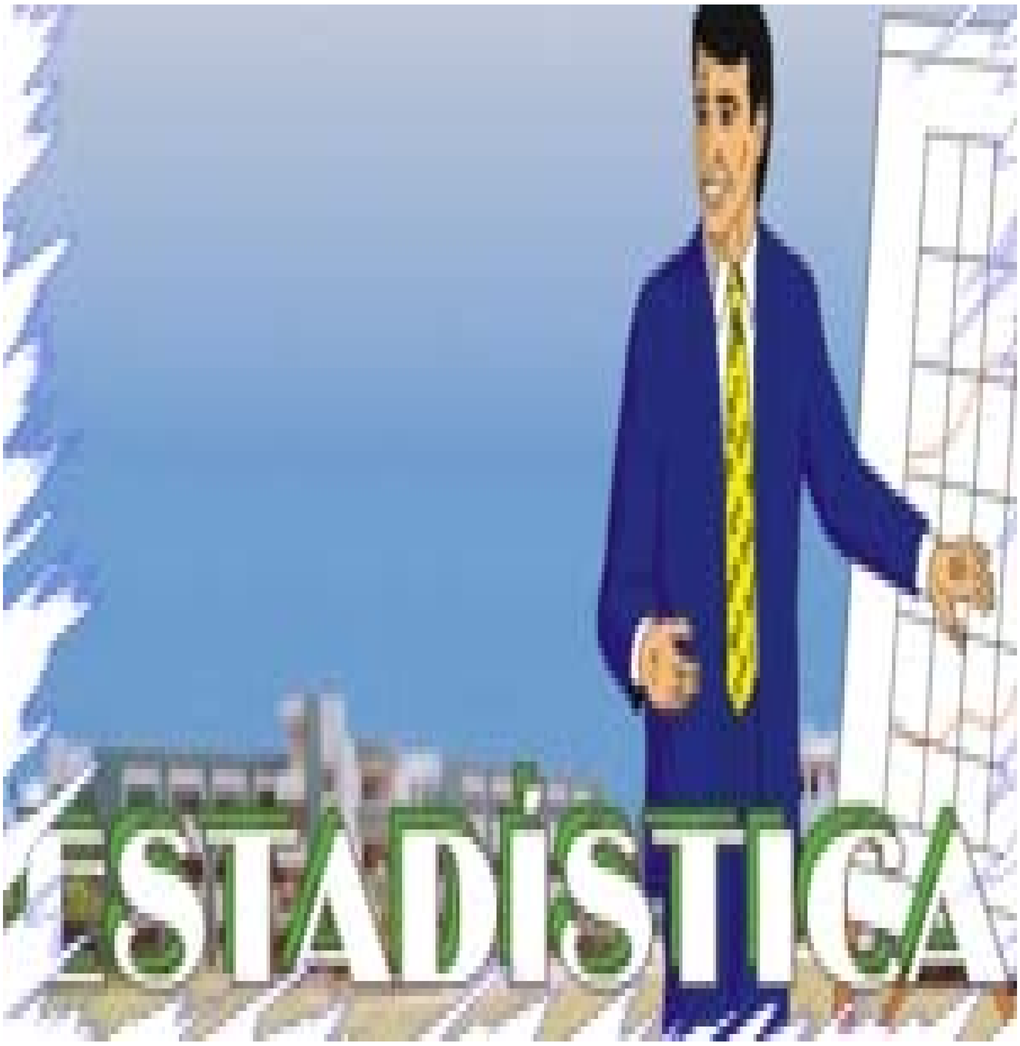
UNIDAD I

¿Qué es la predicción empresarial?

UNIDAD I

¿Qué es la predicción empresarial?

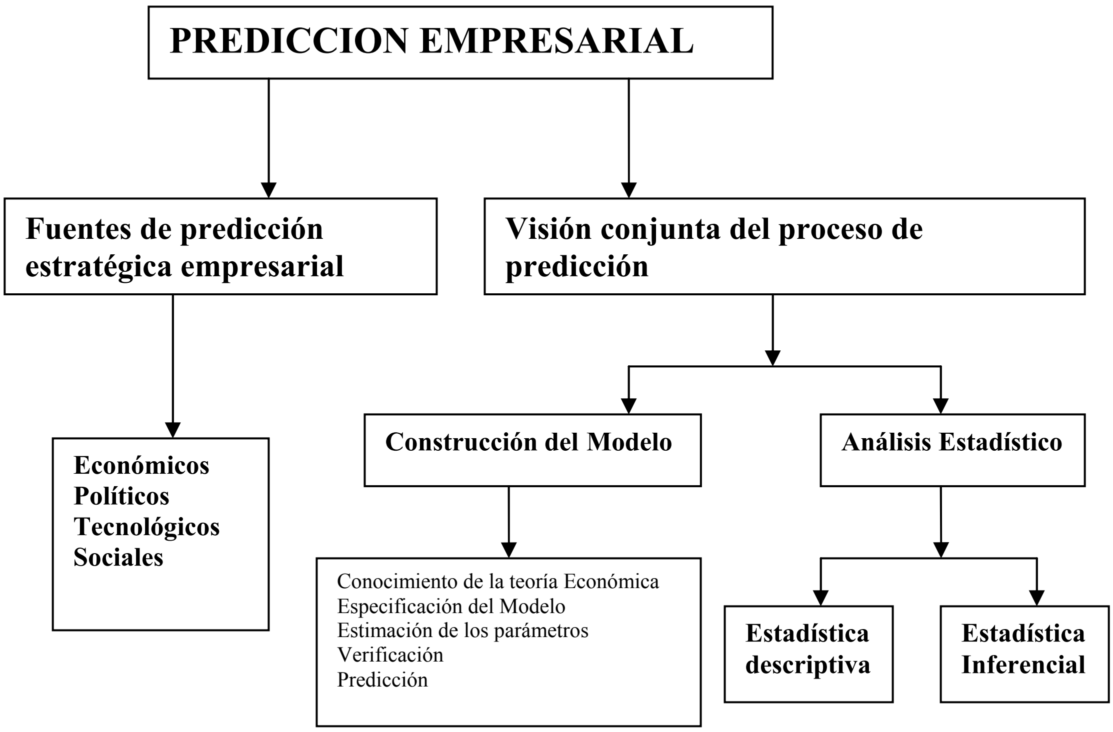


"Predecir es pronosticar con éxito"

- ¿Cuáles son las fuentes de predicción estratégica - empresarial?
- ¿Cuál es la visión conjunta del proceso de predicción?
- ¿Qué es la estadística descriptiva?
- ¿Qué es la estadística inferencial?

PREDICCION EMPRESARIAL

ESQUEMA CONCEPTUAL



COMPETENCIAS A LOGRAR

CONCEPTUAL	PROCEDIMENTAL	ACTITUDINAL
Explica que es la Predicción empresarial, sus fuentes y la visión conjunta del proceso de predicción	Evalúa las competencias para ser un administrador con éxito en el campo de la predicción.	Promueve la responsabilidad, al tomar decisiones después de analizar la predicción como una técnica estadística…

Conceptos –clave

Fuentes de predicción estratégica – empresarial: económicos, políticos económicos, tecnológicos y sociales

LECCIÓN 1

PREDICCIÓN EMPRESARIAL

Generalidades

El éxito o fracaso de una gestión empresarial se juzga sobre la base de los resultados obtenidos y la eficacia en la ejecución de los proyectos. De acuerdo con esto, el responsable de dicha gestión deberá presumirse de la mayor cantidad de instrumentos que le permitan obtener sus objetivos de la mejor manera.

La importancia de la predicción es que "*el futuro sólo puede ser inferido a partir de un profundo estudio del pasado"* y aceptando que *"el comportamiento de los agentes no se modifica sustancialmente"*

La toma de decisiones en asuntos relativos al ambiente empresarial ocupa un lugar relevante y se convierte en una actividad permanente. En este proceso de toma de decisiones, se puede recurrir adicionalmente a un conjunto de técnicas o instrumentos cuantitativos que otorgan elementos de medición para evaluar el impacto de las decisiones potenciales sobre las variables objetivo, incorporando la menor proporción de elementos subjetivos. Ambos procesos –el experimental y el cuantitativo- se debe complementar para definir la dirección de la decisión a tomar, y es en el segundo proceso mencionado que la estadística proporciona un vehículo útil para el administrador, el empresario o en general, el responsable de la gestión. La predicción, analizada como una técnica estadística, cumple justamente este papel.

Predecir es, entonces, encontrar el valor que puede tomar una variable objetivo fuera del ámbito (temporal o espacial) de análisis.

La predicción como una técnica estadística define cuál de los posibles valores futuros de la variable es más probable. Dicho de otro modo: se asigna una probabilidad de ocurrencia a cada uno de los posibles valores de la variable objetivo, con lo que se podrá hacer una lista priorizada sobre la base de su factibilidad probabilística. (Cortez Cumpa, 1992).

En consecuencia, se puede afirmar que el objetivo del aprendizaje es brindar a los responsables de la gestión empresarial un conjunto de técnicas estadísticas, que puedan complementar con su experiencia para el pronóstico de las principales variables de decisión, así como otorgar criterios para formular modelos que permitan explicar y predecir el comportamiento de los agentes económicos.

En esta unidad se exponen además, las fuentes de predicción estratégica y empresarial. También se desarrolla una visión conjunta del proceso de predicción. Por otro lado, se presenta los principales fundamentos de la estadística descriptiva e inferencial.

