite central es cierto para todas las distribuciones.

El teorema del límite central dice que *al seleccionar muestras aleatorias simples de tamaño n de una población, la distribución muestral de la media muestral* $\overline{X}$ *se puede aproximar a una distribución de probabilidad normal, cuando el tamaño de la muestra es grande.*

Si la población está distribuida normalmente, entonces, para cualquier tamaño de muestra, la distribución de la media muestral también será normal. Si la no es normal, es posible que se necesiten muestras de al menos 30 elementos para observar el aspecto de normalidad. La mayor parte de los estadísticos consideran que una muestra de 30 o mayor, es suficiente para que se emplee el teorema de límite central.

5.4 Determinación del tamaño de la muestra

Partiendo que el muestreo es la forma científica para que a partir de los resultados obtenidos en una muestra se infiera valores para la población, la muestra debe ser representativa de la población en lo que se refiere a la característica en estudio, o sea, la distribución de la característica en la muestra debe ser aproximadamente igual a la distribución de la característica en la población. Cualquier información obtenida de una muestra se le conoce como **estadístico** y a partir de él se puede estimar los valores equivalentes poblacionales o **parámetro.**

La **Inferencia estadística** se trata de estimar los parámetros poblacionales a partir de las estadísticas obtenidas en la muestra. Para que los resultados obtenidos de los datos muestrales se puedan extender a la población, la muestra debe ser representativa de la población en lo que se refiere a la característica en estudio, o sea, la distribución de la característica en la muestra debe ser aproximadamente igual a la distribución de la característica en la población.

La diferencia entre el estadístico y el parámetro poblacional se le conoce como el error muestral o de estimación. Midiendo la variabilidad de las estimaciones de muestras repetidas en torno al valor de la población, nos da una noción clara de hasta dónde y con qué probabilidad una estimación basada en una muestra se aleja del valor que se hubiera obtenido por medio de un censo completo. Siempre se habrá un error, pero así podemos saber hasta qué medida podemos cometerlo (los resultados se someten a error muestral e intervalos de confianza que varían muestra a muestra). El estadístico será más preciso en cuanto y tanto su error es más pequeño.

La representatividad en estadística se logra con el tipo de muestreo adecuado que siempre incluye la aleatoriedad en la selección de los elementos de la población que formaran la muestra. Sin embargo, tales métodos solo nos garantizan una representatividad muy probable pero no completamente segura.

Ahora trataremos de resolver a pregunta ¿Cuál es el número adecuado de elemento que se debe analizar u menor error?

Si con el muestreo se trata de estimar la media poblacional de la variable característica de población, se debe tener en cuenta la varianza, bien sea que se conozca de antemano no. El error estándar del muestreo está asociado con la varianza de la población.

Esto nos lleva a que en la determinación del tamaño d e la muestra de debe tener en cuenta los siguientes aspectos

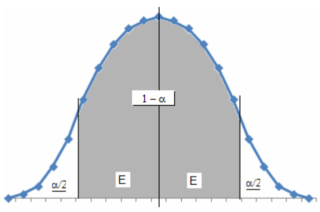
Nivel de Confianza. Es la probabilidad de que el parámetro poblacional esté en el error estimado Cualquier información que queremos recoger está distribuida según una ley de probabilidad (Gauss o Student), así llamamos nivel de confianza a la probabilidad de que el intervalo construido en torno a un estadístico capte el verdadero valor del parámetro.

Varianza Poblacional. Cuando una población es más homogénea la varianza es menor y el número de entrevistas necesarias para construir un modelo reducido del universo, o de la población, será más pequeño. Generalmente es un valor desconocido y hay que estimarlo a partir de datos de estudios previos.

La representatividad en estadística se logra con el tipo de muestreo adecuado que siempre incluye la aleatoriedad en la selección de los elementos de la población que formaran la muestra. No obstante, tales métodos solo nos garantizan una representatividad muy probable pero no completamente segura.

Después de estos preliminares imprescindibles es posible pasa a tratar algunas de las formas que desde el punto de vista científico se puede extraer una muestra.

En la determinación del tamaño de la muestra se busca que con un nivel de confianza y un cierto error asociado con el tipo de población



El primer factor se llama nivel de confianza (1-α) , que en la probabilidad que la media poblacional esté en un cierto intervalo asociado con el error que se puede permitir.

Tamaño de la muestra para estimar la media poblacional para poblaciones infinitas está dada por

$$n = \frac{Z_{1-\alpha/2}^{2} * \sigma^{2}}{E^{2}}$$

Donde Z es el nivel de confianza y E es el error estándar permitido.

Tamaño de la muestra para estimar la media poblacional para poblaciones finitas está dada por

$$n = \frac{N^2 * Z_{1-\alpha/2}^2 * \sigma^2}{E^2 * (N-1) + Z_{1-\alpha/2}^2 * \sigma^2}$$

Donde Z es el nivel de confianza y E es el error estándar permitido y N es el tamaño de la población

Tamaño de la muestra para estimar la proporción poblacional para población infinitas está dada por.

$$n = \frac{Z_{1-\alpha/2}^2 * P * Q}{\varepsilon^2}$$

Donde Z es el nivel de confianza, P es la probabilidad de éxito, Q=1-P, la probabilidad de fracaso. y ε es el error estándar permitido.

Tamaño de la muestra para estimar la proporción poblacional para población infinitas está dada por.

$$n = \frac{N^2 * Z_{1-\alpha/2}^2 * P * Q}{\varepsilon^2 * (N-1) + Z_{1-\alpha/2}^2 * P * Q}$$

Donde Z es el nivel de confianza, P es la probabilidad de éxito, Q=1-P, la probabilidad de fracaso, ε es el error estándar permitido y N es el tamaño de la población

Ejemplo 1. Encuestas pasadas indican que los clientes de una empresa pedían en promedio 2500 unidades con una desviación de 600 unidades. Para actualizar la opinión de los clientes hacia la empresa se va a aplicar una encuesta a una parte de los elementos de la población. De qué tamaño debe ser la muestra si se quiere un nivel de confianza del 90% y un error de 100.

En este caso. $\sigma = 600$, 1-a = 0.90; E = 100; $1-\alpha/2 = 0.95$; Z = 1.64

$n = \dfrac{1.64^2 * 600^2}{100^2} = 96.82;$ Se deben aplicar 97 encuestas.

Cuando no se conoce el valor de la varianza σ^2, poblacional inicial, se puede encontrar a través de una prueba piloto.

Ejemplo 2. En una reciente encuesta realizada a los consumidores de un producto se obtuvo que el 80% estaban satisfechos con la calidad. El gerente del producto quiere actualizar esta información, por lo que le ha pedido a usted que realice una encuesta que tenga un nivel de confianza del 0.95 y un error del 5%. ¿De qué tamaño deberá de ser la muestra?

En este caso. P = 0.8, 1-α = 0.95; E = 100; 1-α/2 = 0.975; Z = 1.64

$$n = \frac{1.96^2 * 0.8 * 0.2}{(0.05)^2} = 245.8624.$$ Se deben hacer 246 encuestas.

Cuando no se conoce el valor de la proporción poblacional inicial, se puede encontrar a través de una prueba piloto o simplemente tomando a P = 0.5 y Q = 0.5

Las respuestas de los ejercicios anteriores se encontraron utilizando el Excel.

	A	B	C	D	E	F	G	H	I	J
1	Ejemplo 1					Ejemplo 2				
2	σ =	600				P =	0.8			
3	1 - α =	0.900				Q =	0.2			
4	α / 2 =	0.050	=(1-B3)/2			1 - α =	0.950			
5	1 - α/2 =	0.950	=1-B4			α / 2 =	0.025	=(1-B3)/2		
6	E =	100				1 - α/2 =	0.975	=1-B4		
7	Z =	1.64485363	=DISTR.NORM.ESTAND.INV(B5)			ε =	0.05			
8						Z =	1.959963985	=DISTR.NORM.ESTAND.INV(B5)		
9	n =	97.3995643	=B6^2*B1^2 /B5^2							
10						n =	245.8533645	=B6^2*B1^2 /B5^2		
11										

El siguiente figura se muestra un análisis de sensibilidad en Excel para determinar el tamaño de la muestra cuando se quiera estimar la media poblacional.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
22															
23		Nivel de confianza													
24		1 – α	0.70	0.71	0.72	0.73	0.74	0.75	0.76	0.77	0.78	0.79	0.80	0.81	0.82
25		α	0.3	0.29	0.28	0.27	0.26	0.25	0.24	0.23	0.22	0.21	0.2	0.19	0.18
26		α/2	0.15	0.145	0.14	0.135	0.13	0.125	0.12	0.115	0.11	0.105	0.1	0.095	0.09
27		Probabilidad	0.85	0.855	0.86	0.865	0.87	0.875	0.88	0.885	0.89	0.895	0.9	0.905	0.91
28		Z	1.036433389	1.05812162	1.08031934	1.10306256	1.12639113	1.15034938	1.174987	1.200359	1.226528	1.253565	1.281552	1.310579	1.340755
29															
30			Nivel de confianza												
31			0.70	0.71	0.72	0.73	0.74	0.75	0.76	0.77	0.78	0.79	0.80	0.81	0.82
32		1.00%	=C28^2*0.5*0.5/B32^2	2799.1	2917.7	3041.9	3171.9	3308.3	3451.5	3602.2	3760.9	3928.6	4105.9	4294.0	4494.1
33		2.00%	671.4	699.8	729.4	760.5	793.0	827.1	862.9	900.5	940.2	982.1	1026.5	1073.5	1123.5
34		3.00%	298.4	311.0	324.2	338.0	352.4	367.6	383.5	400.2	417.9	436.5	456.2	477.1	499.3
35		4.00%	167.8	174.9	182.4	190.1	198.2	206.8	215.7	225.1	235.1	245.5	256.6	268.4	280.9
36		5.00%	107.4	112.0	116.7	121.7	126.9	132.3	138.1	144.1	150.4	157.1	164.2	171.8	179.8
37		6.00%	74.6	77.8	81.0	84.5	88.1	91.9	95.9	100.1	104.5	109.1	114.1	119.3	124.8
38		7.00%	54.8	57.1	59.5	62.1	64.7	67.5	70.4	73.5	76.8	80.2	83.8	87.6	91.7
39		8.00%	42.0	43.7	45.6	47.5	49.6	51.7	53.9	56.3	58.8	61.4	64.2	67.1	70.2
40		9.00%	33.2	34.6	36.0	37.6	39.2	40.8	42.6	44.5	46.4	48.5	50.7	53.0	55.5
41	Error	10.00%	26.9	28.0	29.2	30.4	31.7	33.1	34.5	36.0	37.6	39.3	41.1	42.9	44.9
42		11.00%	22.2	23.1	24.1	25.1	26.2	27.3	28.5	29.8	31.1	32.5	33.9	35.5	37.1
43		12.00%	18.6	19.4	20.3	21.1	22.0	23.0	24.0	25.0	26.1	27.3	28.5	29.8	31.2
44		13.00%	15.9	16.6	17.3	18.0	18.8	19.6	20.4	21.3	22.3	23.2	24.3	25.4	26.6
45		14.00%	13.7	14.3	14.9	15.5	16.2	16.9	17.6	18.4	19.2	20.0	20.9	21.9	22.9
46		15.00%	11.9	12.4	13.0	13.5	14.1	14.7	15.3	16.0	16.7	17.5	18.2	19.1	20.0
47		16.00%	10.5	10.9	11.4	11.9	12.4	12.9	13.5	14.1	14.7	15.3	16.0	16.8	17.6
48		17.00%	9.3	9.7	10.1	10.5	11.0	11.4	11.9	12.5	13.0	13.6	14.2	14.9	15.6
49		18.00%	8.3	8.6	9.0	9.4	9.8	10.2	10.7	11.1	11.6	12.1	12.7	13.3	13.9
50		19.00%	7.4	7.8	8.1	8.4	8.8	9.2	9.6	10.0	10.4	10.9	11.4	11.9	12.4
51		20.00%	6.7	7.0	7.3	7.6	7.9	8.3	8.6	9.0	9.4	9.8	10.3	10.7	11.2
52															

El siguiente figura se muestra un análisis de sensibilidad en Excel para determinar el tamaño de la muestra cuando se quiera estimar el promedio poblacional.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
19		Ingresos de la familias Colmbianas													
20		σ =	$ 235,760.00												
21		Nivel de confianza													
22		1 – α	0.70	0.71	0.72	0.73	0.74	0.75	0.76	0.77	0.78	0.79	0.80	0.81	0.82
23		α	0.3	0.29	0.28	0.27	0.26	0.25	0.24	0.23	0.22	0.21	0.2	0.19	0.18
24		α/2	0.15	0.145	0.14	0.135	0.13	0.125	0.12	0.115	0.11	0.105	0.1	0.095	0.09
25		Probabilidad	0.85	0.855	0.86	0.865	0.87	0.875	0.88	0.885	0.89	0.895	0.9	0.905	0.91
26		Z	1.036433389	1.05812162	1.08031934	1.10306256	1.12639113	1.15034938	1.174987	1.200359	1.226528	1.253565	1.281552	1.310579	1.340755
27															
28			Nivel de confianza												
29			0.70	0.71	0.72	0.73	0.74	0.75	0.76	0.77	0.78	0.79	0.80	0.81	0.82
30		5000	2388.3	2489.3	2594.8	2705.2	2820.8	2942.1	3069.5	3203.5	3344.7	3493.8	3651.5	3818.8	3996.7
31		6000	1658.5	1728.7	1801.9	1878.6	1958.9	2043.1	2131.6	2224.6	2322.7	2426.2	2535.8	2651.9	2775.5
32		7000	1218.5	1270.0	1323.9	1380.2	1439.2	1501.1	1566.1	1634.4	1706.5	1782.5	1863.0	1948.4	2039.1
33		8000	932.9	972.4	1013.6	1056.7	1101.9	1149.3	1199.0	1251.4	1306.5	1364.8	1426.4	1491.7	1561.2
34		9000	737.1	768.3	800.9	834.9	870.6	908.1	947.4	988.7	1032.3	1078.3	1127.0	1178.6	1233.5
35		10000	597.1	622.3	648.7	676.3	705.2	735.5	767.4	800.9	836.2	873.4	912.9	954.7	999.2
36		11000	493.4	514.3	536.1	558.9	582.8	607.9	634.2	661.9	691.1	721.9	754.4	789.0	825.8
37		12000	414.6	432.2	450.5	469.7	489.7	510.8	532.9	556.2	580.7	606.6	633.9	663.0	693.9
38		13000	353.3	368.2	383.8	400.2	417.3	435.2	454.1	473.9	494.8	516.8	540.2	564.9	591.2
39	Error	14000	304.6	317.5	331.0	345.1	359.8	375.3	391.5	408.6	426.6	445.6	465.8	487.1	509.8
40		15000	265.4	276.6	288.3	300.6	313.4	326.9	341.1	355.9	371.6	388.2	405.7	424.3	444.1
41		16000	233.2	243.1	253.4	264.2	275.5	287.3	299.8	312.8	326.6	341.2	356.6	372.9	390.3
42		17000	206.6	215.3	224.5	234.0	244.0	254.5	265.5	277.1	289.3	302.2	315.9	330.3	345.7
43		18000	184.3	192.1	200.2	208.7	217.7	227.0	236.8	247.2	258.1	269.6	281.8	294.7	308.4
44		19000	165.4	172.4	179.7	187.3	195.3	203.7	212.6	221.8	231.6	242.0	252.9	264.5	276.8
45		20000	149.3	155.6	162.2	169.1	176.3	183.9	191.8	200.2	209.0	218.4	228.2	238.7	249.8
46		21000	135.4	141.1	147.1	153.4	159.9	166.8	174.0	181.6	189.6	198.1	207.0	216.5	226.6
47		22000	123.4	128.6	134.0	139.7	145.7	152.0	158.5	165.5	172.8	180.5	188.6	197.3	206.4
48		23000	112.9	117.6	122.6	127.8	133.3	139.0	145.1	151.4	158.1	165.1	172.6	180.5	188.9

Unidad 6. Estimación por intervalo.

