1. IDENTIFICACIÓN DE

INSTITUTO TECNOLÓGICO
SUPERIOR JAPÓN

GUÍA METODOLÓGICA DE ESTADÍSTICA

COMPILADO POR:
DARWIN ESPÌN
CARRERA DE ADMINISTRACIÒN
2019

AMOR AL CONOCIMIENTO

www.itsjapon.edu.ec

Nombre de la Asignatura: **ESTADÍSTICA**	**Componentes del Aprendizaje**	

Resultado del Aprendizaje:

COMPETENCIAS Y OBJETIVOS

- Aplica los fundamentos estadísticos de acuerdo a sus necesidades.
- Conoce los métodos tabulares y gráficos para manipular información.
- Calcula e interpreta la información calculada, aplicando procedimientos referentes a variables, cualitativas, cuantitativas y agrupándolos de acuerdo a su necesidad.
- Analiza la dependencia de las variables estadísticas, mediante la aplicación de modelos lineales, acorde a los fenómenos estudiados económicos y/o financieros.

Docente de Implementación:

DARWIN DANIEL ESPÍN SALAS	**Duración:** 25 horas

Unidades	Competencia	Resultados de Aprendizaje	Actividades	Tiempo de Ejecución
Conceptos básicos, importancia. Métodos estadísticos. (Estadística descriptiva e inferencial) Tablas de frecuencias. Gráficos estadísticos (Histogramas y polígonos). La investigación Estadística	Conoce, analiza e interpreta las características que definen el razonamiento Estadístico y el uso adecuado de los métodos estadísticos.	Interpreta de manera correcta las técnicas de elaboración de la Estadística Básica por medio de distribuciones de frecuencias, gráficas y las medidas para aplicarlos en procesos administrativos empresariales.	Desarrollo de ejercicios	**5**

Medidas de Tendencia Central. Medidas de posición. (Media, mediana, moda, los cuartiles, deciles y percentiles)	Aplicación e identificación de uso de las Fórmulas de cálculo de las medidas de posición y dispersión.	Valora la importancia de recopilar, procesar e interpretar la información.	Desarrollo de trabajo Experimental Exposición.	**5**
Medidas de Tendencia Central. Medidas de Dispersión (Rango o amplitud, Desviación Estándar, Varianza, Coeficiente de variación) Muestra	Aplicación e identificación de uso de las Fórmulas de cálculo de las medidas de posición y dispersión. Conoce y utiliza de manera correcta las metodologías y parámetros adecuados para el cálculo de la muestra.	Interpreta las tendencias por medio de las medidas de dispersión, de tendencia externa y de sesgo mediante el uso de estadígrafos. Mediante la identificación de la muestra, conoce la importancia de la simplificación poblacional para realizar una investigación.	Desarrollo de trabajo Experimental Exposición.	**5**

Números Índices	Identifica los procedimientos que se utilizan para conocer la variabilidad de los precios, cantidades y costos.	A través de coeficientes interpreta la evolución de cantidades, valores y precios para el análisis comparativo en procesos de compra venta.	Desarrollo de trabajo Experimental Exposición.	**5**
Regresión y Correlación y las series de Tiempo	Conoce los modelos lineales, que se deben utilizar acorde a los fenómenos estudiados económicos y/o financieros.	Mediante el análisis de regresión y correlación describe la relación entre las diversas variables de archivo histórico, para proyectar próximas relaciones económicas empresariales.	Desarrollo de trabajo Experimental	**5**

2. CONOCIMIENTOS PREVIOS Y RELACIONAD

Co-requisitos

- Conocer las operaciones matemáticas básicas (suma, resta, multiplicación, división, raíz cuadrada).
- Uso de la calculadora.

3. UNIDADES TEÓRICAS

- **Desarrollo de las Unidades de Aprendizaje (contenidos)**

A. Base Teórica

LA ESTADÍSTICA Y DESCRIPCIÓN DE DATOS.

Conceptos básicos, importancia.

Estadística.- es la ciencia que trata de la recolección, organización, presentación, análisis e interpretación de datos numéricos con el fin de realizar una forma de decisión más efectiva.

Población.- conjunto sobre el que estamos interesados en buscar conclusiones. Población Finita y Población Infinita.

Muestra.- subconjunto de la población al que tenemos acceso.

Muestra Aleatoria.- muestra representativa de la población; y cada elemento tiene la misma oportunidad de formar parte de la muestra.

Variable.- característica observable que varía entre diferentes individuos de la población.

La información de cada individuo es resumida en variable.

Dato.- valor particular de la variable**.**

Parámetro.- función definida sobre los valores numéricos de características medibles de una población.