LECCIÓN 2

FACTORES QUE INCIDEN EN EL FUNCIONAMIENTO DE LA EMPRESA

Durante el tiempo de vida de una empresa, el mundo exterior a ella ejerce diferentes tipos de influencia los cuales se originan por la presencia de ciertos factores que unas veces son favorables y otras perjudiciales.

Los factores que influyen en el funcionamiento de una empresa que van a servir como fuentes de predicción estratégica empresarial se agrupan de la siguiente manera:

1. ECONÓMICOS

Entre ellos tenemos la situación de los mercados y de la competencia, índice de crecimiento de la población, la disponibilidad o carencia de divisas, el poder adquisitivo de la población, la disponibilidad de materias primas nacionales y en plano internacional, la situación general del comercio en el mundo, la fuerza competitiva de otros países y la disponibilidad de divisas de los posibles clientes extranjeros.

2. POLÍTICOS

Entre ellos están las políticas adoptadas por los gobiernos o autoridades locales en cuestiones económicas y sociales. Entre las primeras podrían citarse la política general del gobierno respecto a las empresas privadas y estatales y respecto al empleo, ubicación de las industrias, subvenciones o protección a las industrias nacionales, divisas y permisos, impuestos y reducción de impuestos.

3. TECNOLÓGICOS

La influencia de los factores tecnológicos sobre una empresa dependerá en gran medida del campo en que esa empresa esté actuando, así como, del país donde se establezca.

Donde los cambios se producen con mucha rapidez, las empresas que deseen mantener su poder competitivo tienen que invertir mucho dinero en planeamiento, investigación y perfeccionamiento.

4. SOCIALES

Como la empresa es una organización social, a la vez que una unidad económica, está sujeta a las presiones sociales y a las influencias del medio en que está establecida, Entre los factores que repercutirán en sus actividades tenemos: las tradiciones culturales del país, la estructura de los gastos en cuanto a mercancías y servicios, la actitud respecto al dinero, al trabajo, a la mujer; el grado en que se está orientando a la agricultura, los orígenes de la industria. Ellos van a influir en la empresa desde el interior y exterior de la misma.

LECCIÓN 3

VISIÓN CONJUNTA DEL PROCESO DE PREDICCIÓN

El objetivo del proceso de predicción es dar al administrador o al responsable de la gestión de la empresa las técnicas estadísticas de predicción, que le permita pronosticar las principales variables.

El investigador debe describir en forma clara las fuentes de los datos utilizados, así como analizar el conjunto de variables que están interrelacionadas. Todo esto para construir un Modelo, el cual debe seguir las siguientes etapas:

- Conocimiento de la Teoría Económica
- Especificación del Modelo
- Estimación de los parámetros
- Verificación
- Predicción o Pronóstico

Es decir que dado el Modelo, se estimarán los parámetros, la varianza de los términos de perturbación (error), la desviación estándar, las correlaciones entre las variables, etc. Luego se plantearán Pruebas de hipótesis para verificar la "significancia" del modelo.

La técnica estadística conocida como "Análisis de Regresión" es la herramienta principal utilizada para obtener los valores estimados.

Los estimadores de los parámetros estructurales, obtenidos por cualquiera de las técnicas, serán usadas para saber cuál es el valor que el modelo estima para las variables endógenas (dependientes) cuando son conocidos los valores de las variables exógenas (independientes). Esto es el proceso de simulación y se usa para calcular la bondad de ajuste y la capacidad de predicción del modelo estimado

Posteriormente, se encontrará la predicción media y predicción de un valor individual.

Por último, se aplicará el modelo estimado, el cual se utilizará para predecir el o los valores futuros de la variable dependiente o de pronóstico.

LECCIÓN 4

LA ESTADÍSTICA DESCRIPTIVA E INFERENCIAL

1. LA ESTADÍSTICA DESCRIPTIVA

Es el proceso de recopilación, organización y presentación de un conjunto de datos. Mediante este proceso se puede describir el comportamiento de una o más variables.

1.1 Organización de la información

Los valores disponibles de una variable pueden ser referidos a tiempo o espacio en el cual se verifican. El orden resulta importante y los valores constituyen una serie.

Por ejemplo, se recoge los resultados de una encuesta realizada durante una semana a cuarenta alumnos de un colegio, la cual fue ejecutada para determinar el consumo potencial de una nueva leche.

A continuación se presenta la forma en que se ha organizado la información disponible mediante el uso de la tabla de distribución de frecuencias de la variable que se ha considerado.

En la primera columna se representan todos los valores positivos de la variable analizada.

Frecuencia absoluta.- (O distribución de frecuencias de las variables) Representa la intensidad de cada una de las variables; esto es, el número de veces que se repite el valor en la encuesta realizada (Segunda columna de la Tabla: número de alumnos).
La suma de las frecuencias absolutas es igual al número de encuestas.

Frecuencia relativa.- Representa la intensidad de un determinado valor, expresado en términos de porcentaje del total de eventos del análisis (Tercera columna: proporción de alumnos).

Frecuencia acumulada.- Indica el número de veces que se repite un determinado valor y todos los menores a él (Cuarta columna).

Frecuencia relativa acumulada.- Indica el número de las veces que se repite un determinado valor y todos los menores a él, en porcentajes. Sirve de base para el desarrollo de distribuciones muestrales y funciones de densidad, en las cuales se fundamentan las pruebas de hipótesis (Quinta columna: proporción acumulada de alumnos).

Nº de vasos de leche en la semana	Frecuencia absoluta (f_i)	Frecuencia relativa (h_i)	Frecuencia acumulada (f_i)	Frecuencia relativa acumulada(H_i)
3	6	0.15	6	0.15
4	10	0.25	16	0.40
5	8	0.20	24	0.60
6	6	0.15	30	0.75
7	4	0.10	34	0.85
8	3	0.08	37	0.93
9	2	0.05	39	0.98
10	1	0.02	40	1.00
Total	40	1.00		

En el caso que se quiera determinar que aula tiene una mayor demanda de consumo de leche, bastará con comparar los valores promedios de consumo en las diferentes aulas, para lo que será necesario organizar la información de la forma anteriormente mencionada.

1.2 Pasos para la Construcción de Frecuencias

Previamente se definirán algunos términos:

Rango (R): Es la diferencia entre el máximo y el mínimo valor del conjunto de observaciones consideradas

Clase: Es cada una de las agrupaciones en las cuales se ha dividido la información

Amplitud de clase (a): Es la diferencia entre el límite superior con el límite inferior para cada clase. Por lo general se trabajará con clases de igual tamaño.

Numero de clases (m): Es la cantidad de clases que se van a considerar para el agrupamiento de la información. En la práctica puede adoptarse un número que oscile entre 5 y 15. Su valor dependerá de los objetivos del trabajo estadístico que se pretenda realizar.

Los pasos a seguir para la construcción de una tabla de frecuencias es la siguiente:

1º. Determinar el rango o recorrido: R = Vmáx – Vmín
2º. Determinar el número de clase o de intervalo en los cuales se desea representar la información : $m = 1 + 3.3 \log(n)$ (Fórmula de S[1]turges)
3º. Amplitud de cada intervalo: $a = R / m$
4º. Determinar los límites de clases L_i y L_s ; desde el valor mínimo se establece el primer límite de clase, a este se le suma la amplitud para obtener la segunda clase y así sucesivamente.

[1] En la práctica puede adoptarse un número de intervalos que oscile entre 5 y 15, a criterio y según el objetivo del trabajo estadístico.

Donde:

i. L_i : Límite inferior del intervalo

ii. L_s : Límite superior del intervalo

5º. Determinación de la frecuencia: Consiste en contar cuantos datos caen dentro de cada clase.

6º. Calcular las marcas de clase: $mc = (L_i + L_s)/2$

1.3 Estadísticos de Tendencia Central

Media Aritmética.- Es la suma de los valores de X, dividida por el número de valores. La media aritmética tiene un uso difundido debido a sus propiedades:

- La suma de las desviaciones de los valores de la variable en torno a ella es cero.
- La suma de los cuadrados de las desviaciones de los valores de la variable con respecto de ella es menor que con respecto a cualquier otro valor.

Media aritmética de una población

$$\mu = \frac{\sum X_i}{N} \quad , \quad i=1\ldots N$$

X_i : representa al elemento i esimo de la variable X en la población

N : número total de observaciones en la población

Media aritmética de una muestra

$$\overline{X} = \frac{\sum X_i}{n} \quad , \quad i=1\ldots n$$

Xi: representa al elemento i esimo de la variable X en la muestra

n : número total de elementos en la muestra

Mediana.- Es el valor central de un conjunto ordenado de observaciones, por lo tanto su valor resulta menor o igual que la mitad de los valores y a la vez mayor o igual que la mitad del total de los mismos. La mediana posee dos características importantes:

- No es afectada por los valores extremos de la variable, los mismos que podrían no ser representativos.
- La suma del valor absoluto de las desviaciones de los valores de la variable respecto de ella es menor que respecto de cualquier otro valor.

La fórmula de la mediana

$$\text{Posición de la mediana} = \frac{N+1}{2}$$

N: número de observaciones

Moda.- Es aquel valor del conjunto que más se repite o, lo que es lo mismo, aquel que tiene la mayor frecuencia. Se caracteriza por maximizar el número de desviaciones nulas respecto de ella.

Su principal ventaja radica en que es fácilmente obtenible, y una limitación importante es que su papel de valor representativo de un conjunto de valores es muy frágil.