6.1 Intervalos de confianza.

En la determinación del tamaño de la muestra se tuvieron en cuenta el nivel de confianza y el error buscando que al estimar los parámetros poblacionales fueran lo ms confiables posibles.
Una vez aplicada la encuesta a la muestra se debe hacer las estimaciones de los parámetros poblacionales.
Para hacer esas estimaciones se puede utilizar cualquiera de los estadísticos de la muestra. El estadístico que se usa para hacer la estimación del parámetro poblacional se le denomina **estimador puntal**. Usualmente se usan como estimadores puntales la muestra. La media de la muestra es una estimación puntual de la media población. La media muestral no es el único valor que se podría usar para estimar la media poblacional. También se podría usar la mediana muestral, aunque no es tan eficiente, lo que significa que hay más dispersión en la distribución de las dispersiones.
Los estimadores puntuales usados son:

Estadístico muestral		Parámetro poblacional
$\overline{X}$	$\longrightarrow$	μ
$\overline{p}$	$\longrightarrow$	***P***
S	$\longrightarrow$	σ^2

La media muestral es una estimación puntual de la media poblacional. $\overline{p}$ es una puntual de la proporción poblacional y s, la desviación estándar muestral, es una estimación puntual de σ, la desviación estándar poblacional.

Pero como la estimación puntual no da mucha información acerca del parámetro poblacional, se necesita mayor información por lo que el intervalo de confianza cumple este propósito.

El intervalo de confianza Es el conjunto de valores obtenido a partir de los datos muestrales en el que hay una determinada probabilidad de que se encuentre el parámetro poblacional. Esta probabilidad se le conoce como el nivel de confianza.
Por ejemplo, en una encuesta se encontró que en una determinada región el ingreso mensual promedio de los trabajadores de la construcción es 2.5 SMLM. Un intervalo podrá ser que el salario promedio global este entre 2 y 3 SMLM. Y ¿cuál es la seguridad que eso sea así? Se podrá indicar que se tiene una seguridad del 90% que eso es así.

La información que se tiene acerca de la forma de la distribución muestral de la media muestral, es decir de la distribución muestral de X, permite localizar un intervalo que tenca una determinada probabilidad de contener a la media poblacional.
Si el tamaño de la muestra es razonablemente grande, el teorema del límite central permite establecer lo siguiente:

1. El noventa y cinco por ciento de las medias muéstrales obtenidas de una. población se e encuentran a no más de 1.96 desviaciones estándar de la media poblacionalμ.
2. El noventa y nueve por ciento de las medias poblacionales se encuentran a no más de 2.58 desviaciones estándar de la media poblacional

La probabilidad de que el verdadero valor del parámetro se encuentre en el intervalo construido se denomina nivel de confianza, y se denota 1-α. La probabilidad de equivocarnos se llama nivel de significancia y se simboliza con α. Generalmente se construyen intervalos con confianza 1-α = 95% (o significancia =5%). Menos frecuentes son los intervalos con α = 10% o α = 1%.

En la distribución normal P(-1.96 < z < 1.96) = 0.95, lo que indica que si una variable tiene distribución normal N(μ; σ), entonces el 95% de las muestras

cumplen que $-1.96 \leq \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \leq 1.96$, donde $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$, error estándar del

muestreo

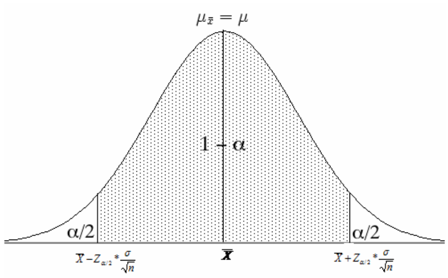
Despejando se tiene que $\overline{X} - 1.96 * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96 * \frac{\sigma}{\sqrt{n}}$ con una probabilidad

de 1-α

Esto quiere decir que de 100 muestras de tamaño n, 95 contienen la media poblacional μ, cuando la variable X es normal y se conoce σ.

6.2 Estimación para la media poblacional. Muestras grandes.

Cuando el tamaño de la muestra es grande o la variable tiene distribución normal, el intervalo de confianza está dado por



$$\bar{X} - Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

$$P\left[\bar{X} - Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

Como el error estándar está afectado por el tamaño de la muestra, este a su vez afecta el intervalo de confianza. Conforme aumenta el tamaño de la muestra, el error estándar disminuye, indicando esto que hay menos variabilidad en la distribución muestral de a media muestral. La estimación obtenida de una muestra grande será más precisa que una estimación obtenida de una muestra pequeña.

Cuando el tamaño de la muestra n es mayor o igual a 30, se aplica el teorema de límite central asegura que la media muestral sigue la distribución normal. Si la media muestral tiene una distribución normal, se puede usar la distribución normal estándar, es decir, z, para los cálculos.

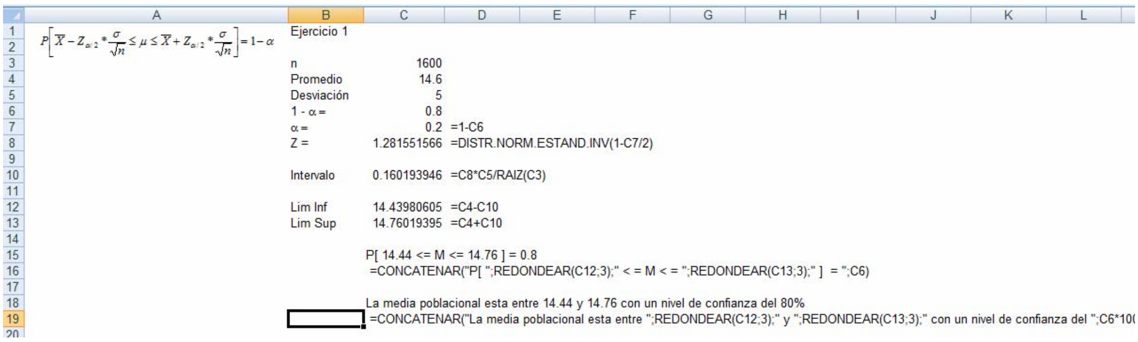
Ejemplo. En una encuesta aplicada a 1600 colombiano se encontró que en promedio ven 14.6 horas de televisión a la semana con una desviación estándar de 5 horas. Realice una estimación de la variable con para un nivel de confianza del 80%.

n = 1600; $\bar{X}$ = 14.6 horas; σ = 5 horas; 1 - α = 0.80; $Z_{0.80}$ = 1.28

$$P\left[14.6 - 1.28 * \frac{5}{\sqrt{1600}} \le \mu \le 14.6 + 1.28 * \frac{5}{\sqrt{1600}}\right] = 0.80$$

$$P[14.44 \le \mu \le 14.76] = 0.80$$

Los colombiano ven televisión en promedio entre 14.4 y 14.76 horas con un nivel de confianza del 80%



	A	B	C	D
1	$P\left[\bar{X} - Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$	Ejercicio 1		
2				
3		n	1600	
4		Promedio	14.6	
5		Desviación	5	
6		1 - α =	0.8	
7		α =	0.2	=1-C6
8		Z =	1.281551566	=DISTR.NORM.ESTAND.INV(1-C7/2)
9				
10		Intervalo	0.160193946	=C8*C5/RAIZ(C3)
11				
12		Lim Inf	14.43980605	=C4-C10
13		Lim Sup	14.76019395	=C4+C10
14				
15			P[14.44 <= M <= 14.76] = 0.8	
16			=CONCATENAR("P[";REDONDEAR(C12;3);" < = M < = ";REDONDEAR(C13;3);"] = ";C6)	
17				
18			La media poblacional esta entre 14.44 y 14.76 con un nivel de confianza del 80%	
19			=CONCATENAR("La media poblacional esta entre ";REDONDEAR(C12;3);" y ";REDONDEAR(C13;3);" con un nivel de confianza del ";C6*10(	

6.3 Estimación para la media poblacional. Muestras pequeñas.

Cuando el número de observaciones es menor de 30, la estimación del intervalo se basa en las suposiciones que si la población es normal o que si se conoce ls desviación estándar de la población.

En caso que la muestra sea pequeña, menor de 30, y se conozca la varianza de la población σ^2, el intervalo de confianza es.

$$\overline{X} - Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

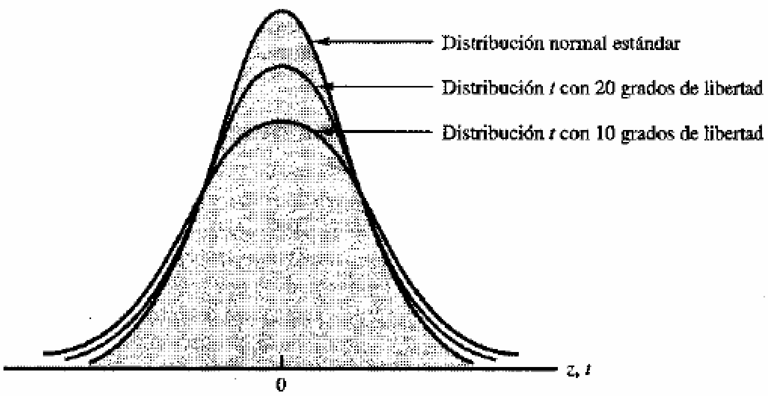
Cauno el tamaño de la muestra sea pequeño y no se conozca la varianza poblacional σ^2 , se utiliza la desviación estándar de la muestra *s*, y la distribución de probabilidad t.

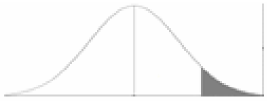
La distribución de probabilidad t o distribución t de Student es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño. Ésta es la base del popular test de la t de Student para la determinación de las diferencias entre dos medias muestrales y para la construcción del intervalo de confianza para la diferencia entre las medias de dos poblaciones.

La distribución t surge, en la mayoría de los estudios estadísticos prácticos, cuando la desviación típica de una población se desconoce y debe ser estimada a partir de los datos de una muestra.

La distribución te tiene las siguientes características.

- Al igual que la distribución Z, es una distribución continua
- La distribución t tiene una media de cero, es simétrica respecto de la media y se extiende de - $\propto$ a + $\propto$ la varianza de t es $\frac{v}{v+2}$ para v > 2. Cuando los grados de libertad son suficientemente grandes la varianza de la distribución t tiende a 1.
- Tiene forma acampanada y simétrica.
- No hay una distribución t, sino una "familia" de distribuciones t. todas con la misma media cero, pero con su respectiva desviación estándar diferente de acuerdo con el tamaño de la muestra n. Existe una distribución t para una muestra de 20, otra para una muestra de 22, y así sucesivamente.
- La distribución t es más ancha y más plana en el centro que la distribución normal estándar como resultado de ello se tiene una mayor variabilidad en las medias de muestra calculadas a partir de muestras más pequeñas. Sin embargo, a medida que aumenta el tamaño de la muestra, la distribución t se aproxima a la distribución normal estándar.





GL	Estimación intervalos de confianza									
	70%	75%	80%	85%	90%	95%	96%	98%	98%	99.00%
1	1.96261	2.41421	3.07768	4.16530	6.31375	12.70620	15.89454	31.82052	31.82052	63.65674
2	1.38621	1.60357	1.88562	2.28193	2.91999	4.30265	4.84873	6.96456	6.96456	9.92484
3	1.24978	1.42263	1.63774	1.92432	2.35336	3.18245	3.48191	4.54070	4.54070	5.84091
4	1.18957	1.34440	1.53321	1.77819	2.13185	2.77645	2.99853	3.74695	3.74695	4.60409
5	1.15577	1.30095	1.47588	1.69936	2.01505	2.57058	2.75651	3.36493	3.36493	4.03214
6	1.13416	1.27335	1.43976	1.65017	1.94318	2.44691	2.61224	3.14267	3.14267	3.70743
7	1.11916	1.25428	1.41492	1.61659	1.89458	2.36462	2.51675	2.99795	2.99795	3.49948
8	1.10815	1.24032	1.39682	1.59222	1.85955	2.30600	2.44898	2.89646	2.89646	3.35539
9	1.09972	1.22966	1.38303	1.57374	1.83311	2.26216	2.39844	2.82144	2.82144	3.24984
10	1.09306	1.22126	1.37218	1.55924	1.81246	2.22814	2.35931	2.76377	2.76377	3.16927

Para la estimación del intervalo de confianza, el valor de t depende de los grados de libertad, n-1, (fila) y del nivel de confianza.

Por ejemplo para un niel de confianza de 90% y 5 grados de libertad, t = 2.01505

GL	Estimación interval				
	70%	75%	80%	85%	90%
1	1.96261	2.41421	3.07768	4.16530	6.31375
2	1.38621	1.60357	1.88562	2.28193	2.91999
3	1.24978	1.42263	1.63774	1.92432	2.35336
4	1.18957	1.34440	1.53321	1.77819	2.13185
5	1.15577	1.30095	1.47588	1.69936	2.01505

6.4 Intervalo de confianza para muestras pequeñas donde no se conoce la varianza poblacional.

$$\overline{X} - t*\frac{s}{\sqrt{n}} \leq \mu \leq \overline{X} + t*\frac{s}{\sqrt{n}} \quad ; \quad P\left[\overline{X} - t*\frac{s}{\sqrt{n}} \leq \mu \leq \overline{X} + t*\frac{s}{\sqrt{n}}\right] = 1 - \alpha$$

Ejemplo. En una encuesta aplicada a 10 personas de la calle se encontró que en promedio ven 2.5 horas de televisión al día con una desviación estándar de 0.8 horas. Con un nivel de confianza del 95% estime el intervalo d confianza para la media poblacional.

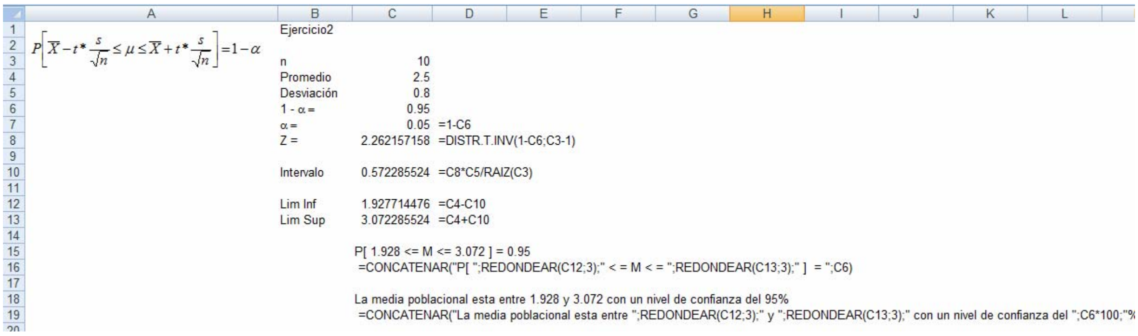
En ese caso

$$n = 10;\ \overline{X} = 2.5;\ s = 0.8\ horas; 1-\alpha = 0.95,\ por\ lo\ \tan to;\quad t_{9;0.95} = 2.262$$

$$P\left[2.5 - 2.262 * \frac{0.8}{\sqrt{10}} \le \mu \le \overline{2.5} + 2.262 * \frac{0.8}{\sqrt{10}}\right] = 0.95$$

$$P[1.92775 \le \mu \le 3.0722] = 0.95$$

Se estima que la población debe ver en promedio entre 1.93 horas y 3.07 horas de televisión al día, con un nivel de confianza del 95%



	A	B	C	D
1		Ejercicio2		
2	$P\left[\overline{X} - t^* \frac{s}{\sqrt{n}} \le \mu \le \overline{X} + t^* \frac{s}{\sqrt{n}}\right] = 1-\alpha$			
3		n	10	
4		Promedio	2.5	
5		Desviación	0.8	
6		1 - α =	0.95	
7		α =	0.05	=1-C6
8		Z =	2.262157158	=DISTR.T.INV(1-C6;C3-1)
9				
10		Intervalo	0.572285524	=C8*C5/RAIZ(C3)
11				
12		Lim Inf	1.927714476	=C4-C10
13		Lim Sup	3.072285524	=C4+C10
14				
15			P[1.928 <= M <= 3.072] = 0.95	
16			=CONCATENAR("P[";REDONDEAR(C12;3);" < = M < = ";REDONDEAR(C13;3);"] = ";C6)	
17				
18			La media poblacional esta entre 1.928 y 3.072 con un nivel de confianza del 95%	
19			=CONCATENAR("La media poblacional esta entre ";REDONDEAR(C12;3);" y ";REDONDEAR(C13;3);" con un nivel de confianza del ";C6*100;"%	

6.4 Estimación para la proporción poblacional. Muestras grandes

Todo el análisis para la estimación de la media poblacional para muestras grandes se aplica para determinar la proporción poblacional

$$\overline{p} - Z * \sqrt{\frac{\overline{p} * (1-\overline{p})}{n}} \le \mu \le \overline{p} + Z * \sqrt{\frac{\overline{p} * (1-\overline{p})}{n}} \quad ; \quad P\left[\overline{p} - Z * \sqrt{\frac{\overline{p} * (1-\overline{p})}{n}} \le \mu \le \overline{X} + Z * \sqrt{\frac{\overline{p} * (1-\overline{p})}{n}}\right] = 1-\alpha$$