Estadístico.- función definida sobre los valores numéricos de una muestra.

Caracteres.- propiedad, rasgos o cualidades de los elementos de la población.

Cualitativos y Cuantitativos.

Modalidades.- diferentes situaciones posibles de un carácter. Exhaustivas y mutuamente excluyentes.

Clases.- conjunto de una o más modalidades, en el que se verifica a que modalidad pertenece a una y solo una clase.

Variable estadística.- es una representación a través de números u otros símbolos, de una variable.

Variables Cualitativas.- son todas aquellas que expresa atributos.

Indiquemos con E = {1,2, 3,…,k} una posible codificación de las observaciones realizadas y sea m, el número de individuos que tienen resultado i = 1,….,k, al mismo que se denomina frecuencia absoluta fi. Se indica fr frecuencia relativa dada por fr = mi/m.

Variables Cuasi Cuantitativa.- son las que demuestran orden o jerarquía. (Ejem: deserción escolar-Alto, Medio y Bajo).

Variables Cuantitativa.- son todas aquellas que expresa cantidades numéricas, cuyos valores numéricos tienen su importancia.

Dentro de estas tenemos las variables Discretas y Continuas.

Variables Discretas.- Aquellas a las que se les puede asociar un número entero, es decir, aquellas que por su naturaleza no admiten un fraccionamiento de la unidad, por ejemplo, número de hermanos, páginas de un libro, etc.

Variables Continuas.- Aquellas que no se pueden expresar mediante un número entero, es decir, aquellas que por su naturaleza admiten que entre dos valores cualesquiera la variable pueda tomar cualquier valor intermedio, por ejemplo, peso, tiempo. etc.

Censo.-corresponde a un listado de una o más características de todos los elementos de la población.

Encuesta.- corresponde a un listado de una o más características de todos los elementos de una muestra.

Métodos Estadísticos

La estadística para su mejor estudio se ha dividido en dos ramas las cuales son: estadística descriptiva y estadística inferencial.

Estadística Descriptiva.- consiste en la presentación de datos en forma de tablas y gráficas. Esta comprende cualquier actividad para resumir o describir los mismos factores pertinentes adicionales, esto se refiere a no intentar nada que vaya más allá de los datos.

Estadística Inferencial.- se deriva de las observaciones hechas solo a una parte de un conjunto numeroso de elementos; implicando así que su análisis requiera de generalizaciones que van más allá de los datos, como consecuencia la característica más importante del crecimiento de la estadística ha sido un cambio en el énfasis de los métodos que sirven para generalizarlas. En otras palabras, la estadística inferencial investiga y analiza una población partiendo de una muestra tomada.

Estadísticas Primarias.- son aquellos datos obtenidos ya sean por encuestas directas, mediante la utilización de cuestionarios, o como resultado de la observación directa; esta utilización es una técnica de estudio de carácter científico.

Estadísticas Secundarias.- En éstas los datos se obtienen de publicaciones, los cuales pueden ser reproducidos totales o parciales. Son fuentes valiosas utilizadas en cualquier tipo de investigación.

Estadísticas Temporales.- Denominadas series de tiempo o series cronológicas. Son las obtenidas y ordenadas en forma cronológica, siendo el resultad de investigaciones u observaciones periódicas: días, meses, años. Cuando las investigaciones son aisladas, es decir, no presentan periodicidad continua, las estadísticas son atemporales.

Tabla de Frecuencias:

Una herramienta que sirve para ordenar y organizar datos estadísticos, entre estos tenemos Datos no agrupados y datos agrupados.

Estructura de una tabla de frecuencias.

X_i: Valores que tiene la variable en el conjunto de datos.

Frecuencia Absoluta n_i.- de un valor Xi es el número de veces que el valor está en el conjunto (X1, X2,…, XN).

Frecuencia Absoluta acumulada N_i: de un valor Xi del conjunto (X1, X2,…, XN) es la suma de las frecuencias absolutas de los valores menores o iguales a Xi, es decir:

Frecuencia Relativa (fi=ni/N).- de un valor Xi es la proporción de valores iguales a Xi en el conjunto de datos (X1, X2,…, XN).

Frecuencia Relativa Acumulada (Fi=Ni/N).- de un valor Xi como la proporción de valores iguales o menores a Xi en el conjunto de datos (X1, X2,…, XN).