Ejemplo Ilustrativo de Media Aritmética, Mediana y Moda

Se desea calcular **la media aritmética** de los cinco exámenes de este semestre del curso "Técnicas estadísticas de predicción". Las notas conseguidas son las siguientes: 20, 18, 15, 19 15.

$$\mu = \frac{\sum X_i}{N}$$

$$\mu = \frac{20+18+15+19+15}{5} = 17.4$$

Utilizado en el cálculo las cinco notas obtenidas, la media aritmética de las notas del curso es 17.4.

Si se desea calcular la **mediana,** primero se coloca las notas en serie ordenada ascendente y tenemos: 15, 15, 18, 19, 20

$$\text{Posición de la mediana} = \frac{N+1}{2}$$

$$\textit{Posición de la mediana} = \frac{5+1}{2} = 3$$

La posición de la puntuación de la mediana es 3, por lo consiguiente la mediana es 18, que ocupa el tercer lugar de los datos colocados ordenadamente.

Si se desea calcular la **moda,** la nota modal es 15 porque se repitió más veces que ninguna otra nota

1.4 Estadísticos de Dispersión

La dispersión de los datos en la población afecta la representatividad del promedio en una muestra de valores. En la medida que estén más disperso los datos se requiere una muestra mayor para asegurar que la media muestral sea representativa de la media poblacional.

Los estadísticos de dispersión más utilizados son los siguientes:

Amplitud de Rango.- Es la diferencia entre el mayor y el menor valor del conjunto de datos.

Ejemplo:

Valor mayor – Valor menor = 20 – 15 = 5

Rango Intercuartílico.- Se puede obtener de la diferencia entre el tercer y primer cuartil. Se entiende por cuartiles aquellos valores que dividen el total del conjunto ordenado ascendentemente en cuatro grupos. Así, el primer cuartil será aquel valor menor o igual al 75% del total de datos y mayor o igual al 25% del mismo total. El tercer cuartil es mayor o igual al 75% de los mismos.

Varianza y Desviación Estándar

La Varianza.-Se define como el promedio de las desviaciones respecto de la media elevadas al cuadrado. Para hallar la varianza se realizan las siguientes operaciones:

- Restar la media aritmética de cada una de las observaciones.
- Elevar al cuadrado estas desviaciones de la media
- Hallar la media aritmética de estas desviaciones

$$\sigma^2 = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + (X_3 - \mu)^2 + ... + (X_N - \mu)^2}{N}$$

La fórmula de la varianza de una población es:

$$\sigma^2 = \frac{\Sigma(X_i - \mu)^2}{N}$$

La Desviación Estándar.-Se define como la raíz cuadrada de la varianza. La fórmula de la desviación estándar es:

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

La varianza de una muestra.- Se calcula de manera similar a la de una población. La varianza de una muestra es:

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

La desviación estándar de la muestra.- La fórmula de la desviación estándar de la muestra es:

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

El Coeficiente de Variación.- Es un número abstracto que, denotado por CV, se obtiene como cociente entre la desviación estándar y su media aritmética.

$$CV = \frac{\sigma}{\mu} \times 100 \quad para\, una\ \ población$$

$$CV = \frac{S}{\overline{x}} \times 100 \quad para\, una\, muestra$$

σ : desviación estándar poblacional.
S : desviación estándar muestral.
μ : media aritmética poblacional.
$\overline{x}$: media aritmética muestral.

1.5 La Media Aritmética, la Mediana y la Moda a partir de Datos Agrupados

Media Aritmética.- Es el cálculo de la media aritmética a partir de datos agrupados, se adopta la hipótesis de que las observaciones de cada clase son iguales a la marca de clase. Una vez aceptada la hipótesis, se tiene en cuenta la frecuencia y la marca de clase de cada una de las clases al calcular la media a partir de datos agrupados.

Fórmula:

$$\overline{X}_g = \frac{\sum f_i X_i}{n}$$

Donde:

- f_i : Frecuencia o número de observaciones de cada clase
- X_i : Marca de clase
- n : Tamaño de la muestra, es igual a la suma de frecuencias en todas las clases

Mediana.- Es la clase cuya frecuencia acumulada es mayor o igual que n/2

Fórmula:

$$Me = L + a\,\frac{n/2 - F_{i-1}}{f_i}$$

Donde:

- L : Límite inferior de la clase mediana
- a : Amplitud del intervalo de clase
- F_{i-1} : Frecuencia acumulada de la clase que precede inmediatamente a la clase que contiene a la mediana
- f : Frecuencia de la clase mediana

Moda.- Es el valor que se presenta más a menudo, por lo tanto se encuentra en la clase de frecuencia más alta.

Fórmula:

$$Mo = L + a\left(\frac{d_1}{d_1 + d_2}\right)$$

L : Límite inferior de la clase modal

d_1 : Es la diferencia entre la frecuencia de la clase modal y de la clase que le precede $(f_i - f_{i-1})$

d_2 : Es la diferencia entre la frecuencia de la clase modal y de la clase que la sigue. $(f_i - f_{i+1})$

f_i : Frecuencia de la clase modal

Ejemplo Ilustrativo de Medidas de la Tendencia Central a partir de Datos Agrupados

La Empresa "El Batán", dedicada a la Agroindustria, desea conocer sobre el rendimiento laboral de sus trabajadores, para lo cual se tiene una muestra de 30 trabajadores con la siguiente distribución de horas laboradas:

138	40	117	164	129	58
57	157	95	110	46	89
122	132	78	139	160	80
71	101	165	138	48	119
134	15	114	50	70	75

Se pide

1. Elaborar la distribución de frecuencias absolutas y relativas.
2. Calcular la media aritmética, mediana, moda, varianza, desviación estándar y el coeficiente de variación, e interpretar sus resultados.

Desarrollo

Para elaborar la distribución de frecuencias se seguirán los pasos siguientes:

1º Determinando el rango o recorrido:

$$R = Vmáx - Vmín = 165 - 15 = 150$$

2º Número de Clases en las cuales se va a presentar los datos:

$$m = 1 + 3.3 \log(n) = 1 + 3.3 \log(30) = 5.87 \approx 6$$

3º Amplitud de cada una de las clases, las mismas que son de igual tamaño.

$$a = R / m = 150 / 6 = 25$$

Los dos extremos del cuadro de distribución serán:

$L'_i = 15$ y $L'_s = 165$

El cuadro será:

Rendimiento (Horas)	X_i	f_i	F_i	h_i	H_i	h_i %	H_i %
[15 - 40 >	27,50	1	1	0,03	0,03	3	3
[40 - 65 >	52,50	6	7	0,20	0,23	20	23
[65 - 90 >	77,50	6	13	0,20	0,43	20	43
[90 - 115 >	102,50	4	17	0,13	0,57	13	57
[115 - 140 >	127,50	9	26	0,30	0,87	30	87
[140 - 165 >	152,50	4	30	0,13	1,00	13	100
		30		1,00		100	

Interpretación

f_i : Es la frecuencia absoluta del intervalo i, representa al número de observaciones que pertenece a dicho intervalo.

X_i : Es el punto medio del intervalo i que representa a dicha clase

h_i : indica la proporción de número de trabajadores que tienen de 15 a 39 horas de rendimiento

F_i : Es la frecuencia absoluta acumulada "menor que" de la clase i. Número de observaciones menores que el extremo superior del intervalo

H_i : Es la frecuencia relativa acumulada

Rendimiento (Horas)	X_i	f_i	$X_i\ f_i$	$(X_i - X)^2$	$(X_i - X)^2$ fi	
[15 - 40 >	27,50	1	27,50	5136,59	5136,59	
[40 - 65 >	52,50	6	315,00	2178,09	13068,53	
[65 - 90 >	77,50	6	465,00	469,59	2817,53	
[90 - 115 >	102,50	4	410,00	11,09	44,36	Mediana
[115 - 140 >	127,50	9	1147,50	802,59	7223,30	Moda
[140 - 165 >	152,50	4	610,00	2844,09	11376,36	
		30	2975,00		39666,67	

a) Calculando la Media Aritmética:

$X = 2975.00 / 30 = 99.17$ horas

El rendimiento laboral promedio de la empresa "El Batán" es aproximadamente 99 horas.

b) Mediana:
Como $n/2 = 30/2 = 15$, el rendimiento de 90 a 115 horas de los trabajadores corresponde a la clase mediana, por lo que a partir de ella es que realizan los cálculos siguientes:

$$Me = L + a\frac{n/2 - F_{i-1}}{f_i}$$

L	=	90
a	=	25
n	=	30
F_{i-1}	=	13
f_i	=	4

Me	=	102,5 horas

El 50% de los trabajadores tienen un rendimiento menor igual a 102.50 horas y el 50% restante tienen un rendimiento laboral mayor o igual a 102.50 horas.

c) Moda:
El rendimiento de 115 a 140 horas de los trabajadores corresponde a la clase modal, ya que posee la mayor frecuencia, por lo que a partir de ésta se realizarán los cálculos siguientes:

$$Mo = L + a\frac{d_1}{(d_1 + d_2)}$$

L	=	115
a	=	25
d_1	=	5
d_2	=	5
Mo	=	127,5 horas

El rendimiento laboral más frecuente de los trabajadores de la empresa "El Batán" es de 127.50 horas.