Recuerde que $p = \dfrac{número\ de\ éxitos}{n}$

Ejemplo. La cadena de televisión TVK está considerando la posibilidad de sustituir una de sus series policiales por una serie de comedia con orientación familiar. Antes de tomar una decisión, los directores de la red toman una muestra de 400 televidentes. Después de ver la nueva serie, 250 indican que sí la verían y sugieren que reemplace a la serie policíaca. Estime el valor de la proporción poblacional con un nivel de 99%

n = 400;   X = 250; p = 400/250 = 0.625; q = 1-0.625 = 0.375; 1- α = 0.99; $Z_{0.995}$ = 2.58

$$P\left[0.625 - 2.58 * \sqrt{\frac{0.625 * 0.375}{400}} \le \mu \le 0.625 + 2.58 * \sqrt{\frac{0.625 * 0.375}{400}}\right] = 0.99$$

$$P[0.6008 \le \mu \le 0.6492] = 0.99$$

Entre el 60.08% y el 64.92% de los televidentes aceptna el cambio de programa

	A	B	C	D
1		Ejercicio 3		
2				
3		n	400	
4		x	250	
5	$P\left[\overline{p}-Z*\sqrt{\frac{\overline{p}*(1-\overline{p})}{n}}\leq\mu\leq\overline{X}+Z*\sqrt{\frac{\overline{p}*(1-\overline{p})}{n}}\right]=1-\alpha$	$\overline{p}$	0.625	=C4/C3
6		1 - α =	0.99	
7		α =	0.01	=1-C6
8		Z =	2.575829304	=DISTR.NORM.ESTAND.INV(1-C7/2)
9				
10		Intervalo	0.0623509	=C8*RAIZ((C5*(1-C5)/C3))
11				
12		Lim Inf	0.5626491	=C5-C10
13		Lim Sup	0.6873509	=C5+C10
14				
15			P[0.5626 < = P < = 0.6874] = 0.99	
16			=CONCATENAR("P[";REDONDEAR(C12;4);" < = P < = ";REDONDEAR(C13;4);"] = ";C6)	
17				
18			Entre el 56.2649% y 68.7351% esta la proporción de la población con un nivel de confianza del 99%	
19			=CONCATENAR("Entre el ";REDONDEAR(C12*100;4);"% y ";REDONDEAR(C13*100;4);"% esta la proporción de la pobal	

6.5 Estimación para la proporción poblacional. Muestras pequeñas

Todo el análisis para la estimación de la proporción poblacional para muestras grandes se aplica para determinar la proporción poblacional par muestras pequeñas.

$$\overline{p}-t*\sqrt{\frac{\overline{p}*(1-\overline{p})}{n}}\leq\mu\leq\overline{p}+t*\sqrt{\frac{\overline{p}*(1-\overline{p})}{n}} \quad ; \quad P\left[\overline{p}-t*\sqrt{\frac{\overline{p}*(1-\overline{p})}{n}}\leq\mu\leq\overline{p}+t*\sqrt{\frac{\overline{p}*(1-\overline{p})}{n}}\right]=1-\alpha$$

Ejemplo. Un periodista estaba preguntando en la calle sobre la opinión que tienen los transeúntes sobre la pena de muerte. De las 25 persona a las que se les preguntó, 20 estuvieron de acuerdo. Encontrar el intervalo de confianza con un nivel de confianza del 95%.

n = 425; X = 20; p = 20/25 = 0.80; q = 1-0.8 = 0.20; 1- α = 0.95; $t_{24;0.95}$ = 2.0639

$$P\left[0.80-2.0639*\sqrt{\frac{0.8*0.2}{25}}\leq\mu\leq 0.80+2.0639*\sqrt{\frac{0.8*0.2}{25}}\right]=0.95$$

$$P\left[0.634888\leq\mu\leq .965112\right]=0.95$$

Según los transeúntes, entre el 63.49% y el 96.51% de los transeúntes están de acuerdo con la pena de muerte.

C27

	A	B	C	D
1		Ejercicio 4		
2	$P\left[\overline{p}-t*\sqrt{\frac{\overline{p}*(1-\overline{p})}{n}}\leq\mu\leq\overline{p}+t*\sqrt{\frac{\overline{p}*(1-\overline{p})}{n}}\right]=1-\alpha$			
3		n	25	
4		x	20	
5		$\overline{p}$	0.8	=C4/C3
6		1 - α =	0.95	
7		α =	0.05	=1-C6
8		t =	2.063898547	=DISTR.T.INV(1-C6;C3-1)
9				
10		Intervalo	0.165111884	=C8*RAIZ((C5*(1-C5)/C3))
11				
12		Lim Inf	0.634888116	=C5-C10
13		Lim Sup	0.965111884	=C5+C10
14				
15			P[0.6349 < = P < = 0.9651] = 0.95	
16			=CONCATENAR("P[";REDONDEAR(C12;4);" < = P < = ";REDONDEAR(C13;4);"] = ";C6)	
17				
18			Entre el 63.4888% y 96.5112% esta la proporción de la población con un nivel de confianza del 95%	
19			=CONCATENAR("Entre el ";REDONDEAR(C12*100;4);"% y ";REDONDEAR(C13*100;4);"% esta la proporción de la pobalción cc	

6.7 Análisis de sensibilidad en Excel para la estimación del intervalo.

UNIDAD 7. PRUEBA DE HIPÓTESIS

7.1 Definición de Hipótesis Nula y Alterna

La estadística inferencial es el proceso de usar la información de una muestra para describir el estado de una población. Sin embargo es frecuente que usemos la información de una muestra para probar un reclamo o conjetura sobre la población. El reclamo o conjetura se refiere a una hipótesis. El proceso que corrobora si la información de una muestra sostiene o refuta el reclamo se llama prueba de hipótesis.

En la prueba de hipótesis se pone a prueba un reclamo hecho sobra la naturaleza de una población a base de la información de una muestra. El reclamo se llama hipótesis estadística.

Hipótesis es una aseveración acerca de una población.

Hipótesis Estadística: Una hipótesis estadística es una afirmación acerca de un parámetro poblacional, hecho con el propósito de ponerlo a prueba.

Por ejemplo, la premisa formulada por un productor de baterías para autos de que su batería dura en promedio 48 meses, es una hipótesis estadística porque el manufacturero no inspecciona la vida de cada batería que él produce. Si surgieran quejas de parte de los clientes, entonces se pone a prueba el reclamo del manufacturero.

El procedimiento para mostrar si la afirmación es válida o no, se conoce como prueba de hipótesis. Ese procedimiento esta basado sobre los resultados de un muestra realizada para tal fin.

La hipótesis estadística sometida a prueba se llama la hipótesis nula, y se denota como H_0.

La hipótesis Nula (H_0) es la afirmación o conjetura que se hace sobre el parámetro poblacional.

En caso que no se logre tomar como válida la hipótesis nula se debe tener una afirmación alterna que se debe dar como válidas y conocida cono la hipótesis alterna que se denota como H_1

Hipótesis Alterna: Una premisa que es cierta cuando la hipótesis nula es falsa.
Por ejemplo, para el productor de baterías

Por ejemplo, para probar o desaprobar el reclamo pronunciado por el productor de baterías debemos probar la hipótesis estadística de que $\mu \geq 48$. Por lo tanto, la hipótesis nula es:

$H_0 : \geq 48$ $\qquad$ $H_1 : < 48$

Luego de tener H_0 y H_1 se procede a tomar una muestra aleatoria de baterías y medir su vida media.

Para probar si la hipótesis nula es cierta, se toma una muestra aleatoria y se calcula la información, como el promedio, la proporción, etc. Esta información muestral se llama **estadística de prueba.**

7.2 Tipos de Errores

A base de la información de una muestra nosotros podemos cometer dos tipos de errores en nuestra decisión.

- Rechazar $\mathbf{H_0}$ siendo que es cierta. Error tipo 1
- Aceptar $\mathbf{H_0}$ siendo que es falsa. Error tipo 2

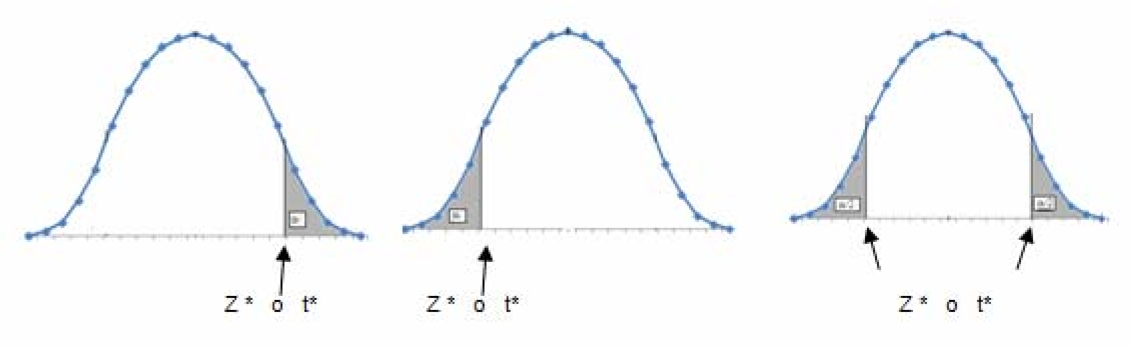
El error Tipo 1 se da **c**uando se rechaza la Hipótesis Nula siendo que es válida.
El error Tipo 2 se da cuando aceptamos la Hipótesis Nula siendo que es falsa.

7.3 Nivel de Significancia (α)

Para ser muy cuidadosos en no cometer el error tipo 1, debemos especificar la probabilidad de rechazar H_0, denotada por α.
El nivel de significancia es la probabilidad de cometer el error tipo I. Este valor debe ser pequeño.

Los valores más usados para hacer prueba de hipótesis es 1%, 5% y 10%.

Usando un valor preasignado de α se construye una región de rechazo o región crítica en la curva normal estándar o en la curva *t* que indica si debemos rechazar H_0 o aceptarla.



Región Crítica o de Rechazo. Una región crítica o de rechazo es una parte de la curva de *z* o de la curva *t* donde se rechaza H_0. La región puede ser de una cola o de dos dependiendo de la Hipótesis Alterna. Si la hipótesis alterna es H1: $\mu > k$, la cola es hacia la derecha, si H1: $\mu < k$, la cola es hacia la derecha, o si H1: $\mu = k$, son dos coas.

7.4 Prueba de Hipótesis Unilaterales y Bilaterales sobre la Media

Si queremos decidir entre dos hipótesis que afectan a un cierto parámetro de la población, a partir de la información de la muestra usaremos el contraste de hipótesis, cuando optemos por una de estas dos hipótesis, hemos de conocer una medida del error cometido, es decir, cuantas veces de cada cien nos equivocamos.
En primer lugar, veremos cómo se escribirían las hipótesis que queremos contrastar:
H_0 se llama hipótesis nula y es lo contrario de lo que sospechamos que va a ocurrir (suele llevar los signos igual, mayor o igual y menor o igual)
H_1 se llama hipótesis alternativa y es lo que sospechamos que va a ser cierto (suele llevar los signos distinto, mayor y menor). Los contrastes de hipótesis pueden ser de dos tipos:

Bilateral: En la hipótesis alternativa aparece el signo =.

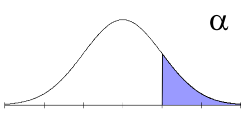
Unilateral: En la hipótesis alternativa aparece o el signo > o el signo <.

Nivel de significancia α. Es la probabilidad de cometer un error de tipo I. β es la probabilidad de cometer un error de tipo II. De los dos, el más importante es α

Debido a que los dos errores anteriores a la vez son imposibles de controlar, vamos a fijarnos solamente en el nivel de significación, este es el que nos interesa ya que la hipótesis alternativa que estamos interesados en probar y no queremos aceptarla si en realidad no es cierta, es decir, si aceptamos la hipótesis alternativa queremos equivocarnos con un margen de error muy pequeño.

El nivel de significación lo marcamos nosotros. Si es grande es más fácil aceptar la hipótesis alternativa cuando en realidad es falsa. El valor del nivel de significación suele ser un 5%, lo que significa que 5 de cada 100 veces aceptamos la hipótesis alternativa cuando la cierta es la nula.

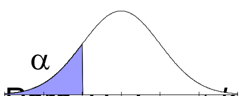
Ejemplos Para H_1: $\mu >$ valor aceptado, la región de rechazo está dada por:



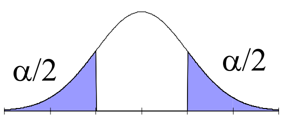
(cola derecha, z* ó t*)

Para H_1 : μ < valor aceptado, la región de rechazo está dada por

(cola izquierda, z* ó t*)



Para H_1 : $\mu \neq k$ valor aceptado, la región de rechazo es de dos colas y está dada por



(2-colas, Z* ó t*)

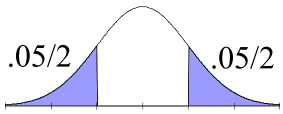
Ejemplo 1: Determine si la región de rechazo es de la cola derecha, de la cola izquierda o de dos colas.

a. $H_0 : \mu = 15, \quad H_1 : \mu \neq 15$, α=.05

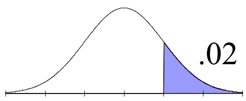
b. $H_0 : p \leq 0.7, \quad H_1 : p > 0.7$, α=.02

Solución: La forma de la región de rechazo está determinada por la hipótesis alterna.

a. $H_1 : \mu \neq 15$ significa que la región está en ambas colas.



b. $H_1 : p > 7$ significa que la región está en la cola derecha.

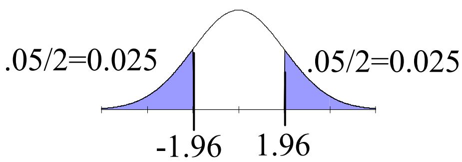


Ejemplo 2: En el Ejemplo 1a, presumamos que la región de rechazo es parte de la curva normal estándar. Complete el dibujo de la región crítica para los valores α siguientes:

a. α = .05

Solución:

a. Del ejemplo 1(*a*), tenemos:



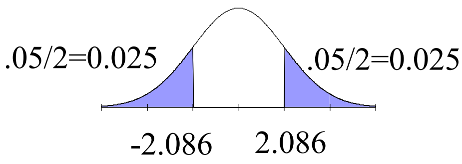
De la tabla de la distribución normal, la P(Z<z) =.025 corresponde a un valor Z= -1.96. Por simetría la

Ejemplo 3: En el ejemplo 1a, presumamos que la región de rechazo es parte de la curva *t*. Complete el dibujo de la región de rechazo para:

a. α = .05 y υ = 14

Solución:

a. Del ejemplo 1(a), α = .05, y υ = 14, tenemos:



De la tabla de la distribución t, la P(T<t) =.025 corresponde a un valor t= -2.086. Por simetría la P(T>t)=.025 corresponde a t=

Ejemplo 4: Establezca las hipótesis nula y alterna.

a. Las millas por galón (mpg) promedio de un nuevo modelo de automóvil es 32.

b. Más del 65% de los empleados de un colegio aportan a Fondos Unidos.

c. En promedio, los empleados de cierta compañía viven a no más de 15 millas de la misma.

d. Al menos un 60% de la población adulta de una comunidad votará en las próximas elecciones Presidenciales.

e. El peso promedio de un pollo para asar es de al menos cuatro libras.

Solución:

a. $H_0 : \mu = 32$, $H_1 : \mu \neq 32$ **b.** $H_0 : p \geq .65$, $H_1 : p < .65$ **c.** $H_0 : \mu \leq 15$, $H_1 : \mu > 15$

d. $H_0 : p \geq .6$, $H_1 : p < .6$ **e.** $H_0 : \mu \geq 4$, $H_1 : \mu < 4$

EJERCICIOS

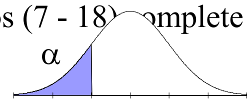
En los ejercicios (1-6) determine si la región de rechazo para la hipótesis nula está en la cola izquierda, en la cola derecha, o ambas colas. Para el nivel de significancia α dibuje la región de rechazo.