EJEMPLO:

Los siguientes datos son los lapsos, en minutos, necesarios para que 50 clientes de un banco comercial lleven a cabo una transacción bancaria:

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

Construya un diagrama de tallo y hoja

Solución. Se trata de una variable aleatoria continúa puesto que el tiempo de duración (en nuestro caso el lapso de tiempo para efectuar una transacción bancaria) se define en general sobre la recta real no negativa esto es $x_i \geq 0$

Su representación mediante un diagrama de tallo y hoja ordenada es

Tallo	Hoja	Frecuencia
0	124455789	9
1	12233445689	11
2	34457889	8
3	133478	6
4	2467	4
5	2568	4
6	238	3
7	268	3

8		0
9	57	2

En nuestro ejemplo en ***número de clases*** que se puede elegir es de ***k*** = 10, es decir se toma en cuenta el número de tallos descritos en el diagrama Tallo y Hoja.

Para determinar el número de clases ***k*** se puede utilizar la ***fórmula de Sturgues*** (Fórmula empírica)

$$k = 1 + 3.322 \ln(n)$$

donde n es el número de observaciones, generalmente la fórmula se aplica cuando n>50. se establecen las clases de forma que tengan las mismas diferencias entre ellas y el número de clases aumenta en función de n.

Nota. En la práctica se tienen buenos resultados si se hace la selección del número de clases considerando la raíz cuadrada de n.

Cada clase está definida por dos valores, estos valores constituyen los ***límites reales*** de las clases. El límite real superior de una clase es el límite inferior de la siguiente.

La diferencia entre los límites reales de una clase constituye el ***intervalo de clase.*** Se llama ***marca de clase o punto medio*** al valor correspondiente al punto central del intervalo.

Cuando tenemos ya determinado las clases clasificamos y contamos los individuos incluidos en cada clase. El número resultante se denomina ***frecuencia absoluta*** de la clase respectiva. El número de i9ndividuos de una clase se puede expresar también mediante su ***frecuencia relativa,*** bien en forma de proporción (cociente entre la frecuencia absoluta de esa clase y el número total de individuos de la muestra) o bien en forma de porcentaje (frecuencia referida a 100 individuos de la muestra).

Con los datos del problema construiremos una tabla de ***distribución de frecuencias.***

Reglas generales para construir una tabla de distribución de frecuencias:

1. Determinar el mayor y menor entre los datos registrados para luego calcular el rango (diferencia entre el valor mayor y el menor)

   Rango = 9.7 – 0.1 = 9.6

2. Dividir el rango en un número conveniente de intervalos de clase del mismo tamaño (igual longitud). Si esto no es posible, entonces utilizar intervalos de diferente tamaño. El número

de clases que se emplea para clasificar los datos depende del total de observaciones. Como anteriormente se dijo, si el número de observaciones es relativamente pequeño, la experiencia muestra que el número de clases a emplear es generalmente mayor o igual que a 5. Si existe una cantidad grande de datos, el número de clases debe encontrarse entre 8 y 12 y generalmente no existirán más de 15 o 20 clases. Para este problema se decidió utilizar 10 clases (intente aplicar la fórmula de Sturgess o utilice la raíz cuadrada de n). entonces la ***longitud de cada clase*** será aproximadamente 1, pues 9.6/10 = 0.96 = 1

3. Determinar el número de observaciones que caen en cada clase (recuento de datos). La tabla de frecuencias que se obtienen es:

No *i*	Clase	Límite Inferior	Límite Superior	Punto Medio (*xm*)	Frec. de Clase (*fi*)	Frec. Relativa (*fr*)	Porcentaje %
1	(0 - 1]	0	1	0,5	9	0,18	18
2	(1 - 2]	1	2	1,5	11	0,22	22
3	(2 - 3]	2	3	2,5	8	0,16	16
4	(3 - 4]	3	4	3,5	6	0,12	12
5	(4 - 5]	4	5	4,5	4	0,08	8
6	(5 - 6]	5	6	5,5	4	0,08	8
7	(6 - 7]	6	7	6,5	3	0,06	6
8	(7 - 8]	7	8	7,5	3	0,06	6
9	(8 - 9]	8	9	8,5	0	0	0
10	(9 -10]	9	10	9,5	2	0,04	4
	Total				50	1	100

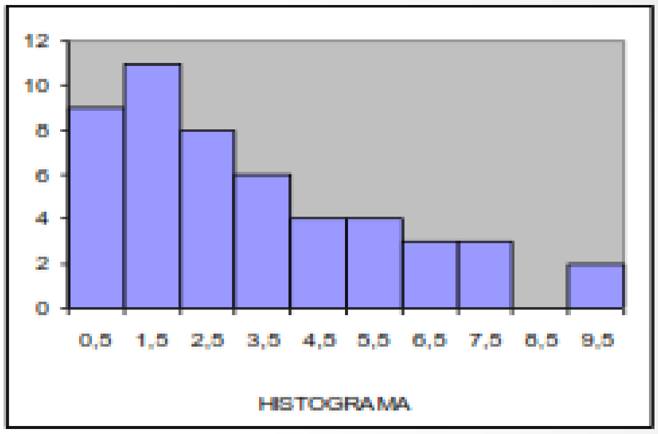
La columna de las frecuencias relativas se denomina nuevamente distribución de ***frecuencias relativas*** o simplemente ***distribución.***

Las tablas estadísticas presentadas son la expresión escrita de la distribución de frecuencias de los individuos de una muestra respecto a una variable.