Para analizar la dispersión de los datos alrededor del promedio se utiliza la desviación estándar. Previamente se calculará la varianza:

d) Varianza:

$$S^2 = \frac{\sum (X_i - \mu)^2 f_i}{(n-1)}$$

$$S^2 = \frac{\sum 39666.67}{(30 - 1)} = 1367.82$$

e) Desviación Estándar

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2 f_i}{n-1}}$$

$$\sigma = \sqrt{1367.82} = 36.98$$

La dispersión promedio entre los rendimientos laborales de la empresa "El Batán" es aproximadamente 37 horas.

f) Coeficiente de variación

$$\text{CV} = \frac{36.98}{99.17} \times 100 = 37.29$$

Por lo tanto el 37.29% de los rendimientos laborales de la empresa "El Batan" se encuentran dispersos alrededor del valor promedio

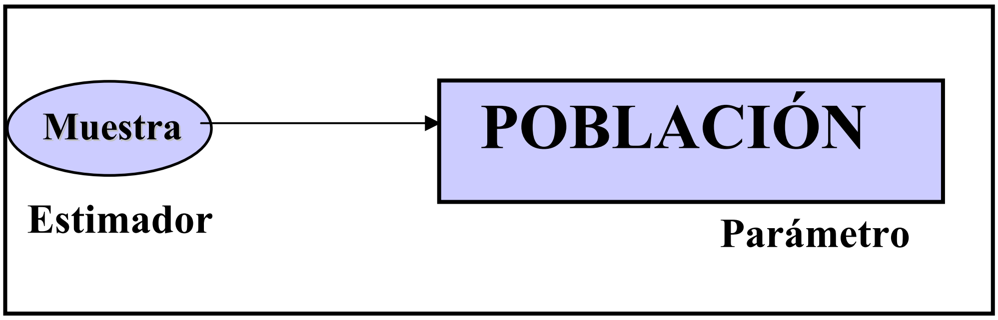
2. LA ESTADÍSTICA INFERENCIAL

Es la utilización de datos muestrales para extraer inferencias o conclusiones sobre la población.

- **Población**.- Es el conjunto de todas las observaciones posibles de un fenómeno. Si la totalidad de ellas es ilimitada, se dice que la población es infinita; en caso contrario, se dice que la población es finita.
  Ejemplo: El ingreso de todos los asalariados del Perú.

- **Muestra**.- Es un subconjunto seleccionado de la población. La muestra debe ser representativa de la población, para que se pueda efectuar inferencias que tengan sentido.
  Ejemplo: El ingreso promedio de todos los asalariados, en base a una pequeña parte de esa población, ya que el estudio de toda la población exigiría demasiado tiempo y dinero.

- **Parámetros**.- Son los indicadores que caracterizan a la población y para cuyo cálculo habría que investigar a todas las unidades que la conforman.
  Ejemplo: El ingreso promedio de todos los asalariados del Perú.



- **Estadígrafos o estadísticos o estimadores.-** Son los indicadores que caracterizan a la muestra.
  Ejemplo: El ingreso promedio de los asalariados de la muestra.

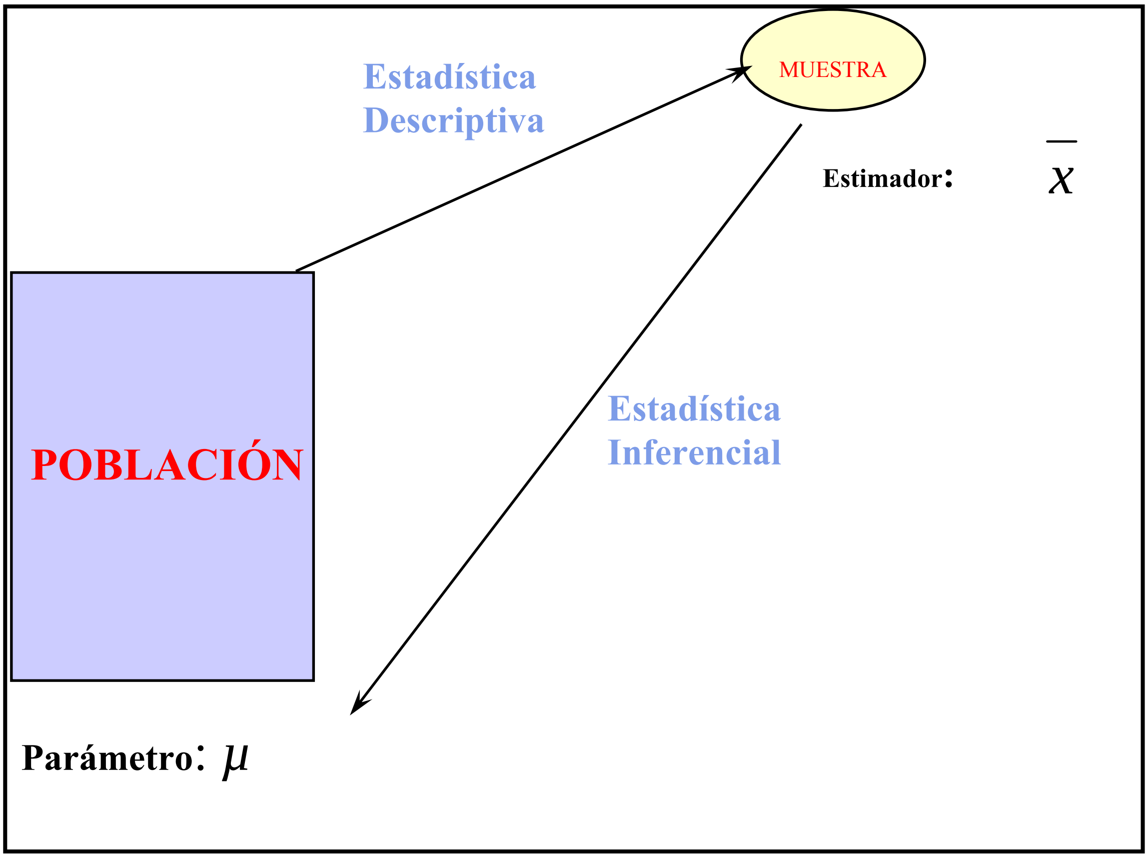
- **Distribución muestral.-** Indica la distribución de los valores que tomará el estimador al seleccionar distintas muestras de la población.
  Ejemplo: Los ingresos promedios de las diferentes muestras.

- **Proceso de inferencia**.- Consiste en establecer la validez de determinadas afirmaciones acerca de los parámetros utilizando un estimador obtenido a partir de una muestra, pero del cual se puede determinar su distribución muestral.

- Los caminos para obtener alguna información sobre los parámetros poblacionales son:

  - **La estimación**.- Consiste en designar un estadígrafo muestral que pueda ser considerado representativo del parámetro.
    Ejemplo: La media muestral de los ingresos de los asalariados sería una estimación de la media poblacional.

  - **Las pruebas de hipótesis**.- Consisten en presumir valores de los parámetros poblacionales y verificar si son coherentes con la información extraída de la muestra.



2.1 Variable Aleatoria

Una Variable Aleatoria es una función que asigna un valor numérico a cada ocurrencia de un experimento. A la lista de todos los posibles valores de X de un experimento y sus correspondientes probabilidades de ocurrencia se llama Distribución o Función de Probabilidad.

Por ejemplo: Sea el experimento aleatorio de averiguar la marca de zapatillas que preferirá un individuo entre las posibles marcas: <<A>>, <<B>>, <<C>><<D>>.

En este caso la asociación de un número para cada suceso elemental posible del experimento no es inmediata. En consecuencia, se establece una correspondencia entre el conjunto de los sucesos elementales posibles y el conjunto de los números reales, del modo siguiente:

Al suceso elemental <<preferir la marca A>> se le hace corresponder el número 1; al suceso elemental <<preferir la marca B>> se le hace corresponder el número 2; al suceso elemental <<preferir la marca C>> se le hace corresponder el número 3; al suceso elemental <<preferir la marca D>> se le hace corresponder el número 4.

La variable aleatoria X será: X = (1, 2, 3, 4).

Media y Varianza de una Variable Aleatoria

Media.- Sea X una variable aleatoria discreta con función de probabilidad p(xi). Entonces, el valor esperado de X (media o esperanza matemática de X), E(X), está definido por:

$$E(X) = \sum_{i} X_i p(X_i)$$

Si $p(X_i)$ es una caracterización exacta de la distribución de frecuencias de la población, entonces $E(X) = \mu$, es la media poblaciónal.

Varianza y Desviación Estándar.- La varianza de una variable aleatoria X está definida como el valor esperado de $(\mu - x)^2$. Es decir:

$$Var(X) = E[(X-\mu)^2] = \sum\left[(X-\mu)^2 p(X_i)\right]$$

La desviación estándar de X es la raíz cuadrada positiva de Var(X).

Si $p(X)$ es una caracterización exacta de la distribución de frecuencias de una población, entonces $E(X) = \mu$, es la media de la poblaación; $VAR(X) = \sigma^2$ es la varianza de la población y σ es la desviación estándar de la población.

Ejercicio Aplicativo
En la empresa Invermax trabajan 14 analistas de inversión. Todos los días se les encarga a cada uno de ellos que evalué de uno a cuatro valores. En un día cualquiera se les dio el siguiente encargo:

Resultado de Xi (Nº de valores)	**Frecuencia de Xi (Nº de analistas)**
1	2
2	3
3	4
4	5

Invermax elaboró una distribución de probabilidad para la variable aleatoria X (Nº de valores asignados a los analistas en ese día.