1. $H_0 : \mu \leq 11$; $H_1 : \mu > 11$ 2. $H_0 : \mu \geq 5.8$; $H_1 : \mu < 5.8$

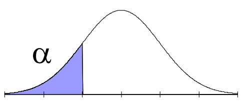
3. $H_0 : p = 0.4$; $H_1 : p \neq 0.4$ 4. $H_0 : \mu = 110$; $H_1 : \mu \neq 110$

5. $H_0 : p \geq 0.3$; $H_1 : p < 0.3$ 6. $H_0 : p \geq 0.8$; $H_1 : p < 0.8$

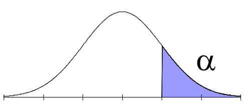
En los ejercicios (7 - 18) complete la región de rechazo (encuentre el valor de z y t).

7. 

a) z, si $\alpha = .05$ b) *t*, si $\alpha = .025$ y $\upsilon = 9$

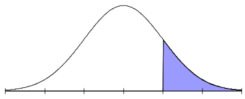
8. 

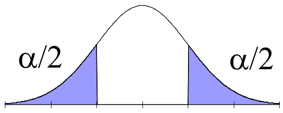
a) z, si $\alpha = .01$ b) *t*, si $\alpha = .05$ y $\upsilon = 13$

9. 

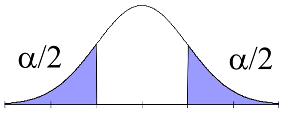
a) z, si $\alpha = .02$ b) *t*, si $\alpha = .01$ y $\upsilon = 5$

α

10.  a) z, si $\alpha = .025$ b) t, si $\alpha = .01$ y $\upsilon = 9$

11. 

a) z, si $\alpha = .05$ b) t, si α=.05 y $\upsilon = 10$

12. 

a) z, si $\alpha = .01$ b) t, si α=0.1 y $\upsilon = 7$

En los ejercicios (13 - 18) establezca las hipótesis nula y alterna.

13. Los automóviles estacionados en el estacionamiento de periodo prolongado del aeropuerto internacional de Baltimore permanecen un promedio de 2.5 días.

14. Una nueva marca de llantas radiales dura en promedio más de 48,000 millas.

15. El balance promedio de una cuenta de cheques en el First State Bank es de al menos $150.

16. Se reclama que al menos el 60% de las compras realizadas en cierta tienda por departamentos son artículos de especiales.

17. Se reclama que el 20% de los graduados de cierto colegio privado solicitan admisión a escuelas de medicina.

18. Un dentista reclama que el 5% de sus pacientes sufren enfermedades en las encías.

7.4.1 Muestras Grandes

En las pruebas de hipótesis para la media (μ), cuando se conoce la desviación estándar (σ) poblacional, o cuando el valor de la muestra es grande (30 o más), el valor estadístico de prueba es z y se determina a partir de:

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

El valor estadístico z, para muestra grande y desviación estándar poblacional desconocida se determina por la ecuación:

$$Z = \frac{\bar{x} - \mu}{S / \sqrt{n}}$$

7.4.2 Muestras Pequeñas

En la prueba para una media poblacional con muestra pequeña y desviación estándar poblacional desconocida se utiliza el valor estadístico t.

$$t = \frac{\bar{x} - \mu}{S / \sqrt{n}}$$

Procedimiento para una prueba de hipótesis

Los pasos a seguir son:

1. Formular la hipótesis nula H_0 y la alternativa H_1, de acuerdo al problema.
2. Escoger un nivel de significación o riesgos α.
3. Elegir la estadística de prueba apropiada, cuya distribución por muestreo sea conocida en el supuesto de que H_o es cierta.
4. En base a α y H_1, determinar el valor (o los valores) críticos y con ello se establecen las regiones de aceptación o rechazo.
5. Calcular los valores de la prueba estadística a partir de una muestra aleatoria de tamaño n, Ho y reemplazarlos en la estadística de prueba elegida en el paso 3, para hallar el valor experimental.
6. Tomar la decisión de aceptar H_o si el valor experimental cae en la región de aceptación y rechazarla si dicho valor cae en la región crítica o de rechazo.
7. Opcional: Si se rechaza H_0, se puede hallar un intervalo de confianza para el parámetro de interés.

Prueba de hipótesis sobre la media poblacional

Caso A: Cuando la varianza poblacional es conocida.

Deseamos contrastar la hipótesis de que el parámetro poblacional μ = $\overline{X}$ toma un determinado valor K Conocemos que la población se distribuye normalmente y conocemos también su varianza , o bien si nos es desconocida, el tamaño muestral es lo suficientemente grande cómo para poder utilizar la muestral cómo poblacional.

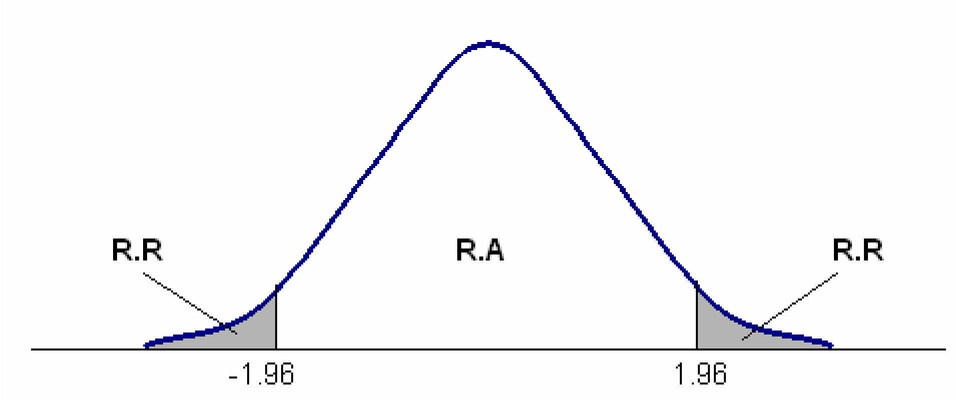
Hemos determinado un nivel de significación para la realización del contraste y vamos a plantearlo en el supuesto de realizar una muestra aleatoria de tamaño n.

Así: conocemos que $\bar{x} \Rightarrow N\left[u, \sigma/\sqrt{n}\right]$ de lo que deducimos que $\frac{\bar{x}-u}{\sigma/\sqrt{n}} \Rightarrow N[0,1]$ de forma que la hipótesis nula es: H_0: $\mu = K$.

El estadístico está dado por: $Z = \frac{\bar{x}-u}{\sigma/\sqrt{n}}$.

Ejemplo Nº 1. De 100 observaciones de una población normal se obtiene que $\bar{x}$ = 5 y que σ=2.Contrastar con un nivel de significación del 5% la hipótesis de que la media de la población sea 7.

Aplicando el procedimiento para probar una hipótesis tenemos:

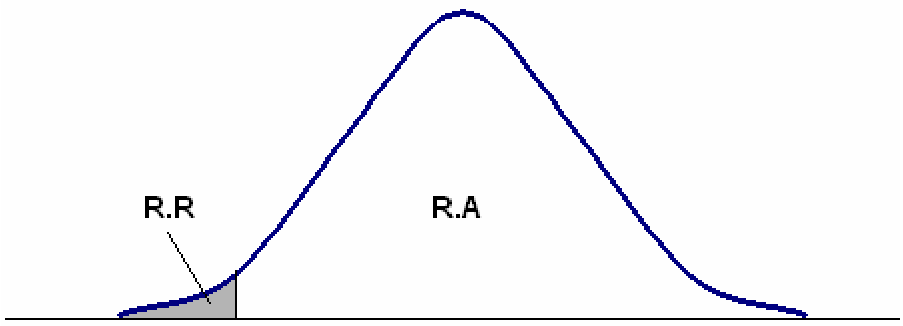
1. H_0: μ=7
   H_1: $\mu \neq 7$
2. El nivel de significancia es del 5%. (α=5%)
3. $Z = \frac{\bar{x}-u}{\sigma/\sqrt{n}}$
4. Establecemos la región de aceptación y de rechazo:
5. 

6. Realizamos la prueba estadística: $Z = \frac{5-7}{2/\sqrt{100}} = -10$
7. Dado que Z=-10 y no pertenece a la región de aceptación estamos en condiciones de rechazar la hipótesis nula, luego aceptar la alternativa : $\mu_0 \neq 7$.

Ejemplo Nº 2. Un empresario está considerando la posibilidad de ampliar su negocio mediante la adquisición de un pequeño bar. El dueño actual del bar afirma que el ingreso diario del establecimiento sigue una distribución normal de media $675 y una desviación estándar de $75 s. Para comprobar si decía la verdad, tomó una muestra de treinta días y ésta reveló un ingreso diario promedio de $625. Utilizando un nivel de significación del 10

%. ¿Hay evidencia de que el ingreso diario promedio sea menor del que afirma el presente dueño?.

Aplicando el procedimiento para probar una hipótesis tenemos:

1. H_0: $\mu \geq 675$
   H_1: $\mu < 675$
2. El nivel de significancia es del 10%. (α=10%)
3. $Z = \dfrac{\bar{x} - u}{\sigma / \sqrt{n}}$
4. Establecemos la región de aceptación y de rechazo:
5. 

6. Realizamos la prueba estadística: $Z = \dfrac{625 - 675}{75 / \sqrt{30}} = -3.65$
7. Dado que Z=-3.65 y no pertenece a la región de aceptación estamos en condiciones de rechazar la hipótesis nula, luego aceptar la alternativa: $\mu < 7$.

Caso B: Cuando no se conoce la varianza poblacional y para una muestra pequeña.

Deseamos contrastar la hipótesis de que el parámetro poblacional μ toma un determinado valor K. Desconocemos la varianza de la población y, dado que el tamaño muestral es pequeño (menor o igual a 30), podemos utilizar varianza en su lugar.

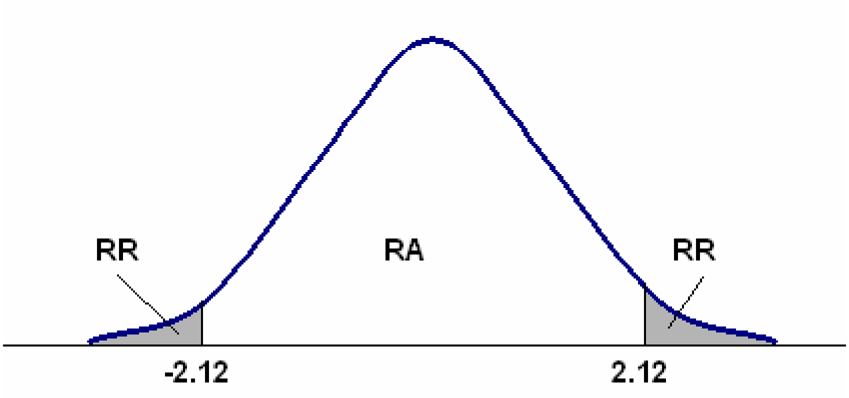
Hemos determinado un nivel de significación para la realización del contraste y vamos a plantearlo en el supuesto de realizar una muestra aleatoria de tamaño n menor o igual a 30.

En este caso el estadístico de prueba será $t = \dfrac{\bar{X} - \mu}{s / \sqrt{n}}$

Ejemplo . Se escoge a 17 individuos al azar y se les mide, resultando que su estatura media es de 1,71 metros con desviación típica de 0,02 .Contrastar la hipótesis de que la estatura media nacional sea de 1.75 metros si utilizamos un nivel del significación del 5%. Se supone normalidad

Aplicando el procedimiento para probar una hipótesis tenemos:

1. H_0: $\mu = 1.75$
   H_1: $\mu \neq 1.75$
2. El nivel de significancia es del 5%. (α=5%).
3. Como n es meor de 30 se busca en la tabla para un nivel de caso t* = +/- 2.11991.



4. Establecemos la región de aceptación y de rechazo Utilizando la tabla T.
5. Se calcula el estadístico de prueba. $t = \frac{1.71 - 1.75}{0.02/\sqrt{17}} = -8.25$
6. Dado que t=-8.25 y no pertenece a la región de aceptación estamos en condiciones de rechazar la hipótesis nula, luego aceptar la alternativa: $\mu \neq 1.75$.

7.5 Prueba de Hipótesis sobre la Proporción de una Población p

Se trata de efectuar una prueba de hipótesis acerca de la proporción de elementos con cierto atributo en una población, hipótesis de la forma:

H_0: $P = p_0$.
H_1: $p \neq p_0$.

H_0: $p \leq p_0$.
H_1: $p > p_0$.

H_0: $p \geq p_0$.
H_1: $p < p_0$.

Cuando el tamaño de la muestra es mayor o igual a 30 el estadístico de prueba es Z.

$$Z = \frac{P - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Donde $P = \frac{x}{n}$ (proporción muestral)

Ejemplo 1. Una empresa de publicidad desea comprobar si un determinado programa de televisión es visto por el 30% de la audiencia potencial .Para ello se escoge al azar una muestra de 200 familias resultando que de ellas 50 lo ven asiduamente. Contrastar la hipótesis con un nivel de significación del 5%.

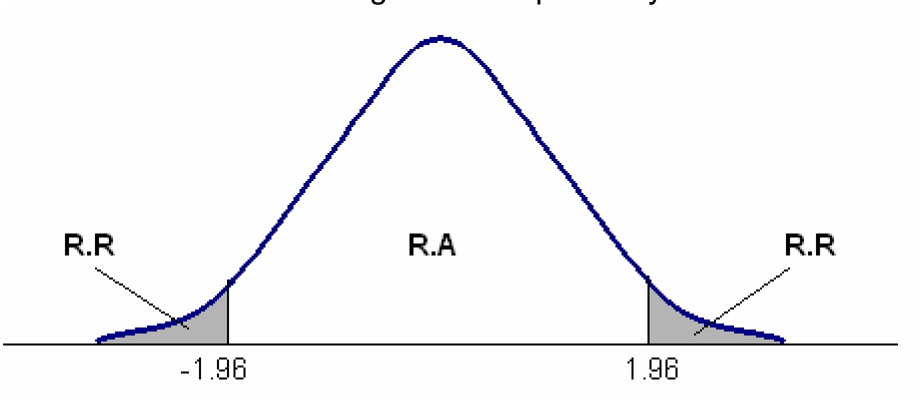
Aplicando el procedimiento para probar una hipótesis tenemos

1. H_0: p = 0.3
   H_1: p $\neq 0.30$

2. El nivel de significancia es del 5%. (α=5%).

3. Z* = +/- 1.96

4. Establecemos la región de aceptación y de rechazo



5. Realizamos la prueba estadística:

$$P = \frac{50}{200} = 0.25$$

$$Z = \frac{P - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.25 - 0.30}{\sqrt{\frac{0.3(1-0.3)}{200}}} = -1.54$$

6. Dado que Z=-1.54 y pertenece a la región de aceptación estamos en condiciones de acepta la hipótesis nula, es decir: p=0,3

Ejemplo. Un fabricante de refrescos sin burbujas desea sacar al mercado una variedad de su producto que tenga burbujas. Su director comercial opina que al menos el 50 % de los consumidores verá con buenos ojos la innovación. Se realiza un sondeo de mercado y resulta que de 100 consumidores encuestados 40 son favorables a la innovación.

a) Contrastar la hipótesis del director comercial frente a la alternativa de que el % de aceptación es inferior, con un nivel de significación del 1%.

b) Si el aceptable la hipótesis de que el % de aceptación del nuevo producto es inferior o igual al 30 % el fabricante decidirá no fabricarlo. Si es aceptable el criterio del director comercial entonces sí fabricarán el refresco con burbujas. Y si ninguna de las 2 hipótesis es aceptable procederán a hacer otro sondeo. Para tomar esta decisión trabajarán con un nivel de significación del 5 %. ¿ Por qué optarán?.

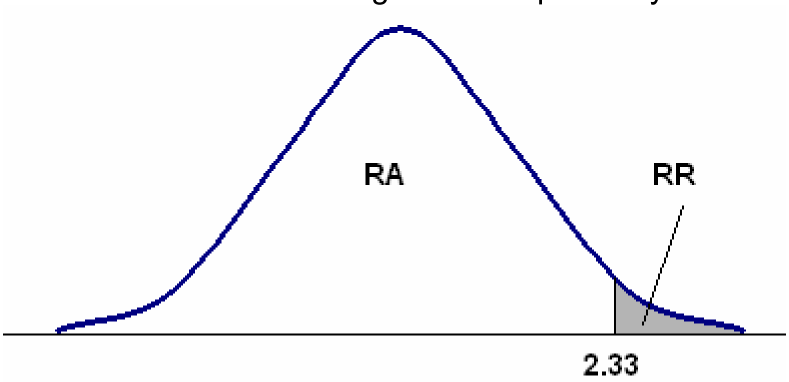
Para el punto a)
Aplicando el procedimiento para probar una hipótesis tenemos:

1. H_0: p ≤ 0.5
   H_1: p > 0.5

2. El nivel de significancia es del 1%. (α=1%).

3. Z = 2.33

4. Establecemos la región de aceptación y de rechazo:



5. Realizamos la prueba estadística: $P = \frac{40}{100} = 0.4$

$$Z = \frac{P - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.4 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{100}}} = -2$$

6. Dado que Z=-2 y pertenece a la región de aceptación estamos en condiciones de aceptar la hipótesis nula, es decir: p$\leq 0{,}5$.