Gráficos estadísticos

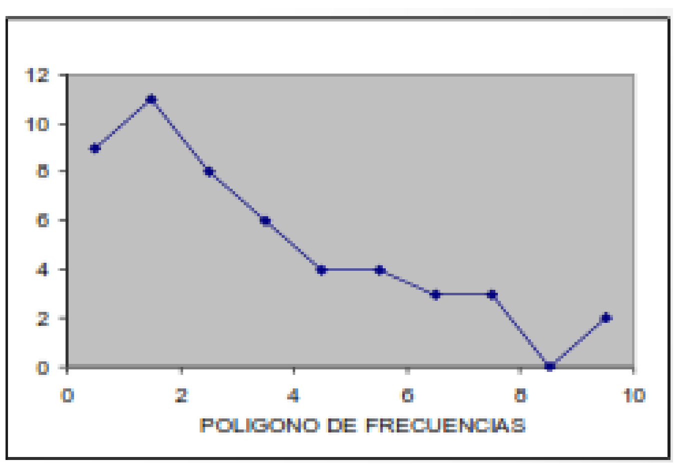
Histogramas

Es un conjunto de rectángulos que se determinan representando las frecuencias relativas en el eje vertical contra los límites reales inferiores, para cada una de las clases en el eje horizontal del plano cartesiano.



Polígono de Frecuencias

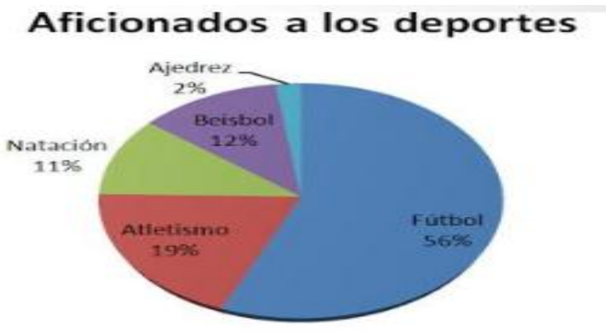
Se utiliza básicamente para mostrar la distribución de frecuencias de variables cuantitativas. Para construir el polígono de frecuencia se toma la marca de clase que coincide con el punto medio de un histograma.



Sectores Circulares

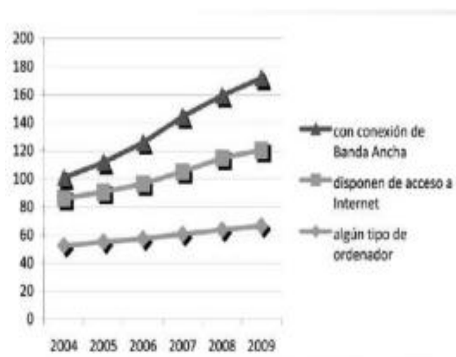
Gráfico de torta, debido a su forma característica de una circunferencia dividida en sectores, por medio de radios que dan la sensación de un pastel cortado en porciones.

Se usa para representar variables cualitativas en porcentajes o cifras absolutas cuando el número de ítems no es superior a 5 y se quiere resaltar uno de ellos.



Líneas o Tendencia

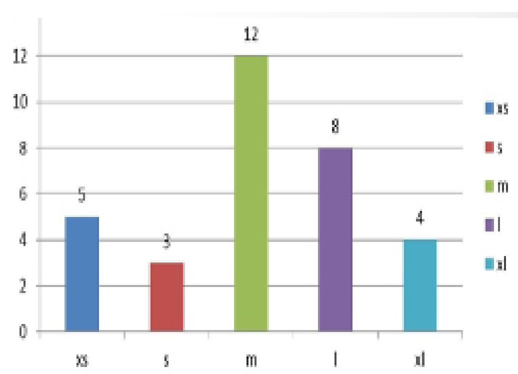
Usado básicamente para mostrar el comportamiento de una variable cuantitativa a través del tiempo. El gráfico de líneas consiste en segmentos rectilíneos unidos entre sí, los cuales resaltan las variaciones de la variable por unidad de tiempo.