El cuadro siguiente muestra la distribución de probabilidad de la Variable X:

Resultado de Xi (Nº de valores)	**Frecuencia de Xi (Nº de analistas)**	**P(X = Xi)**
1	2	2/14=0.14
2	3	3/14=0.21
3	4	4/14=0.29
4	5	5/14=0.36

Entonces si se elige un analista al azar, la probabilidad que tenga que evaluar 4 valores es 0.36.

La media, la varianza y la desviación estándar de la variable aleatoria X

Entonces:
La media de X es:

$$\mu = E(X) = \sum_i X_i p(X_i) = (1*2/14) + (2*3/14) + (3*4/14) + (4*5/14) = 2.86$$

La varianza de X es:

$$Var(X) = \sigma^2 = (1-2.86)^2(2/14) + (2-2.86)^2(3/14) + (3-2.86)^2(4/14) + (4-2.86)^2(5/14) = 1.12$$

La desviación estándar de X es:

$$\sqrt{\sigma^2} = \sqrt{VAR(X)} = \sqrt{1.12} = 1.06$$

Entonces, a los analistas se les asignó una media de 2.86 valores, para que sean evaluados y analizados. La desviación estándar, es decir la dispersión respecto a la media es de 1.06.

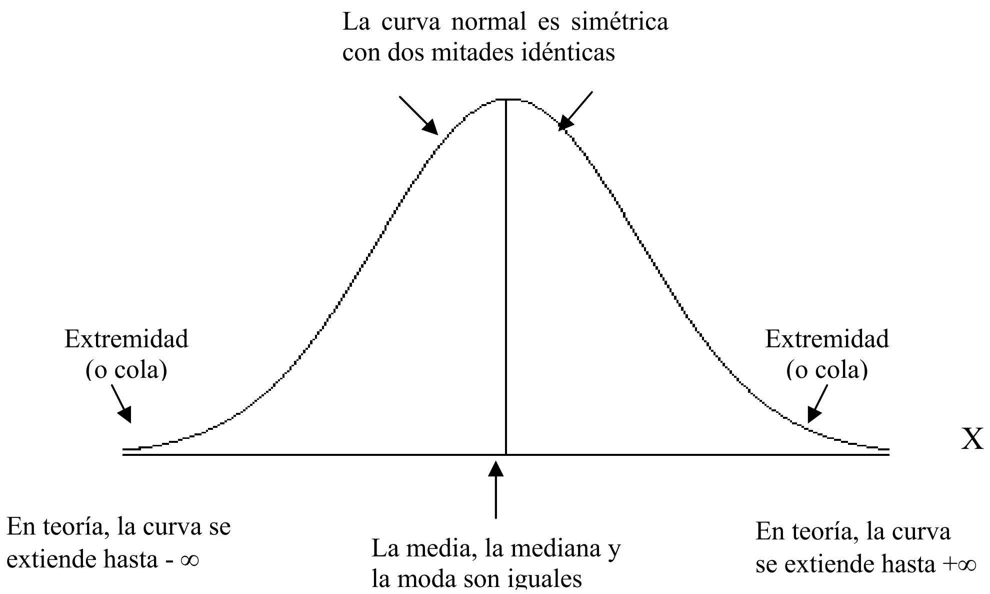
2.2 Distribución Normal

La importancia de la distribución normal se debe principalmente a que hay muchas variables asociadas a fenómenos naturales que siguen el modelo de la normal. Por ejemplo, el consumo de cierto producto por un mismo grupo de individuos.

Se dice que X es una variable aleatoria que tiene una distribución normal con parámetros μ y σ^2, y se escribe: $X \sim N(\mu, \sigma^2)$, y su función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}, \text{ con } -\infty < \mu < +\infty, \quad \sigma > 0, \quad x \in \Re$$

Cada par (μ, σ^2), da lugar a una distribución diferente y cuando se tiene valores dados de μ y σ^2 se determina completamente a la distribución normal de interés.



Características de la distribución normal:

- Es acampanada; y el valor de la media, la mediana y la moda son iguales.
- La distribución probabilística es simétrica con respecto a la media (eje Y)
- La curva normal decrece uniformemente en ambas direcciones a partir del valor central. Además es asintótica, lo que significa que la curva se acerca cada vez más al eje X, pero en realidad nunca llega a tocarlo.
- Es descrita completamente por la media y la desviación estándar.
- Existe una familia de distribuciones normales. Por lo que cada vez que cambian la media o la desviación estándar, se origina una nueva distribución normal.
- La variable asume todos los valores reales, es decir va de $-\infty$ a ∞

Distribución de Probabilidad Normal Estándar

La distribución normal estándar es un caso especial de la distribución normal.

Sea Z una v.a con media 0 y desviación estándar de 1, es decir, $Z \sim N(0,1)$ Entonces Z es una variable con distribución Normal Estándar.
Cualquier distribución normal puede convertirse a una distribución normal estándar mediante la siguiente fórmula:

$$z = \frac{x-\mu}{\sigma}, \quad \text{donde:} \quad \begin{array}{l} X \sim N(\mu, \sigma^2) \\ Z \sim N(0,1) \end{array}$$

Donde:

x : es el valor de cualquier observación específica de la distribución N
μ : es la media de la distribución N
σ^2 : es la desviación estándar de la distribución N

Estandarizando una distribución normal podemos apreciar la distancia en unidades de la media y de la desviación estándar.

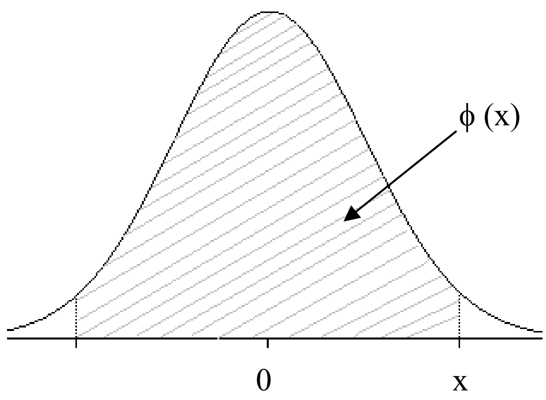
La función de densidad de la distribución de probabilidad normal estándar será:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, x \in \Re ;$$

La función de distribución Acumulada es:

$$\Phi(x) = P[X \leq x] = \int_{-\infty}^{x} f(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

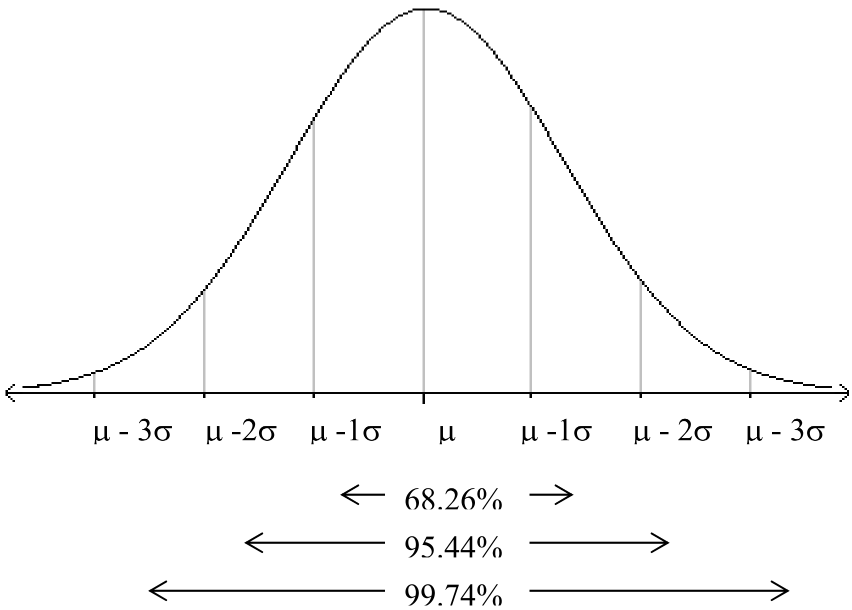
Gráficamente:



Áreas Bajo la Curva

Se tiene que:

- Aproximadamente 68% del área bajo la curva normal está dentro de más una y menos una desviación estándar respecto de la media ($\mu \pm 1\sigma$.)
- Aproximadamente 95% del área bajo la curva normal está dentro de más dos y menos dos desviaciones estándares respecto a la media ($\mu \pm 2\sigma$).
- Prácticamente toda el área (99.74%) bajo la curva normal está dentro de tres desviaciones estándares respecto de la media ($\mu \pm 3\sigma$)



Otras Propiedades:

- Si x tiene distribución N(0,1), entonces para todo x real positivo, se cumple: $\phi(x) = 1 - \phi(x)$
- Si la variable X tiene distribución $N(\mu,\sigma^2)$, entonces la variable Z, definida por $Z = \frac{x-\mu}{\sigma}$, tiene distribución N(0,1) esta propiedad indica lo siguiente: cualquiera que sean los valores de los parámetros de la distribución $N(\mu,\sigma^2)$, puede ser transformada a N(0,1). Según la transformación Z anterior, las distribuciones de probabilidades correspondientes a X pueden ser calculados a partir de la distribución de la variable $z = \frac{x-\mu}{\sigma}$, a la que se denomina variable normal estandarizada. A este proceso de transformación se le denomina "estandarización".
- $E(x) = \mu$ y $Var(x) = \sigma^2$
- Si $X \sim N(\mu,\sigma^2)$, entonces la variable Y= aX + b tiene distribución $N(a\mu + b, a^2\sigma^2)$.