Para el punto b)

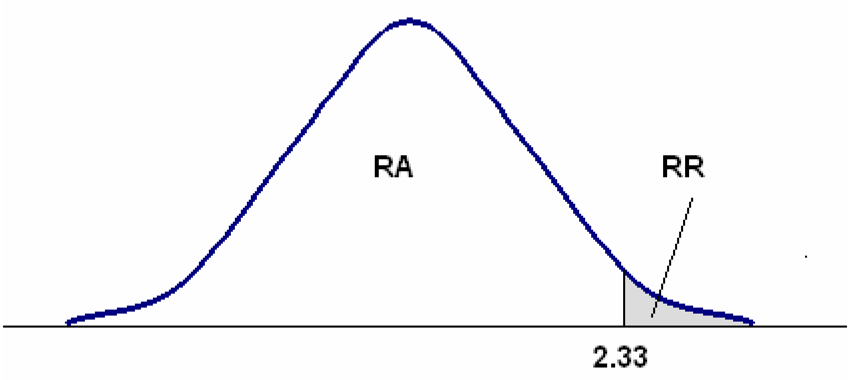
Aplicando el procedimiento para probar una hipótesis tenemos:

1. H_0: p $\leq$ 0.3
   H_1: p > 0.3

2. El nivel de significancia es del 1%. (α=1%).

3. Z = 2.33

4. Establecemos la región de aceptación y de rechazo:



5. Realizamos la prueba estadística:

$$P = \frac{40}{100} = 0.4$$

$$Z = \frac{P - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.4 - 0.3}{\sqrt{\frac{0.3(1-0.3)}{100}}} = 2.18$$

6. Dado que Z=2.18 y pertenece a la región de aceptación estamos en condiciones de aceptar la hipótesis nula, es decir: p$\leq 0,3$. Por lo tanto se recomiendo no fabricar el refresco.

Cálculo de la Probabilidad para el error tipo II

Analizaremos en forma completa los errores Tipo **I** y Tipo **II**, con respecto a las prueba de una media hipotética. Sin embargo los conceptos que se ilustran aquí son aplicables también a otros modelos de pruebas de hipótesis.

La probabilidad del error Tipo **I** es siempre igual al nivel de significancia que se utiliza al probar las hipótesis nulas. Dicho de otra manera, hay todavía una probabilidad (valor de α) de que se pudiera elegir una muestra al azar que diera una media comprendida dentro de la región de rechazo ó las regiones de rechazo. Ello haría que al rechazar la hipótesis nula siendo verdadera cometiéramos el error Tipo **I**.

La única forma en que se puede determinar la probabilidad del error Tipo II (β) es con respecto a un valor específico incluido dentro del rango de la hipótesis alternativa.

News & World Report publicó un artículo sobre la carrera de éxitos de Wal-Mart. Actualmente es la mayor cadena de ventas al por menor de la nación. Empezó con una sola tienda de descuento en la pequeña localidad de Rogers, Arkansas, y a crecido hasta poseer 1300 tiendas en 25 estados. Este éxito le ha valido a Sam Walton, fundador y mayor accionista, el título del hombre más rico de América. Las ventas anuales se cifran en 15 millones de dólares por tienda.

a) Si se elige al azar una muestra de 120 tiendas y se hallan unas ventas medias de 15.39 millones de dólares, con una desviación estándar de 2.9 millones de dólares. Pruebe la hipótesis $\mu =$ 15 millones con un nivel de significancia del 10%
b) Si la μ es en realidad 14.8 millones de dólares, ¿Cuál es la probabilidad de cometer el error Tipo II.

Datos
$n = 120$
$\bar{X} = 15.39$ millones
$\sigma = 2.9$ millones
$\alpha = .10$

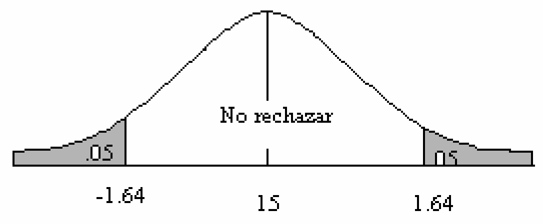
1.Establecer la hipótesis

Ho: $\mu = 15$
H1: $\mu \neq 15$

2. Establecer la estadística de prueba

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

2. Definir el nivel de significancia y la zona de rechazo



3.Nivel de significancia $\beta = 0.10$
Zona de rechazo = { Z/ Z < -1.64 o Z/Z > 1.64 }

4. Calcular la estadística de prueba $Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ el valor de la media poblacional es el que esta a prueba en la hipótesis por lo tanto $Z = \frac{15.39 - 15}{2.9 / \sqrt{120}} = 1.47$

Como $1.47 < 1.64$, No se rechaza Ho

5. Conclusión. Existe evidencia para decir que las ganancias anuales por tienda son de 15 millones de dólares por tienda con un nivel de significancia de 0.10.

Pasos para calcular el error Tipo II

1. Plantear la hipótesis nula y alternativa para la prueba

   Ho: $\mu = 15$

   H_1: $\mu \neq 15$

2. Determinar el valor crítico de la media muestral que debe utilizarse para probar la hipótesis nula con un nivel de significancia dado.

$$\overline{X}_c = \mu \pm Z \ \sigma / \sqrt{n}$$

$$\overline{X} = 15 \pm (1.64)\, 2.9 / \sqrt{120} = 14.5659$$

$$\overline{X} = 15 \pm (1.64)\, 2.9 / \sqrt{120} = 15.4341$$

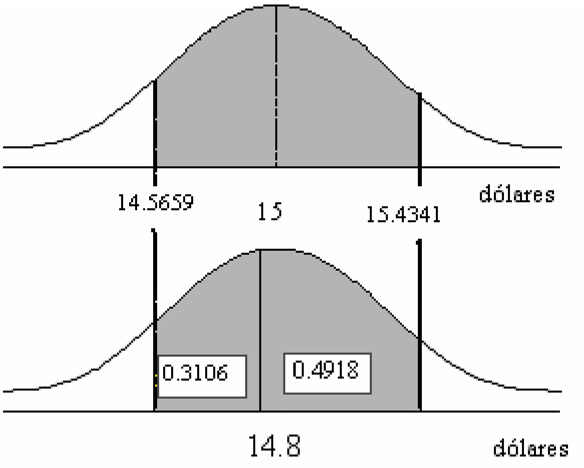
Identificar la probabilidad del error Tipo I correspondiente al valor crítico que se calculó antes, como base para la regla de decisión. $\alpha = 0.10$

Identificar la probabilidad del error Tipo II correspondiente a la regla de decisión, dada una media alternativa especificada.

P (error Tipo II) = P ($14.56 < \overline{X} < 15.44$)

$$Z = \frac{15.4341 - 14.8}{2.9 / \sqrt{120}} = 2.40 \quad ; \quad Z = \frac{15.5659 - 14.8}{2.9 / \sqrt{10}} = 0.88$$

P (error Tipo II) = P ($-0.88 < Z < 2.40$) = 0..8024


14.5659
15
15.4341
dólares
0.3106
0.4918
14.8
dólares

UNIDAD 8. REGRESIÓN Y CORRELACIÓN SIMPLES.

En la Estadística descriptiva se analizó el comportamiento de una variable aleatoria. En cada variable se hace un análisis y luego se buscó hacer un estimado a cerca de un parámetro poblacional utilizando las técnicas de Estimación por intervalo o un aprueba de hipótesis utilizando las distribuciones de probabilidad.

Ahora se buscará analizar dos variables que de alguna forma podrían tener un relación donde el resultado de una esta ligado al resultado de la otra variable.

El espacio muestral de un experimento con dos variables consiste en cierto conjunto de pares ordenados de medidas. Es decir, se hacen dos observaciones en cada prueba. Por ejemplo, en un almacén se puede observar los descuentos y el volumen de ventas. Otro ejemplo podría ser que a cada estudiante se observara su estatura y el peso. La tabulación de los datos podría tener las siguientes tablas de resultados.

Ejemplo 1:

Variable 1	Descuentos (%)	1,1	2,5	3,1	4,2	2,0	2,5	2,8	2,7	1,8	3,8
Variable 2	Ventas (millones $)	13,64	24,05	26,78	30,63	21,22	24,05	25,49	25,03	19,88	29,36

Ejemplo 2:

Variable 1	Estatura (Cms)	167	185	168	161	178	166	178	160	158	172
Variable 2	Peso (kgs)	64	92	63	66	85	65	80	56	50	67

El problema principal de la inferencia estadística en una distribución de dos variables es el de determinar la verdadera relación entre X y Y; es decir, cómo se comportan las dos variables, una con respecto a la otra. Los almacenes con alto descuento tendrían mayores ventas. Estudiantes altos deberían tener mayores pesos corporales.

El problema que se plantea es como se establece un tipo de relación en forma de ecuación, de tal manera que basados en el valor de una X, se pueda hallar una estimación de la otra variable.

Una ecuación de este tipo se conoce como una relación de estimación. El procedimiento de estimación es también una técnica de predicción, lo cual es función fundamental de la ciencia, natural o social. En las ciencias naturales, la predicción se hace posible porque existe la relación de causa y efecto entre dos o más variables.

Si existe una relación funcional entre dos variables nos lleva a un tema importante de la estadística, el análisis de regresión.

Si son dos variables las que se relacionan, la variable independiente se designa por X; en tanto que la variable cuyo valor se va a estimar se llama la variable dependiente y se designa por Y. Cuando se formula una ecuación para estimar Y a partir de X, ésta se denomina una regresión de Y respecto de X.

La técnica del análisis de la regresión no es otra cosa que un procedimiento de estimación o predicción. El término "regresión" es un término estadístico. Fue introducido por primera vez en 1877 por Sir Francis Galton, quien encontró en sus estudios sobre la herencia que los padres altos tendían a engendrar hijos altos y que los padres bajos mostraban tendencia a tener hijos bajos. Pero con todo, la estatura promedio de los hijos de padres muy altos era menor que la estatura promedio de sus padres, en tanto que los hijos de padres muy bajos, en promedio, eran más altos que sus padres. Galton llamó esta tendencia hacia la estatura promedio de todos los hombres con el nombre de regresión. Asociaciones parecidas en otros varios fenómenos fueron observadas por Galton, quien entonces las generalizó como una ley universal de comportamiento entre dos o más variables asociadas.

El análisis de la regresión se clasifica generalmente en dos tipos, simple y múltiple. La regresión simple es aquella en que entran solamente dos variables, tales como la regresión de Y respecto a X antes mencionada. La regresión múltiple es aquella en la que intervienen tres o más variables, una de las cuales es la variable dependiente, la que se va a asociar con los valores de todas las demás.

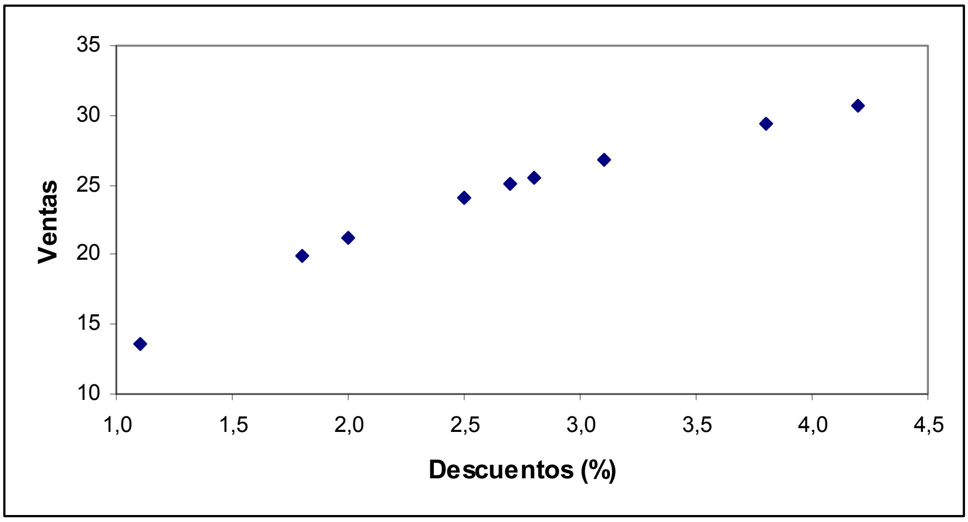
Por demás, el estudio se restringirá a la regresión simple solamente o sea aquella en que la ecuación que describe la relación entre X y Y es lineal y se representa gráficamente por una recta. A veces ocurre que una ecuación de regresión que describe de la mejor manera posible la relación entre variables resulta curvilínea; su representación geométrica es entonces una curva en vez de una recta.

Cuando se encuentra que unas variables están relacionadas entre sí, suele ser útil averiguar cuan estrecha es la relación. El grado de relación entre éstas se denomina también correlación entre las variables. El problema de correlación está íntimamente asociado al de la regresión y es parte integrante del análisis de dos variables.

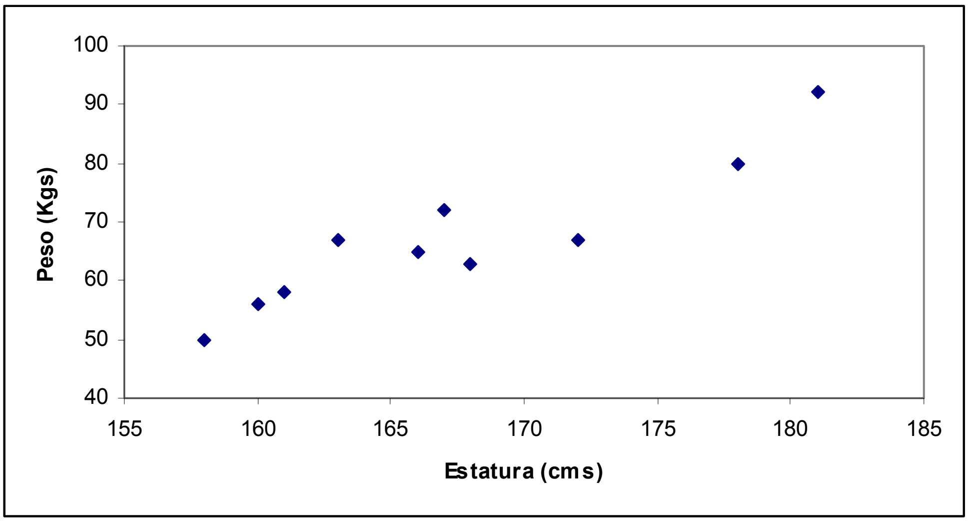
El análisis de correlación es el conjunto de técnicas estadísticas empleados para medir la intensidad de la asociación entre dos variables. El objetivo es determinar que tan intensa es la relación entre las dos variables.

Diagramas de dispersión. Es la gráfica que representa la relación entre las dos variables.

La siguiente gráfica muestra en una como están relacionados la variable descuento con el volumen de ventas. Es claro que entre mayor descuento se ofrezca a los clientes mayor será el volumen de venta. Hay una tendencia creciente.



La siguiente gráfica muestra en una como están relacionados la variable estatura con el volumen de peso. Se observa que a mayor estatura mayor pesos. Hay una tendencia creciente.