Tabla de la Distribución Normal

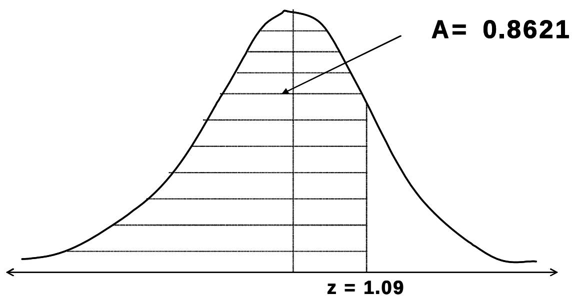
Para valores de x, que varían a intervalos de un centésimo, generalmente desde 0 hasta 3.49, el cuerpo de la tabla presenta valores de $\phi(x)$. Esta tabla se usa de dos maneras:

Uso directo: dado x se halla $\phi(x)$.
Uso inverso: dado $\phi(x)$, se halla x.

X	0.00 0.01 0.02 0.03...............................0.09
0.0	
0.1	.... $\phi(0.11)$= 0.5438
.	
.	
1.0	....$\phi(1.09)$= 0.8621
1.1	.... $\phi(1.13)$= 0.8686
.	
.	
.	
3.4	........................$\phi(3.42)$= 0.9997

Donde:

$\phi(0.11) = 0.5438$, $\phi(1.09) = 0.8621$, $\phi(1.13) = 0.8686$, $\phi(3.42) = 0.9997$



ϕ(1.09)=0.8621

2.3 Distribución de la Variable Aleatoria β_i

Considerando el modelo $Y_i = \beta_1 + \beta_i X_i + \mu_i$, se debe tener presente que el parámetro β_i en la población es estimado a partir de una muestra de observaciones "$\hat{\beta}_i$."

Como la estimación de los β_i se hacen a partir de una muestra de estimaciones su valor diferirá de una muestra a otra que dependerá del grado de aleatoriedad (μ) que venga con la información.

Como la variable aleatoria β_i se distribuye normalmente con promedio β_i y varianza $\sigma^2_{\beta_i}$, entonces:

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2_{\beta_i})$$

Con el propósito de estandarizar la v.a $(\hat{\beta}_i)$ se le resta su valor promedio y se divide entre la desviación estándar de este estimador, ello permite construir una nueva variable aleatoria:

$$Z_1 = \frac{\hat{\beta}_i - \beta_i}{\sigma_{\beta_i}} \sim N(0,1).$$

donde σ_{β_i} es la desviación estándar poblacional.

La construcción de una variable aleatoria con distribución "estadístico t" se obtiene dividiendo una variable aleatoria con distribución normal estandarizada entre una variable con distribución Chi-cuadrado dividido entre su grado de libertad.

$$\frac{Z_i}{\sqrt{\frac{X_n^2}{n}}} \sim t_n$$

Observaciones:

Si $Z_i \sim N(0,1) \Rightarrow \sum_{i=1}^{n} Z_i^2 \sim X_n^2$ grados de libertad

Si $Z_1 ... Z_n$ son v.a. $\sim X_n^2 \Rightarrow \sum_{i=1}^{n} Z_i \sim X^2_{\sum k_i \, G.L.}$

Generalmente no se conoce la desviación estándar poblacional, es por ese motivo que se construye una nueva variable aleatoria "**t**" con $\hat{\sigma}_{\beta_i}$, la cual es obtenida de los datos muestrales. La nueva variable aleatoria será:

$$t = \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_{\beta_i}} \sim t_{n \text{ grados de libertad}}$$

2.4 Pruebas de Hipótesis

Es un procedimiento basado en la evidencia muestral y en la teoría de probabilidad. Se emplea para determinar si la hipótesis es un enunciado razonable y no debe ser rechazada, o si es irrazonable y debe ser rechazada.

Hipótesis.- Son supuestos o enunciados que pueden o no ser verdaderas, relativas a una o más poblaciones, pueden ser:

(a) **Hipótesis nula:** Denotada con H_0, determina supuestos o conjeturas de la población o poblaciones bajo estudio, con el propósito de rechazar. En ella se indica que no hay cambios, que no hay diferencias o se propone un modelo teórico determinado. Por lo común es una afirmación de que el parámetro de población tiene un valor específico.

(b) **Hipótesis alternativa:** Denotado con H_1, determina supuestos o conjeturas de la población o poblaciones bajo estudio con el propósito de no rechazarla. Esta afirmación se aceptará si los datos muestrales proporcionan amplia evidencia de que la hipótesis nula es falsa.

Tipos de Error

Error Tipo I: Se refiere a la probabilidad de rechazar la hipótesis nula, H_0, cuando en realidad es verdadera.

1- α: Se refiere a la probabilidad de no rechazar la hipótesis nula, H_0, cuando en realidad es verdadera.

Error tipo II: Se refiere a la probabilidad de aceptar la hipótesis nula, H_0 cuando en realidad es falsa.

1- β: Se refiere a la probabilidad de rechazar la hipótesis nula, H_0, cuando en realidad es falsa. No se busca maximizarlo por que nunca se va aceptar la H_0.
Lo ilustramos mejor en el siguiente cuadro:

Hipótesis Nula	**El investigador**	
	No Rechaza H_0	**Rechaza H_0**
Si H_0 es verdadera	Decisión Correcta =(1-α)	Error Tipo I = α **Nivel de significación**
Si H_0 es falsa	Error Tipo II = β	Decisión Correcta = (1- β)

Por tanto las probabilidades de error tipo I y tipo II están dadas por las siguientes proposiciones:

α= P (rechazar $H_0 \mid H_0$ es verdadera) $\Rightarrow$ 1-α = P (no rechazar H_o / H_0 verdadera)

β = P (no rechazar $H_0 \mid H_0$ es falsa) $\Rightarrow$ 1-β = P (rechazar H_0 / H_0 es falsa)

Procedimiento de Cinco pasos para Probar una Hipótesis

Paso 1: Plantear Hipótesis Nula y Alternativa

Es el paso inicial en la prueba de hipótesis, en donde se busca corroborar el cumplimiento de ciertas características específicas:

Por ejemplo si los parámetros coinciden con unos valores "C" que se consideran como referencia del estudio a probar un caso particular de la dócima:

$H_0 : \beta_i = C$ Hipótesis nula
$H_1 : \beta_i \neq C$ Hipótesis alternativa

Otro ejemplo podría ser si se esta interesado en averiguar si las variables que intervienen en el modelo de regresión lineal son explicativas del mismo, en este caso será

$H_0 : \beta_i = 0$ Hipótesis nula
$H_0 : \beta_i \neq 0$ Hipótesis alternativa

Paso 2: Seleccionar un Nivel de Significación

Consiste en la elección del riesgo que se asume si se rechaza la hipótesis nula cuando en realidad debe aceptarse por ser verdadera, por lo general se considera el 1% ó 5%.

Actualmente los software que nos permiten realizar pruebas de hipótesis consideran un "p value", el cual es el valor que nos permite rechazar o no la hipótesis nula H_0 sin necesidad de recurrir al uso de tablas, si su valor se encuentra por debajo del nivel de significación escogido (1% ó 5%) se rechazará la hipótesis nula.

Paso 3: Identificación y Obtención del Estadístico de Prueba

Los estadísticos más usados en las pruebas de hipótesis son la normal estandarizada"**Z**", el estadístico "**t**", el Chi cuadrado "χ^2 "y el "**F**".

El estadístico de prueba**,** es un valor determinado a partir de la información muestral y que tiene una distribución característica, es utilizado para determinar los puntos críticos partir de los cuales se rechazará o no la hipótesis nula, un ejemplo es el llamado valor Z (o desvío normal), que se determina a partir de datos muestrales:

$$z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Donde:

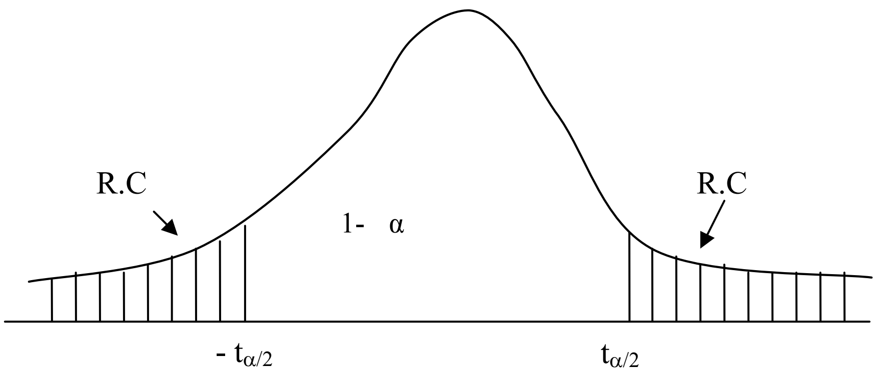
σ : Desviación estándar de la población
n : Número en la muestra
$\overline{X}$: Media muestral
μ : Media poblacional

Para tamaños de muestra n < 30, no se necesita un análisis preliminar del valor estadístico de la variable, pero para un tamaño de muestra n > 30, si se necesitaría un análisis preliminar de las variables.

Paso 4: Formular una Regla de Decisión

Consiste en comparar el valor del estadístico calculado con el de la tabla con un nivel de significación α y un grado de libertad (n-k), que depende del número de observaciones y del número de parámetros.

Ya calculado el valor del estadístico y escogido el nivel de significancia el paso que sigue es determinar la región critica (RC.) pues es a partir de ella que se rechaza a hipótesis nula. Para el caso de la distribución "t", por ejemplo:



En donde el $1-\alpha$ es la zona de aceptación de hipótesis nula, con lo cual su región crítica será:

$$RC = \{ |t_c| \geq t_{n-k,\alpha/2} \}$$

Para el ejemplo el estadístico "**t_c**" obtenido fue 2.34, el cual se encuentra en la región de rechazo, es decir, es mayor a 1.645 ($t_{\alpha/2}$), por lo tanto la hipótesis nula es rechazada.

Paso 5: Toma de una Decisión

Es el resultado final de la prueba de hipótesis donde se Acepta H_0, o bien se rechaza H_0 y se acepta H_1

2.5 Pruebas para la Media Poblacional

Muestra grande conociendo la desviación estándar de la población

Para el caso en que se tiene una muestra grande conociendo la desviación estándar de la población, se tendrá que el valor del estadístico de prueba posee una distribución aproximadamente normal, con lo cual:

$$\boxed{z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}} \sim N(0,1)$$

H_0	H_1	RC
$\mu=\mu_o$	$\mu=\mu_1$ ($<\mu_o$)	$Z<-z_{1-\alpha}$
	$\mu \neq \mu_1$ ($>\mu_o$)	$Z> z_{1-\alpha}$
	$\mu\neq \mu_o$	$Z<-z_{1-\alpha/2}$ ó $Z> z_{1-\alpha/2}$ $\lvert Z\rvert> z_{1-\alpha/2}$
$\mu\geq\mu_o$	$\mu<\mu_o$	$Z<-z_{1-\alpha}$
$\mu\leq\mu_o$	$\mu>\mu_o$	$Z> z_{1-\alpha}$

Ejemplo Ilustrativo

Un comprador de ladrillos cree que la calidad de los ladrillos está disminuyendo. De experiencias anteriores, la resistencia media al desmoronamiento de tales ladrillos es 200 Kg, con una desviación típica de 10 Kg. Una muestra de 100 ladrillos arroja una media de 195 Kg. Probar la hipótesis de que la calidad media no ha cambiado, contra la alternativa que ha disminuido.

Solución:

1. Planteamiento de hipótesis H_0: $\mu = 200$kg. y $H_1 : \mu < 200$ kg.

2. Se escoge un nivel de significación $\alpha = 0.05$

3. Identificación del tipo de prueba, en este caso la estadística de prueba es para la $\overline{X}$, por lo tanto se hará uso de la distribución normal estandarizada.

4. Desde que la muestra es grande n = 100, la distribución de $\overline{X}$ es:

$$N\left(200, \frac{10^2}{100}\right) = N\ (200,\ 1)$$

5. Cálculo de la región crítica (R.C.)

R .C. =$<-\infty, \overline{x}_C>$ donde $\overline{x}_c$ es tal que $P\left[\overline{X} < \overline{x}_c / H_o\right] = \alpha$

ó $$P\left[\frac{\overline{X} - \mu_O}{\sigma/\sqrt{n}} < \frac{\overline{x}_c - 200}{10/10}\right] = P\ \left[Z < \overline{x} - 200\right] = 0.05$$

Donde $Z_\alpha = \overline{x}_c - 200 = -1.64$, luego $\overline{x}_c = 198.36$

R.C. = $< -\infty, 198.36>$

6. Cálculo de media muestral: Según el enunciado para una muestra de n = 100, se tiene $\bar{x}$ = 195.

7. Conclusión: Puesto que $\bar{x}$ =195 $\in$ R.C. = < - ∞, 198.36>. Rechazamos H_0, por lo que se concluye que la calidad media de los ladrillos ha disminuido.

Muestra Grande no Conociendo la Desviación Estándar de la Población

Para el caso en que se tiene una muestra grande desconociendo la desviación estándar de la población, suponiendo que la población tiene distribución aproximadamente normal, usamos el estadístico de prueba T, el cual tiene una distribución "t-student", con lo cual :

$$\boxed{T = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}} \sim t_{(n-1)}$$

H_0	H_1	RC
$\mu=\mu_o$	$\mu=\mu_1$ ($<\mu_o$)	$T<-t_{1-\alpha}$
	$\mu=\mu_1$ ($>\mu_o$)	$T> t_{1-\alpha}$
	$\mu\neq\mu_o$	$T<-t_{1-\alpha/2}$ o $T> t_{1-\alpha/2}$
$\mu\geq\mu_o$	$\mu<\mu_o$	$T<-t_{1-\alpha}$
$\mu \leq\mu_o$	$\mu>\mu_o$	$T> t_{1-\alpha}$

Ejemplo Ilustrativo

Una máquina para enlatar conservas de pescado ha sido regulada para que el contenido de una lata sea 16 onzas. Usando α = 0.05, ¿Diría usted que la máquina ha sido adecuadamente regulada, si una muestra de 20 latas dio un peso medio de 16.05 onzas y una desviación típica de 1.5 onzas?.

Solución:

1. Planteamiento de hipótesis H_o: μ = 16 y H_1: $\mu \neq 16$

2. Escogiendo un nivel de significación α = 0.05

3. Puesto que n = 20 es pequeño y suponiendo que la población tiene distribución aproximadamente normal, usamos la variable aleatoria.

$T = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$ que tiene una distribución t con n – 1 = 19 grados de libertad como estadística de prueba.

4. Región Crítica (RC): $T < - t_{\alpha/2} = - 2.093$ ó $T > t_{\alpha/2} = 2.093$, con $\alpha/2 = 0.025$, para buscar en la tabla tomamos

$1 - \alpha/2 = 0.975$ luego, R.C. = < - 2.093, 2.093>

5. De los datos $\bar{x} = 16.05$, s = 1.5 para n = 20, entonces

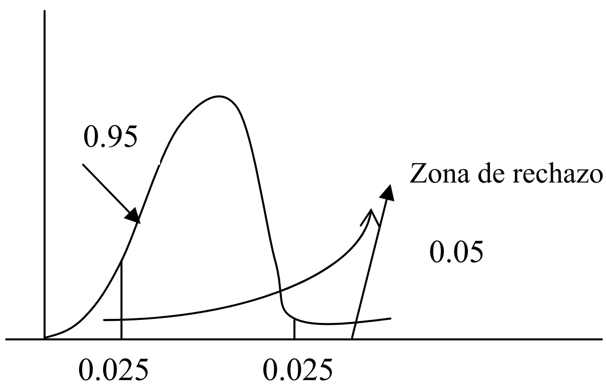
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{16.05 - 16}{1.5/\sqrt{20}} \cong 0.148$$

6. Conclusión: desde que t = 0.148 ∈ R.C, aceptamos H_0; es decir se acepta que la máquina ha sido adecuadamente regulada.

2.6 Consideraciones para el Uso de la Tabla Chi – Cuadrada

$$\hat{\beta}_i \sim N(\beta_i, \sigma_R^2)$$

$$\boxed{z_1 = \frac{\hat{\beta}_i - \beta_i}{\sigma_\beta} \sim N(0,1)}$$



Como la distribución Chi - cuadrada tiene asimetría positiva y crece con los grados de libertad, la Chi – cuadrado se aproxima a una distribución normal si los grados de libertad son mayores a 40. Asimismo se puede calcular la probabilidad aproximando a una normal.

Ejemplo

1)

$$G.L. = 15$$
$$P(X_0 > X) = 0.95 \quad \text{probabilidad de rechazo}$$
$$P(X_0 < X) = 0.05 \quad \text{probabilidad de aceptación}$$

$$X_i^2 > \chi_{tabla}^2$$

Ejercicio de autoconocimiento

¿Porqué hacer una predicción empresarial?

	SI	NO	NO SÉ
1. Porque tendré éxito en la gestión empresarial			
2. Porque permite dar un buen pronóstico de las variables de decisión			
3. Para analizar el pasado y predecir el futuro de la empresa			
4. Para ejecutar los proyectos con eficacia			
5. Para una buena toma de decisiones económicas			
6. Para aprender de los éxitos y fracasos anteriores			
7. Porque la estadística es útil para el administrador			
8. Para cumplir una meta trazada			
9. Para resolver problemas y solucionarlos			
10. Para tomar precauciones al usar la predicción			

CALIFICACION

Puntuar con un punto cada respuesta "SI".
Si obtienes de 1 - 3 puntos tienes pocas expectativas de hacer una buena predicción empresarial.
Si tienes entre 4 – 7, tienes buenas expectativas de hacer una buena predicción empresarial.
Y si tienes entre 8 – 10, denotas excelentes expectativas de hacer una buena predicción empresarial.

RESUMEN

La predicción como una técnica estadística define cuál de los posibles valores futuros de la variable es más probable. Dicho de otro modo: se asigna una probabilidad de ocurrencia a cada uno de los posibles valores de la variable objetivo, con lo que se podrá hacer una lista priorizada sobre la base de su factibilidad probabilística.

Los factores que influyen en el funcionamiento de una empresa que van a servir como fuentes de predicción estratégica empresarial se agrupan de la siguiente manera: económicos, sociales, políticos y tecnológicos.

Para construir un Modelo se deben seguir las siguientes etapas:

Conocimiento de la Teoría Económica
Especificación del Modelo
Estimación de los parámetros
Verificación
Predicción o Pronóstico

La técnica estadística conocida como "Análisis de Regresión" es la herramienta principal utilizada para obtener los valores estimados.