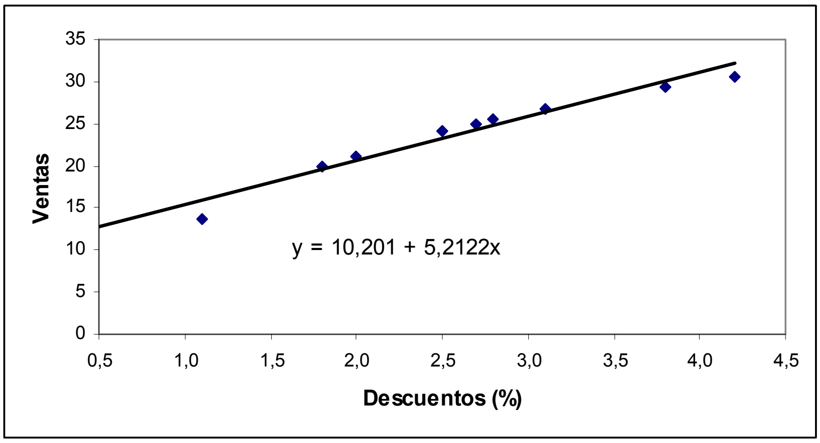


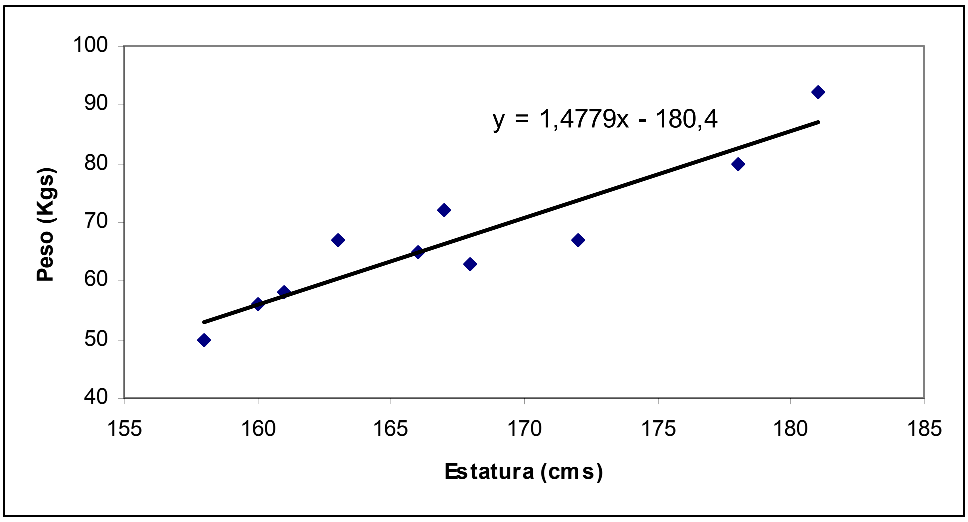
Para hacer una estimación se parte de la propuesta de un modelo. En el análisis de dos variables interesa un modelo particular, la ***recta de regresión*** de la población. Ésta, la cual se refiere a la población, no puede ser conocida y por tanto, debe ser estimada con base en los datos muéstrales y se obtiene la recta de regresión estimada. Como es de esperarse, para dar validez a las conclusiones acerca de la citada recta, hay que fijar ciertos supuestos. Entre los de mayor relevancia están los siguientes:

1. Los valores de la variable dependiente X y Y se toman previamente. Se consideran variables determinísticas. si su valor está prefijado de antemano en el experimento.

2. Se supone que la relación entre X y Y. está dada por la ecuación: $Y_i = \alpha + \beta X_i + \epsilon_i$, donde α, β son los parámetros poblacionales y ϵ_i es el error por la diferencia entre Y_i y el valor esperado de Y como variable aleatoria determinada por el valor de X en particular. La variable dependiente es la que se predice o calcula mientras que la independiente es la base del cálculo.

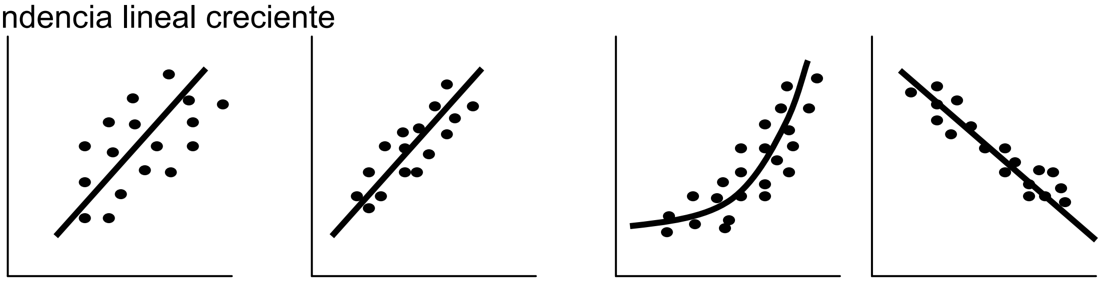
α es la intersección de la recta con el eje Y y β es la pendiente de la recta de regresión. Cuando β es negativo, la recta es decreciente y si β es positivo, la recta de regresión es creciente. Otros autores se refieren al modelo de regresión lineal con la ecuación Y = A + B*X





La representación de los datos muestra lo que se conoce como nube de puntos y el ubicar una línea describe con la técnica de mano alzada el modelo de regresión, lineal o no lineal. Al analizar la gráfica de una serie de datos se puede observar que la tendencia tiene una de los siguientes formas.

1. Tendencia lineal creciente



1 2 3 4

2. Tendencia lineal creciente
3. Tendencia no-lineal creciente
4. Tendencia lineal decreciente.

8.2 Ecuación del modelo de regresión lineal simple

$\hat{Y} = \alpha + \beta * \hat{X}; \quad donde\, \hat{Y}\ se\, conoce\, como\, Y\, estimado\ y\ \hat{X}, se\, conoce\, como\ estimado$

$Y = A + B * \hat{X}$

La técnica para estimar los valores de α y β se conoce como análisis de regresión lineal. La ecuación de regresión define la relación entre dos variables.

Utilizando el método de los mínimos cuadrados se puede llegar a determinar el valor del coeficiente β y el término independiente α

El método de los mínimos cuadrados consiste en minimizar la suma de los cuadrados de las distancias verticales entre los valores verdaderos de Y y los valores estimados de $\hat{Y}$.

Las fórmulas de α y β
Donde

$$\beta = \frac{n\sum(X_iY_i) - \left(\sum X_i\right)\left(\sum Y_i\right)}{n\left(\sum X_i^2\right) - \left(\sum X_i\right)^2} \quad ; \quad \alpha = \frac{\sum Y_i}{n} - \beta * \frac{\sum X_i}{n} = \overline{Y} - \beta\overline{X}$$

Ejemplo 1: Descuentos comparados con las ventas

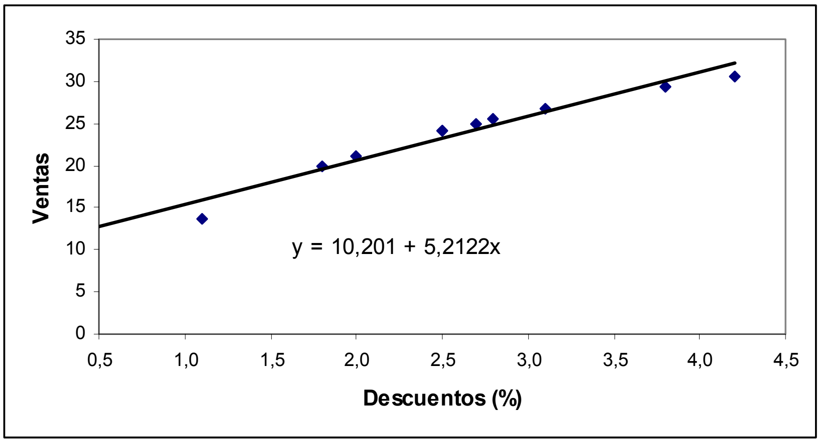
Nº	Xi	Yi	X_i^2	Y_i^2	X_i*Y_i
1	1.1	13.64	1.210	186.0496	15.004
2	2.5	24.05	6.250	578.4025	60.125
3	3.1	26.78	9.610	717.1684	83.018
4	4.2	30.63	17.640	938.1969	128.646
5	2.0	21.22	4.000	450.2884	42.44
6	2.5	24.05	6.250	578.4025	60.125
7	2.8	25.49	7.840	649.7401	71.372
8	2.7	25.03	7.290	626.5009	67.581
9	1.8	19.88	3.240	395.2144	35.784
10	3.8	29.36	14.440	862.0096	111.568
Sumas	**26.5**	**240.13**	**77.770**	**5981.9733**	**675.6630**

$$\beta = \frac{10 * 675{,}6630 - 26.5 * 240.13}{10 * 5981{,}9733 - 77{,}770} = 5.21119947$$

$$\alpha = \frac{240.13}{10} - 5{,}21119947 * \frac{26{,}5}{10} = 10.2033214$$

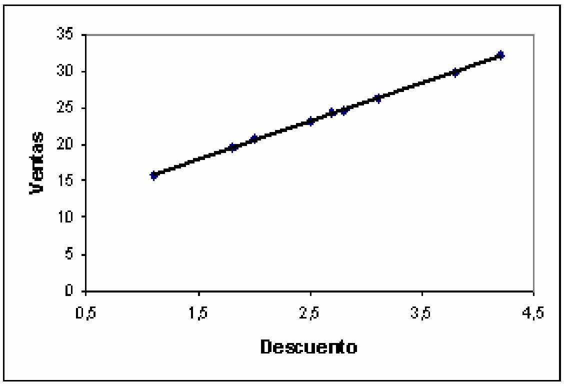
$$Modelo\ de\ regresión\ lineal \quad \hat{Y} = 10{,}2033214 + 5{,}21119947 * \hat{X}$$

Trazado de la línea de regresión



Obsérvese en el gráfico de dispersión anterior no todos los puntos quedan con exactitud en la recta de regresión. Si todos hubieran quedado en la línea y si el número de observaciones hubiera sido suficientemente grande, no existiría error en el cálculo del número de unidades vendidas. Dicho de otra forma, si todos los puntos estuvieran en recta de regresión, las ventas podrían pronosticarse con una precisión de 100% Entonces no habría error al pronosticar la variable Y con base en la variable X. Podemos tomar como ejemplo los siguientes datos históricos de ventas.

Nº	Descuentos	Ventas
1	1,1	16
2	2,5	23
3	3,1	26
4	4,2	32
5	2,0	21
6	2,5	23
7	2,8	25
8	2,7	24
9	1,8	20
10	3,8	30



Debido a que no hay diferencia entre los valores observados y los valores pronosticados, no existe error en esta estimación.

Obtener una predicción perfecta en los aspectos de economía y administración es prácticamente imposible. Por ejemplo, los ingresos anuales provenientes de ventas de gasolina (Y) con base en los registros de automóviles (X) hasta cierta fecha, sin duda podrían aproximarse con gran exactitud, pero el pronóstico no sería preciso con redondeo a unidades monetarias enteras, o tal vez hasta el millar de unidades monetarias. Aun los pronósticos de resistencia a la tensión mecánica de los alambres de acero, con base en el diámetro externo de los mismos, no siempre son exactos, debido a ligeras diferencias en la composición del acero.

Entonces, lo que se necesita es una medida que indique qué tan preciso es el pronóstico de Y con base en X o, por el contrario, cuan inexacta podría ser la predicción. A esta medida se le denomina ***error estándar de estimación***, el cual se representa por $s_{x,y}$ (es el mismo concepto que el de la desviación estándar). La desviación estándar mide la dispersión respecto a la línea de regresión.

Error estándar de estimación. Es la medida de la dispersión de los valores observados, con respecto a la línea de regresión.

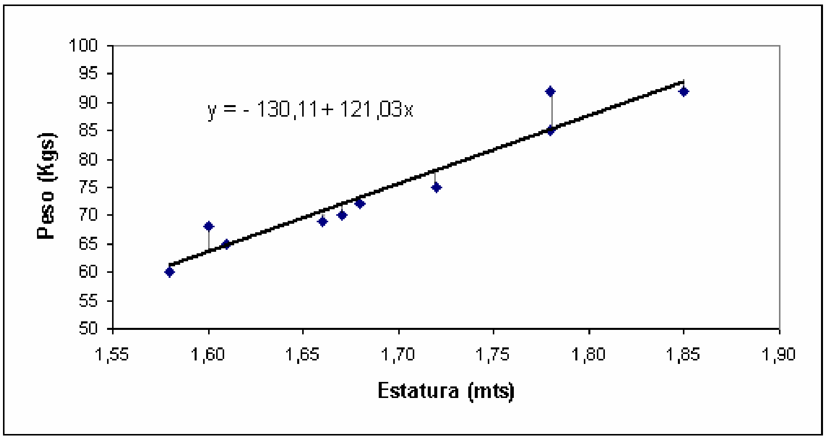
El error estándar de estimación se determina aplicando la siguiente ecuación. Obsérvese que ésta es muy semejante a la de la desviación estándar de una muestra.

$$ERROR\ ESTANDAR\ DE\ LA\ ESTIMACION:\quad S_{Y,X} = \sqrt{\frac{\sum\left(Y-\hat{Y}\right)^2}{n-2}}$$

N°	Xi	Yi	X_i^2	Y_i^2	X_i*Y_i	$\hat{Y}$	$(Y_i - \hat{Y})^2$
1	1.1	13.64	1.210	186.0496	15.004	15.9356408217	5.269966782
2	2.5	24.05	6.250	578.4025	60.125	23.2313200795	0.670236812
3	3.1	26.78	9.610	717.1684	83.018	26.3580397614	0.178050443
4	4.2	30.63	17.640	938.1969	128.646	32.0903591783	2.13264893
5	2.0	21.22	4.000	450.2884	42.44	20.6257203446	0.353168309
6	2.5	24.05	6.250	578.4025	60.125	23.2313200795	0.670236812
7	2.8	25.49	7.840	649.7401	71.372	24.7946799205	0.483470013
8	2.7	25.03	7.290	626.5009	67.581	24.2735599735	0.572201514
9	1.8	19.88	3.240	395.2144	35.784	19.5834804506	0.087923843
10	3.8	29.36	14.440	862.0096	111.568	30.0058793903	0.417160187
Sumas	**26.5**	**240.13**	**77.770**	**5981.9733**	**675.6630**	**240.1300000000**	**10.835063645**

$$ERROR\ ESTANDAR\ DE\ LA\ ESTIMACION:\quad S_{Y,X} = \sqrt{\frac{10{,}83506}{8}} = 1.163779599$$

Las desviaciones $Y_i - \hat{Y}$ son las desviaciones verticales con respecto a la recta de regresión [illegible]en sumar cero. Uno puntos estarán por encima de la recta y los otros por debajo.



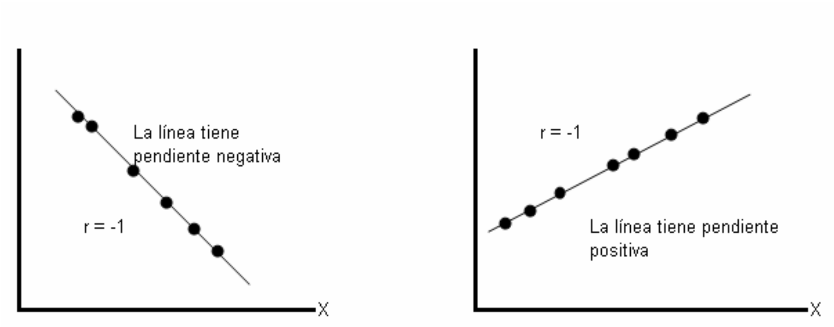
$$ERROR\ ESTANDAR\ DE\ LA\ ESTIMACION:\quad S_{Y,X} = \sqrt{\frac{\sum Y_i^2 - \alpha\left(\sum Y_i\right) - \beta\left(\sum X_i * Y_i\right)}{n-2}}$$

8.3 Coeficiente de correlación.

Es el valor que mide de laguna manera la intensidad de la relación entre dos conjuntos de variables.

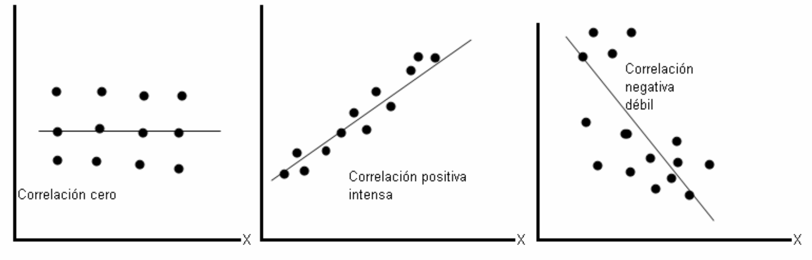
Originado por el investigador Karl Pearson, aproximadamente en el año 1900, el coeficiente de correlación describe la intensidad de la relación entre dos conjuntos de variables.

Como se le denota con r, con frecuencia se menciona también como r de Pearson, o como coeficiente de correlación. Puede tomar cualquier valor de -1.00 a +1.00, inclusive. Un coeficiente de correlación de -1.00 o de +1.00 indica una correlación perfecta. Un valor de -1.00 indicaría que las variables están perfectamente relacionadas en un sentido lineal negativo.



Si no existe en absoluto alguna relación entre los dos conjuntos de variables, la r de será cero. Un coeficiente de correlación r cercano a 0 (por ejemplo, 0.08) indica que relación es muy débil. Se llega a la misma conclusión si r = -0.08. Coeficientes de -0.91 +0.91 tienen igual fuerza; ambos indican una correlación muy intensa entre los dos conjuntos de variables. De modo que la fuerza de la correlación no depende de la dirección (ya sea –/+).

En el diagrama se muestran diagramas de dispersión para r = 0, una r débil (por ejemplo, -0.23) y una r fuerte (por ejemplo, +0.87). Obsérvese que si la correlación es débil, existe una dispersión considerable con respecto a una recta trazada a través del espacio central de los datos. Para que el diagrama de dispersión represente una relación fuerte, debe existir poca dispersión con respecto a la citada línea.



$$r = \frac{n\left(\sum X_i Y_i\right) - \left(\sum X_i\right)\left(Y_i\right)}{\sqrt{\left[n\left(\sum X_i^2\right) - \left(\sum X_i\right)^2\right]\left[n\left(\sum Y_i^2\right) - \left(\sum Y_i\right)^2\right]}}$$

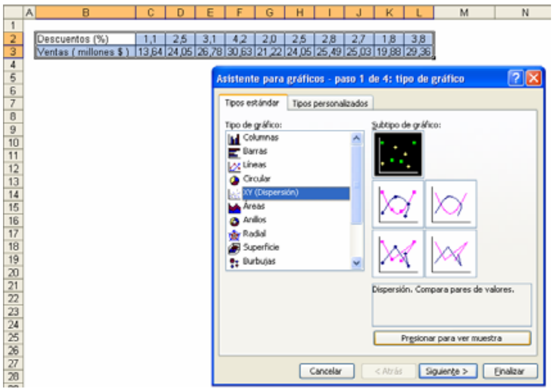
Nº	Xi	Yi	X_i^2	Y_i^2	X_i*Y_i
1	1.1	13.64	1.210	186.0496	15.004
2	2.5	24.05	6.250	578.4025	60.125
3	3.1	26.78	9.610	717.1684	83.018
4	4.2	30.63	17.640	938.1969	128.646
5	2.0	21.22	4.000	450.2884	42.44
6	2.5	24.05	6.250	578.4025	60.125
7	2.8	25.49	7.840	649.7401	71.372
8	2.7	25.03	7.290	626.5009	67.581
9	1.8	19.88	3.240	395.2144	35.784
10	3.8	29.36	14.440	862.0096	111.568
Sumas	**26.5**	**240.13**	**77.770**	**5981.9733**	**675.6630**

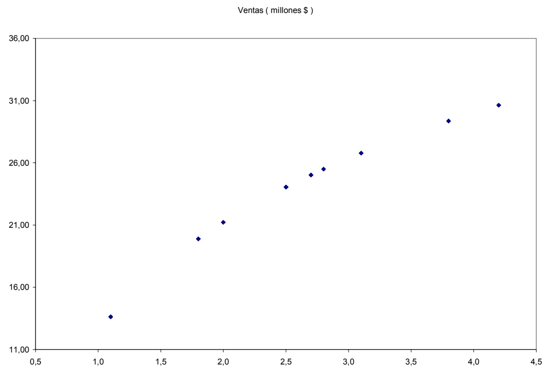
$$r = \frac{10 * 675{,}6630 - 26{,}5 * 240{,}13}{\sqrt{\left(10 * 77{,}770 - 26{,}5^2\right) * (10 * 5981.9733 - 240{,}13^2}} = 0.974564139$$

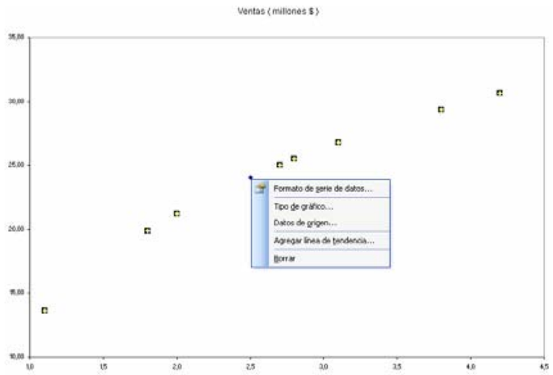
Coeficiente de determinación r^2. La proporción de la variación total en la variable dependiente Y, que se explica por la variación de la variable independiente X

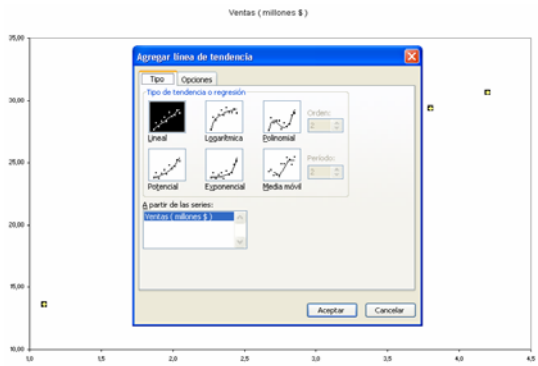
Existen calculadoras que vienen con el modelo de regresión para hacer los cálculos directos, solamente con entrar los datos históricos. Esto nos permite encontrar en forma rápida los parámetros del modelo de regresión lineal y el coeficiente de correlación.

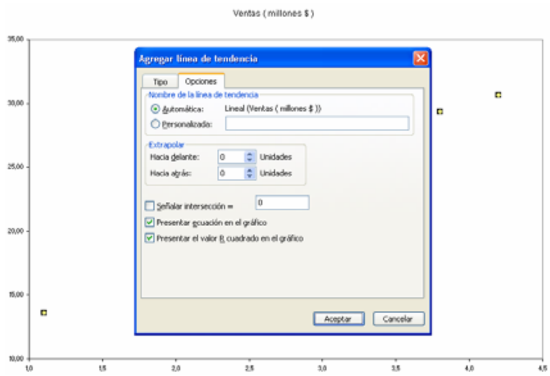
También podemos utilizar las hojas electrónicas como el Excel para encontrar los parámetros del modelo y hacer los pronósticos. Dentro del grupo de las funciones estadísticas están las que nos pueden servir para hacer el pronóstico.

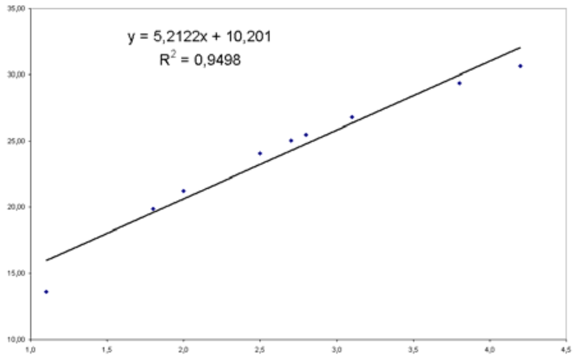



Ventas (millones $)
36,00
31,00
26,00
21,00
16,00
11,00
0,5
1,0
1,5
2,0
2,5
3,0
3,5
4,0
4,5


Ventas (millones $)
Formato de serie de datos...
Tipo de gráfico...
Datos de origen...
Agregar línea de tendencia...
Borrar


Ventas (millones $)
Agregar línea de tendencia
Tipo
Opciones
Tipo de tendencia o regresión
Lineal
Logarítmica
Polinomial
Orden:
Potencial
Exponencial
Media móvil
Período:
A partir de las series:
Ventas (millones $)
Aceptar
Cancelar

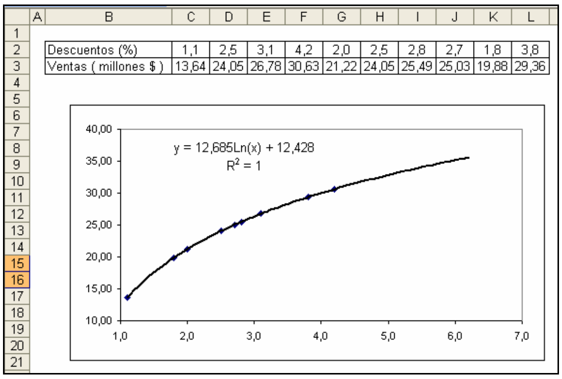

Ventas (millones $)
Agregar línea de tendencia
Tipo
Opciones
Nombre de la línea de tendencia
Automática:
Lineal (Ventas (millones $))
Personalizada:
Extrapolar
Hacia delante:
Unidades
Hacia atrás:
Unidades
Señalar intersección =
Presentar ecuación en el gráfico
Presentar el valor R cuadrado en el gráfico
Aceptar
Cancelar



8.5 Modelos de regresión no lineal

Modelos de regresión Logarítmica

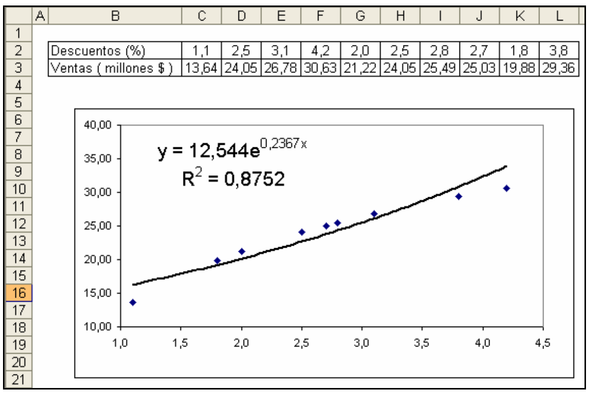
$Y = \alpha + \beta * LN(X)$



Descuentos (%)	1,1	2,5	3,1	4,2	2,0	2,5	2,8	2,7	1,8	3,8
Ventas (millones $)	13,64	24,05	26,78	30,63	21,22	24,05	25,49	25,03	19,88	29,36

Modelos de regresión Exponencial

$Y = \alpha * e^{\beta X}$ Modelo lineal equivalente: $LN(Y) = Ln(\alpha) + \beta X$ $Y' = \alpha' + \beta * X$

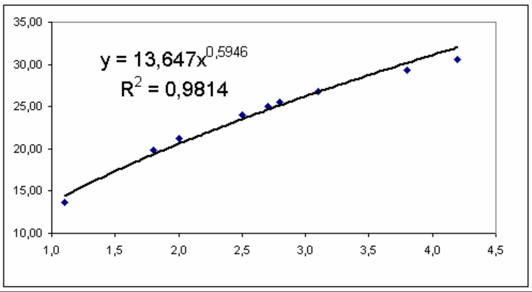


Descuentos (%)	1,1	2,5	3,1	4,2	2,0	2,5	2,8	2,7	1,8	3,8
Ventas (millones $)	13,64	24,05	26,78	30,63	21,22	24,05	25,49	25,03	19,88	29,36

Modelos de regresión Potencial

$Y = \alpha * X^{\beta}$ Modelo lineal equivalente: $LN(Y) = Ln(\alpha) + \beta * ln(X)$ $Y' = \alpha' + \beta * X'$

Descuentos (%)	1,1	2,5	3,1	4,2	2,0	2,5	2,8	2,7	1,8	3,8
Ventas (millones $)	13,64	24,05	26,78	30,63	21,22	24,05	25,49	25,03	19,88	29,36



EJERCICIO

1. Utilizando las calculadora o Excel, complete el siguiente cuadro y realice el pronóstico para los años 6, 7, 8, 9, 10

AÑO	1	2	3	4	5
VENTAS UNIDADES	490	555	585	608	625

Modelo Lineal	$Y = A + B*X$	
Modelo logaritmico	$Y = A + B*Ln(X)$	
Modelo Exponencial	$Y = A*e^{B*X} \leftrightarrow Ln(Y) = Ln(A) + B*X$	
Modelo Potencial	$Y = A*X^{B} \leftrightarrow Ln(Y)= Ln(A) + B*Ln(X)$	

				PRONOSTICO				
	A	B	R^2	6	7	8	9	10
Modelo Lineal								
Modelo logaritmico								
Modelo Exponencial								
Modelo Potencial								

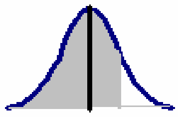
2. Utilizando las fórmulas estadísticas del Excel complete el siguiente cuadro

	A	B	C	D	E	F	G
1	AÑO	1	2	3	4	5	
2	VENTAS UNIDADES	490	555	585	608	625	
3							
4	Ln(X)						
5	Ln(Y)						
6							
7	Modelo Lineal	$Y = A + B*X$					
8	Modelo logaritmico	$Y = A + B*Ln(X)$					
9	Modelo Exponencial	$Y = A*e^{B*X} \leftrightarrow Ln(Y) = Ln(A) + B*X$					
10	Modelo Potencial	$Y = A*X^{B} \leftrightarrow Ln(Y)= Ln(A) + B*Ln(X)$					
11							
12	**PROYECCIONES CON LA FUNCION PRONOSTICO**						
13							
14	AÑO	6	7	8	9	10	
15	VENTAS UNIDADES						
16	Modelo Lineal						
17	Modelo logaritmico						
18	Modelo Exponencial						
19							
20	Modelo Potencial						
21							
22							
23	COEFICIENTE DE CORRELACION						
24	Modelo Lineal						
25	Modelo logaritmico						
26	Modelo Exponencial						
27	Modelo Potencial						
28							
29	COEFICIENTE R^2						
30	Modelo Lineal						
31	Modelo logaritmico						
32	Modelo Exponencial						
33	Modelo Potencial						