La estadística descriptiva es el proceso de recopilación, organización y presentación de un conjunto de datos

La estadística inferencial es la utilización de datos muestrales para extraer inferencias o conclusiones sobre la población.

EXPLORACIÓN ON LINE

1. Técnicas estadísticas de predicción aplicables al campo empresarial de Cortez Cumpa. Universidad del Pacífico. 1994.Tarwi. Lanmolina.edu.pe
   http: www.portalagrario.gob.pe

2. Corvu
   Obtiene un módulo de predicción
   http: icomred.com.pe/corvuhtm-41k

3. Ayuda memoria
   Mayor dinámica por la búsqueda de su sostenibilidad empresarial
   http: www.editoraperu.com.pe/ep 2003

LECTURA

IMPORTANCIA DE LA ESTADISTICA

Casi todos los campos de la investigación científica seria se pueden beneficiar del análisis estadístico. Los responsables de la toma de decisiones sobre política económica, asesores del presidente y de otros altos cargos públicos, tienen en la estadística una herramienta muy valiosa.

Únicamente con ayuda del análisis estadístico pueden tomarse decisiones inteligentes en relación con los tipos tributarios, los programas sociales, los gastos de defensa y muchas otras cuestiones.

Es fundamental también para los empresarios, en su búsqueda incansable del beneficio. Las actividades de control de calidad, minimización de costes, combinación de productos y existencias y multitud de otros aspectos empresariales se pueden gestionar con eficacia mediante procedimientos estadísticos contrastados.

En la investigación de mercados, la estadística representa una ayuda inestimable para determinar si es probable que un nuevo producto tenga éxito.

Su utilidad es evidente también para asesores financieros que han de evaluar las oportunidades de inversión. Contables, directores de personal y fabricantes se benefician igualmente del análisis estadístico. Incluso el investigador médico, preocupado por la eficacia de un nuevo medicamento, encuentra en la estadística un aliado imprescindible.

Fuente: Webster (1998)

ACTIVIDADES

1. Describa con sus propias palabras el proceso de predicción.

2. ¿Cuál es la diferencia entre la estadística descriptiva y la inferencial?

3. ¿Cuál es la importancia de las Técnicas Estadísticas de la Predicción?

4. ¿Cuál es la herramienta principal utilizada para obtener los valores estimados?

5. Caso hipotético, es una encuesta realizada a 100 asistentes a un centro comercial, para determinar el número de veces al mes que acude a realizar compras a ese lugar. La pregunta No 1 fue: ¿Cuál es el número de veces (promedio) al mes que acuden a comprar en este centro comercial?; los entrevistados respondieron:

Nº de veces	Frecuencia absoluta	Frecuencia relativa	Frecuencia acumulada	Frecuencia relativa acumulada
3	14			
4	30			
5	25			
6	18			
7	8			
9	3			
10	1			
11	1			
Total	100			

   Hallar la frecuencia relativa, acumulada y relativa acumulada.

6. El emplazamiento en el cobro de las últimas cien ventas facturadas por un establecimiento comercial se había agrupado en cuatro intervalos, recordándose solo los siguientes datos de la distribución:
   a. El primer intervalo tiene 6 semanas como extremo superior, una frecuencia relativa de 0.2 y una amplitud de 4 semanas.
   b. En el segundo intervalo se acumulan 60 ventas
   c. Las marcas de clase del segundo y cuarto intervalo son 8 y 50 semanas, respectivamente.
   d. El tercer intervalo presenta una frecuencia de 30 ventas y amplitud de 30 semanas.

   Con esta información reconstruya la distribución de frecuencias y represente el histograma correspondiente.

7. En un convenio laboral se acuerda subir un 10% el volumen total de salarios. Un empresario con 250 empleados les paga un total de 200,000 soles mensuales. La subida del 10% representa para el empresario un incremento total de las nóminas del personal de 20,000 mensuales. Dicha subida puede ser:
   a. Proporcional; Se aumenta el sueldo de cada empleado en un 10%.
   b. Lineal; Cada empleado percibe un aumento de 80 soles

   ¿Qué alternativa conduce a reducir las diferencias salariales?

8. Una empresa de hospedería presenta la siguiente distribución de salarios por hora:

Salario por Hora (Soles)	Nº de empleados
5-8	10
8-10	15
10-12	10
12-15	5
15-18	5

Si la empresa tiene 45 trabajadores obtenga:
   a. El salario más frecuente y el salario medio por hora
   b. Calcule e interprete una medida de asimetría de la distribución de salarios por hora en la empresa
   c. En 1998 debido a la crisis del sector por el Fenómeno de "El Niño", la dirección de la empresa reduce en un 20% los salarios. Calcule el salario medio tras dicha reducción sin calcular la nueva distribución.
   d. Suponga ahora que la crisis fue de tal magnitud que en vez de reducir los salarios la decisión que se toma finalmente es la reducir el número de empleados de cada intervalo en un 20%. ¿Cuál es el salario medio? ¿Y si se redujeron los salarios y el número de empleados simultáneamente?

9. Se ha encuestado a 1000 establecimientos de comercio minorista de la ciudad de Lima por sus ventas anuales de un cierto producto alimenticio. Los resultados obtenidos son los que refleja la tabla siguiente:

Ventas (Soles)	Establecimientos
Hasta 600	400
De 600 a 1,500	225
De 1,500 a 3,000	175
De 3,000 a 6,000	120
De 6,000 a 9,000	75
Mas de 9,000	5

Además se sabe que el total de ventas para ese producto ascendió a 2 millones de soles.
   a. Obtenga la media aritmética de esa variable. Analice su representatividad en términos absolutos y relativos. Observando la distribución de frecuencias, ¿considera que es el promedio más adecuado?
   b. Determine el porcentaje de establecimientos, de entre los que menos venden, cuyas ventas acumuladas representan la cuarta parte de las ventas totales. Determine también el valor de la variable del 10% de los establecimientos que más venden.
   c. Si al año siguiente el precio de ese producto aumenta en 5%, determine cual sería el nuevo volumen medio de ventas así como su desviación estándar.

10. Intradevco S.A., reconocida empresa nacional que produce una gama muy variada de productos para el aseo, ha registrado desde la aparición de la pasta dental DENTO en todas sus modalidades, la duración de este dentífrico en los

hogares (expresado en días). Se han obtenido que en promedio duran alrededor de 25 días con una desviación estándar de 5 días.

Si se sabe que Intradevco Industrial S.A vende a un distribuidor 30,000 pastas para los dientes, asegurándole que los valores que ha obtenido son fiables. Suponga que la duración de las pastas para dientes sigue una distribución normal.

a. ¿Cuál es la probabilidad que duren más de 20 días?
b. ¿Cuál es la probabilidad de que duren entre 22 y 32 días?

11. En una fábrica de conservas de fruta se desea verificar si las latas resultan en promedio con un peso no inferior a 1 kg. Se sabe que el tamaño de la fruta puede contribuir con una variación en los pesos de las latas de manera que estos se distribuyan normalmente con una dispersión del 8%. Se efectúa una muestra de 100 latas en la que se determina los pesos, resultando la $\overline{X} == 980$ gr. Deseamos saber si la muestra comprueba o rechaza la Hipótesis inicial planteada. Fijamos un coeficiente de riesgo igual al 5%.

12. Se conoce por investigaciones ya realizadas que el 20% de la población mayor de 15 años fuma. Después de efectuar una fuerte campaña televisiva y radial durante 6 meses, se decide estudiar si la población adulta ha disminuido el hábito de fumar.

    Para ello se selecciona una muestra aleatoria de 1000 personas adultas a las que se somete a una determinada encuesta. Resumida la información proporcionada por el trabajo de campo, se observó que el 12% de las personas encuestadas fumaba habitualmente. Probar la hipótesis de que la campaña publicitaria ha disminuido la cantidad de fumadores

13. Telefónica ha determinado que la duración media mensual de llamadas locales en Lima ha sido de 150 minutos, con una desviación típica de 15 minutos, además ha observado que su duración mensual es una variable que sigue una distribución normal. Sabiendo esto ¿Cuál será la probabilidad de que una comunicación telefónica cualquiera dure entre 150 y 180 minutos?

AUTOEVALUACIÓN

Encierra en un círculo la letra que contenga la alternativa correcta.

1. La predicción empresarial es una técnica estadística que define cuál de los posibles valores futuros de la ______________ es más probable.

   a. Constante
   b. Variable
   c. Parámetro
   d. Investigación

2. El objetivo del proceso de predicción es dar al administrador de la empresa las _______________________ que pueda pronosticar las principales variables.

   a. Las fuentes de datos utilizados
   b. La estimación de parámetros
   c. Las técnicas estadísticas de la predicción
   d. Etapas del modelo

3. La media aritmética. La mediana, la moda son:

   a. Estadísticos de dispersión
   b. Modelos
   c. Estimadores
   d. Estadísticos de tendencia

4. La amplitud de rango, rango intercuartílico, la varianza y la desviación estándar son:

   a. Estadísticos de dispersión
   b. Modelos
   c. Estimadores
   d. Estadísticos de tendencia

5. Población es a parámetro como muestra es a :

   a. Estadístico
   b. Inferencia
   c. Hipótesis
   d. Verificación

RESPUESTAS DE CONTROL

1. b, 2. c, 3. d, 4. a, 5. a