3. Taller. Realice el pronóstico para los años 6, 7, 8, 9 y 10.para los siguientes datos

AÑO	1	2	3	4	5	6
GASTOS	$ 45.000,00	$ 49.000,00	$ 54.500,00	$ 62.110,00	$ 67.079,00	$ 75.128,00

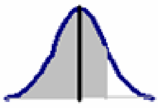
TABLA DE DISTRIBUCION NORMAL



$$Z = \frac{X_i - \bar{X}}{\sigma}$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	0.5000000	0.5039894	0.5079783	0.5119665	0.5159534	0.5199388	0.5239222	0.5279032	0.5318814	0.5358564	**0.0**
0.1	0.5398278	0.5437953	0.5477584	0.5517168	0.5556700	0.5596177	0.5635595	0.5674949	0.5714237	0.5753454	**0.1**
0.2	0.5792597	0.5831662	0.5870644	0.5909541	0.5948349	0.5987063	0.6025681	0.6064199	0.6102612	0.6140919	**0.2**
0.3	0.6179114	0.6217195	0.6255158	0.6293000	0.6330717	0.6368307	0.6405764	0.6443088	0.6480273	0.6517317	**0.3**
0.4	0.6554217	0.6590970	0.6627573	0.6664022	0.6700314	0.6736448	0.6772419	0.6808225	0.6843863	0.6879331	**0.4**
0.5	0.6914625	0.6949743	0.6984682	0.7019440	0.7054015	0.7088403	0.7122603	0.7156612	0.7190427	0.7224047	**0.5**
0.6	0.7257469	0.7290691	0.7323711	0.7356527	0.7389137	0.7421539	0.7453731	0.7485711	0.7517478	0.7549029	**0.6**
0.7	0.7580363	0.7611479	0.7642375	0.7673049	0.7703500	0.7733726	0.7763727	0.7793501	0.7823046	0.7852361	**0.7**
0.8	0.7881446	0.7910299	0.7938919	0.7967306	0.7995458	0.8023375	0.8051055	0.8078498	0.8105703	0.8132671	**0.8**
0.9	0.8159399	0.8185887	0.8212136	0.8238145	0.8263912	0.8289439	0.8314724	0.8339768	0.8364569	0.8389129	**0.9**
1.0	0.8413447	0.8437524	0.8461358	0.8484950	0.8508300	0.8531409	0.8554277	0.8576903	0.8599289	0.8621434	**1.0**
1.1	0.8643339	0.8665005	0.8686431	0.8707619	0.8728568	0.8749281	0.8769756	0.8789995	0.8809999	0.8829768	**1.1**
1.2	0.8849303	0.8868606	0.8887676	0.8906514	0.8925123	0.8943502	0.8961653	0.8979577	0.8997274	0.9014747	**1.2**
1.3	0.9031995	0.9049021	0.9065825	0.9082409	0.9098773	0.9114920	0.9130850	0.9146565	0.9162067	0.9177356	**1.3**
1.4	0.9192433	0.9207302	0.9221962	0.9236415	0.9250663	0.9264707	0.9278550	0.9292191	0.9305634	0.9318879	**1.4**
1.5	0.9331928	0.9344783	0.9357445	0.9369916	0.9382198	0.9394292	0.9406201	0.9417924	0.9429466	0.9440826	**1.5**
1.6	0.9452007	0.9463011	0.9473839	0.9484493	0.9494974	0.9505285	0.9515428	0.9525403	0.9535213	0.9544860	**1.6**
1.7	0.9554345	0.9563671	0.9572838	0.9581849	0.9590705	0.9599408	0.9607961	0.9616364	0.9624620	0.9632730	**1.7**
1.8	0.9640697	0.9648521	0.9656205	0.9663750	0.9671159	0.9678432	0.9685572	0.9692581	0.9699460	0.9706210	**1.8**
1.9	0.9712834	0.9719334	0.9725711	0.9731966	0.9738102	0.9744119	0.9750021	0.9755808	0.9761482	0.9767045	**1.9**
2.0	0.9772499	0.9777844	0.9783083	0.9788217	0.9793248	0.9798178	0.9803007	0.9807738	0.9812372	0.9816911	**2.0**
2.1	0.9821356	0.9825708	0.9829970	0.9834142	0.9838226	0.9842224	0.9846137	0.9849966	0.9853713	0.9857379	**2.1**
2.2	0.9860966	0.9864474	0.9867906	0.9871263	0.9874545	0.9877755	0.9880894	0.9883962	0.9886962	0.9889893	**2.2**
2.3	0.9892759	0.9895559	0.9898296	0.9900969	0.9903581	0.9906133	0.9908625	0.9911060	0.9913437	0.9915758	**2.3**
2.4	0.9918025	0.9920237	0.9922397	0.9924506	0.9926564	0.9928572	0.9930531	0.9932443	0.9934309	0.9936128	**2.4**
2.5	0.9937903	0.9939634	0.9941323	0.9942969	0.9944574	0.9946139	0.9947664	0.9949151	0.9950600	0.9952012	**2.5**
2.6	0.9953388	0.9954729	0.9956035	0.9957308	0.9958547	0.9959754	0.9960930	0.9962074	0.9963189	0.9964274	**2.6**
2.7	0.9965330	0.9966358	0.9967359	0.9968333	0.9969280	0.9970202	0.9971099	0.9971972	0.9972821	0.9973646	**2.7**
2.8	0.9974449	0.9975229	0.9975988	0.9976726	0.9977443	0.9978140	0.9978818	0.9979476	0.9980116	0.9980738	**2.8**
2.9	0.9981342	0.9981929	0.9982498	0.9983052	0.9983589	0.9984111	0.9984618	0.9985110	0.9985588	0.9986051	**2.9**
3.0	0.9986501	0.9986938	0.9987361	0.9987772	0.9988171	0.9988558	0.9988933	0.9989297	0.9989650	0.9989992	**3.0**
3.1	0.9990324	0.9990646	0.9990957	0.9991260	0.9991553	0.9991836	0.9992112	0.9992378	0.9992636	0.9992886	**3.1**
3.2	0.9993129	0.9993363	0.9993590	0.9993810	0.9994024	0.9994230	0.9994429	0.9994623	0.9994810	0.9994991	**3.2**
3.3	0.9995166	0.9995335	0.9995499	0.9995658	0.9995811	0.9995959	0.9996103	0.9996242	0.9996376	0.9996505	**3.3**
3.4	0.9996631	0.9996752	0.9996869	0.9996982	0.9997091	0.9997197	0.9997299	0.9997398	0.9997493	0.9997585	**3.4**
3.5	0.9997674	0.9997759	0.9997842	0.9997922	0.9997999	0.9998074	0.9998146	0.9998215	0.9998282	0.9998347	**3.5**
3.6	0.9998409	0.9998469	0.9998527	0.9998583	0.9998637	0.9998689	0.9998739	0.9998787	0.9998834	0.9998879	**3.6**
3.7	0.9998922	0.9998964	0.9999004	0.9999043	0.9999080	0.9999116	0.9999150	0.9999184	0.9999216	0.9999247	**3.7**
3.8	0.9999277	0.9999305	0.9999333	0.9999359	0.9999385	0.9999409	0.9999433	0.9999456	0.9999478	0.9999499	**3.8**
3.9	0.9999519	0.9999539	0.9999557	0.9999575	0.9999593	0.9999609	0.9999625	0.9999641	0.9999655	0.9999670	**3.9**
4.0	0.9999683	0.9999696	0.9999709	0.9999721	0.9999733	0.9999744	0.9999755	0.9999765	0.9999775	0.9999784	**4.0**
4.1	0.9999793	0.9999802	0.9999811	0.9999819	0.9999826	0.9999834	0.9999841	0.9999848	0.9999854	0.9999861	**4.1**
4.2	0.9999867	0.9999872	0.9999878	0.9999883	0.9999888	0.9999893	0.9999898	0.9999902	0.9999907	0.9999911	**4.2**
4.3	0.9999915	0.9999918	0.9999922	0.9999925	0.9999929	0.9999932	0.9999935	0.9999938	0.9999941	0.9999943	**4.3**
4.4	0.9999946	0.9999948	0.9999951	0.9999953	0.9999955	0.9999957	0.9999959	0.9999961	0.9999963	0.9999964	**4.4**
4.5	0.9999966	0.9999968	0.9999969	0.9999971	0.9999972	0.9999973	0.9999974	0.9999976	0.9999977	0.9999978	**4.5**
4.6	0.9999979	0.9999980	0.9999981	0.9999982	0.9999983	0.9999983	0.9999984	0.9999985	0.9999986	0.9999986	**4.6**
4.7	0.9999987	0.9999988	0.9999988	0.9999989	0.9999989	0.9999990	0.9999990	0.9999991	0.9999991	0.9999992	**4.7**
4.8	0.9999992	0.9999992	0.9999993	0.9999993	0.9999994	0.9999994	0.9999994	0.9999994	0.9999995	0.9999995	**4.8**
4.9	0.9999995	0.9999995	0.9999996	0.9999996	0.9999996	0.9999996	0.9999996	0.9999997	0.9999997	0.9999997	**4.9**
5.0	0.9999997	0.9999997	0.9999997	0.9999998	0.9999998	0.9999998	0.9999998	0.9999998	0.9999998	0.9999998	**5.0**
Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**	**Z**

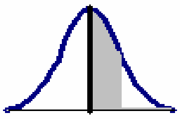
TABLA DE DISTRIBUCION NORMAL



$$Z = \frac{X_i - \overline{X}}{\sigma}$$

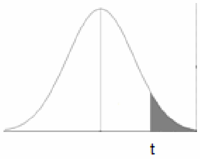
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586	0.0
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535	0.1
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409	0.2
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173	0.3
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793	0.4
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240	0.5
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490	0.6
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524	0.7
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327	0.8
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891	0.9
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214	1.0
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298	1.1
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147	1.2
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774	1.3
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189	1.4
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408	1.5
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449	1.6
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327	1.7
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062	1.8
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670	1.9
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169	2.0
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574	2.1
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899	2.2
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158	2.3
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361	2.4
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520	2.5
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643	2.6
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736	2.7
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807	2.8
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861	2.9
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900	3.0
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929	3.1
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950	3.2
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965	3.3
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976	3.4
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983	3.5
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989	3.6
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992	3.7
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995	3.8
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997	3.9
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998	4.0
4.1	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99998	0.99999	0.99999	4.1
4.2	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	4.2
4.3	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	4.3
4.4	0.99999	0.99999	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	4.4
4.5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	4.5
4.6	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	4.6
4.7	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	4.7
4.8	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	4.8
4.9	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	4.9
5.0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	5.0
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z

TABLA DE DISTRIBUCION NORMAL



$$Z = \frac{X_i - \bar{X}}{\sigma}$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586	**0.0**
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535	**0.1**
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409	**0.2**
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173	**0.3**
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793	**0.4**
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240	**0.5**
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490	**0.6**
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524	**0.7**
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327	**0.8**
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891	**0.9**
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214	**1.0**
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298	**1.1**
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147	**1.2**
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41309	0.41466	0.41621	0.41774	**1.3**
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189	**1.4**
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408	**1.5**
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449	**1.6**
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327	**1.7**
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062	**1.8**
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670	**1.9**
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169	**2.0**
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574	**2.1**
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899	**2.2**
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158	**2.3**
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361	**2.4**
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520	**2.5**
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643	**2.6**
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736	**2.7**
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807	**2.8**
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861	**2.9**
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900	**3.0**
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929	**3.1**
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950	**3.2**
3.3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965	**3.3**
3.4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976	**3.4**
3.5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983	**3.5**
3.6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989	**3.6**
3.7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992	**3.7**
3.8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995	**3.8**
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997	**3.9**
4.0	0.49997	0.49997	0.49997	0.49997	0.49997	0.49997	0.49998	0.49998	0.49998	0.49998	**4.0**
4.1	0.49998	0.49998	0.49998	0.49998	0.49998	0.49998	0.49998	0.49998	0.49999	0.49999	**4.1**
4.2	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	**4.2**
4.3	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	**4.3**
4.4	0.49999	0.49999	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	**4.4**
4.5	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	**4.5**
4.6	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	**4.6**
4.7	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	**4.7**
4.8	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	**4.8**
4.9	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	**4.9**
5.0	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	**5.0**
Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**	**Z**



=DISTR.T.INV(α;GL)

GL	Intervalos de confianza (1- α)													
	80%	85%	86%	88%	90%	92%	94%	92%	95%	96.00%	97%	98%	99%	99.90%
	Nivel de significancia para pruebas de una cola													
	0.10	0.075	0.07	0.06	0.05	0.04	0.03	0.04	0.025	0.02	0.015	0.01	0.005	0.0005
	Nivel de significancia para pruebas de dos colas													
	0.2000	0.1500	0.1400	0.1200	0.1000	0.0800	0.0600	0.0800	0.0500	0.0400	0.0300	0.0200	0.0100	0.0010
1	3.07768	4.16530	4.47374	5.24218	6.31375	7.91582	10.57889	7.91582	12.70620	15.89454	21.20495	31.82052	63.65674	636.61925
2	1.88562	2.28193	2.38338	2.62016	2.91999	3.31976	3.89643	3.31976	4.30265	4.84873	5.64278	6.96456	9.92484	31.59905
3	1.63774	1.92432	1.99502	2.15624	2.35336	2.60543	2.95051	2.60543	3.18245	3.48191	3.89605	4.54070	5.84091	12.92398
4	1.53321	1.77819	1.83752	1.97123	2.13185	2.33287	2.60076	2.33287	2.77645	2.99853	3.29763	3.74695	4.60409	8.61030
5	1.47588	1.69936	1.75289	1.87268	2.01505	2.19096	2.42158	2.19096	2.57058	2.75651	3.00287	3.36493	4.03214	6.86883
6	1.43976	1.65017	1.70021	1.81165	1.94318	2.10431	2.31326	2.10431	2.44691	2.61224	2.82893	3.14267	3.70743	5.95882
7	1.41492	1.61659	1.66430	1.77021	1.89458	2.04601	2.24088	2.04601	2.36462	2.51675	2.71457	2.99795	3.49948	5.40788
8	1.39682	1.59222	1.63827	1.74024	1.85955	2.00415	2.18915	2.00415	2.30600	2.44898	2.63381	2.89646	3.35539	5.04131
9	1.38303	1.57374	1.61854	1.71758	1.83311	1.97265	2.15038	1.97265	2.26216	2.39844	2.57380	2.82144	3.24984	4.78091
10	1.37218	1.55924	1.60308	1.69984	1.81246	1.94810	2.12023	1.94810	2.22814	2.35931	2.52748	2.76377	3.16927	4.58689
11	1.36343	1.54756	1.59063	1.68558	1.79588	1.92843	2.09614	1.92843	2.20099	2.32814	2.49066	2.71808	3.10581	4.43698
12	1.35622	1.53796	1.58040	1.67387	1.78229	1.91231	2.07644	1.91231	2.17881	2.30272	2.46070	2.68100	3.05454	4.31779
13	1.35017	1.52992	1.57184	1.66408	1.77093	1.89887	2.06004	1.89887	2.16037	2.28160	2.43585	2.65031	3.01228	4.22083
14	1.34503	1.52310	1.56458	1.65578	1.76131	1.88750	2.04617	1.88750	2.14479	2.26378	2.41490	2.62449	2.97684	4.14045
15	1.34061	1.51723	1.55833	1.64865	1.75305	1.87774	2.03429	1.87774	2.13145	2.24854	2.39701	2.60248	2.94671	4.07277
16	1.33676	1.51213	1.55291	1.64246	1.74588	1.86928	2.02400	1.86928	2.11991	2.23536	2.38155	2.58349	2.92078	4.01500
17	1.33338	1.50766	1.54815	1.63703	1.73961	1.86187	2.01500	1.86187	2.10982	2.22385	2.36805	2.56693	2.89823	3.96513
18	1.33039	1.50371	1.54395	1.63224	1.73406	1.85534	2.00707	1.85534	2.10092	2.21370	2.35618	2.55238	2.87844	3.92165
19	1.32773	1.50019	1.54020	1.62797	1.72913	1.84953	2.00002	1.84953	2.09302	2.20470	2.34565	2.53948	2.86093	3.88341
20	1.32534	1.49704	1.53685	1.62415	1.72472	1.84433	1.99371	1.84433	2.08596	2.19666	2.33624	2.52798	2.84534	3.84952
21	1.32319	1.49419	1.53383	1.62071	1.72074	1.83965	1.98804	1.83965	2.07961	2.18943	2.32779	2.51765	2.83136	3.81928
22	1.32124	1.49162	1.53109	1.61759	1.71714	1.83542	1.98291	1.83542	2.07387	2.18289	2.32016	2.50832	2.81876	3.79213
23	1.31946	1.48928	1.52860	1.61476	1.71387	1.83157	1.97825	1.83157	2.06866	2.17696	2.31323	2.49987	2.80734	3.76763
24	1.31784	1.48714	1.52633	1.61217	1.71088	1.82805	1.97399	1.82805	2.06390	2.17154	2.30691	2.49216	2.79694	3.74540
25	1.31635	1.48517	1.52424	1.60979	1.70814	1.82483	1.97010	1.82483	2.05954	2.16659	2.30113	2.48511	2.78744	3.72514
26	1.31497	1.48336	1.52232	1.60760	1.70562	1.82186	1.96651	1.82186	2.05553	2.16203	2.29581	2.47863	2.77871	3.70661
27	1.31370	1.48169	1.52054	1.60558	1.70329	1.81913	1.96320	1.81913	2.05183	2.15782	2.29091	2.47266	2.77068	3.68959
28	1.31253	1.48014	1.51890	1.60371	1.70113	1.81659	1.96014	1.81659	2.04841	2.15393	2.28638	2.46714	2.76326	3.67391
29	1.31143	1.47870	1.51737	1.60197	1.69913	1.81424	1.95729	1.81424	2.04523	2.15033	2.28217	2.46202	2.75639	3.65941
30	1.31042	1.47736	1.51595	1.60035	1.69726	1.81205	1.95465	1.81205	2.04227	2.14697	2.27826	2.45726	2.75000	3.64596
40	1.30308	1.46772	1.50570	1.58871	1.68385	1.79631	1.93566	1.79631	2.02108	2.12291	2.25027	2.42326	2.70446	3.55097
60	1.29582	1.45820	1.49560	1.57723	1.67065	1.78085	1.91703	1.78085	2.00030	2.09936	2.22292	2.39012	2.66028	3.46020
100	1.29007	1.45067	1.48761	1.56817	1.66023	1.76866	1.90237	1.76866	1.98397	2.08088	2.20150	2.36422	2.62589	3.39049
200	1.28580	1.44508	1.48168	1.56144	1.65251	1.75963	1.89152	1.75963	1.97190	2.06723	2.18568	2.34514	2.60063	3.33984
500	1.28325	1.44175	1.47814	1.55743	1.64791	1.75425	1.88507	1.75425	1.96472	2.05912	2.17630	2.33383	2.58570	3.31009
1000	1.28240	1.44064	1.47696	1.55610	1.64638	1.75247	1.88293	1.75247	1.96234	2.05643	2.17319	2.33008	2.58075	3.30028

RAFAEL VARGAS BARRERA

Profesor ESAP
Matemático
Especialista en Costos y Control de Gestión
Master en finanzas
rvargas@etb.net.co