Barras

Está constituido por barras rectangulares de igual ancho, conservando la misma distancia de separación entre sí.

Se utiliza básicamente para mostrar y comparar frecuencias de variables cuantitativas o comportamientos en el tiempo, cuando el número de ítems es reducido.



Investigación Estadística

Una investigación se puede denominar "estadística" cuando las hipótesis son del tipo definido: afirmaciones relativas a la distribución de una o más variables aleatorias.

Componentes:

- Universo: un conjunto de entidades (personas, seres vivos, objetos) de los cuales se desea conocer alguna características.
- Variables: características medibles que poseen todas las unidades del universo.
- Objeto de la investigación: se trata de resumir información acerca de la distribución de dichas características.
- Procedimiento de recolección: es posible conocer, el valor de las variables según sus entidades del universo.
- Restricciones: recursos disponibles (humanos, técnicos, financieros)

Actividades:

- Planificación
- Ejecución y procesamiento
- Análisis y divulgación

Planificación:

1. Objetivo de la investigación
2. Universo, unidad a investigar
3. Experiencia en investigaciones similares.
4. Marco legal aplicable (confidencialidad de información)
5. Procedimiento de recolección (Censo, muestreo, Explotación estadística de registro administrativo, Experimentación)
6. Método de recolección (entrevista personal, correo, teléfono, internet).
7. Instrumentos de levantamiento: (cuestionario para el encuestador, grabadora, guías, cuestionario para el encuestado, correo electrónico, bases de datos (e-mail); manuales de levantamiento (supervisor, encuestador)

8. Presupuesto
9. Niveles de desagregación (criterios de análisis de información)
10. Marco Muestral (la muestra puede ser: mediante selección al azar o probabilística o mediante muestreo no probabilístico)
11. Prueba piloto
12. Ajustes a los instrumentos, presupuesto y cronograma.
13. Reclutamiento de personal
14. Capacitación al personal

Ejecución y procesamiento:

1. Cartografía
2. Equipos de encuestadores y supervisores
3. Asignación de cargas de trabajo y distribución del material
4. Manejo de los cuestionarios
5. Realización de las encuestas
6. Supervisión
7. Control administrativo de los trabajos de campo
8. Programación para el ingreso de datos (codificación en software)
9. Programación para la tabulación
10. Crítica y codificación
11. Ingreso de datos
12. Concentración de la información

Análisis y divulgación:

1. Consistencia interna y de completitud (El proceso de crítica pudo permitir que los cuestionarios con algunos datos faltantes siguieran su curso. Esta es la etapa en la que se debe decidir cómo se tratan los datos faltantes).
2. Comparación con otras fuentes.
3. Análisis estadístico (análisis rápido de distribuciones univariadas: mínimos, máximos, media, cuartiles).

4. Tabulados finales (posteriori del análisis es posible decidir sobre la pertinencia de las aperturas de respuesta hasta el máximo nivel de desagregación).
5. Informe final (proporcionar a los usuarios información sobre la metodología utilizada en la investigación, los principales resultados, los problemas enfrentados).
6. Publicación y difusión de resultados (difundir a los usuarios y diferentes medios).
7. Informe técnico. (Informe está destinado a un número muy reducido de usuarios: académicos, periodistas especializados, investigadores y las personas que tendrán en el futuro la responsabilidad de realizar investigaciones similares).

MEDIDAS DE TENDENCIA CENTRAL

- **Medidas de posición**
- **Medidas de dispersión**

Medidas de posición.- Son aquellos valores numéricos que nos permiten dar alguna medida de tendencia central o bien fragmentar la cantidad de datos en partes iguales**.**
Entre las medidas de posición tenemos: Media, la mediana, la moda, los cuartiles, deciles y percentiles.

Medidas de dispersión.- Son aquellas que retrata la distancia de los valores de la variable a un cierto valor central, o que permiten identificar la concentración de los datos en un cierto sector del recorrido de la variable.

Entre las medidas de dispersión tenemos: Rango o amplitud, Desvío estándar y la varianza.

Medidas de posición

Media

La media de n observaciones $x_1, x_1, x_1, \ldots\ldots\ldots\ldots x_n$ es el promedio aritmético de los mismos (suma de todos los datos divididos para el total n) y denotamos por $\bar{x}$, es decir:

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

Esta medida también se llama media aritmética o media muestral.

El símbolo $\bar{x}$ se referirá a la media de una muestra. La media de todas las observaciones de una población se representará con el símbolo µ. Obsérvese que en general no es posible medir µ; más bien µ *es un parámetro desconocido que se desea estimar a partir de la información de una muestra.*

En general si las observaciones x_i tienen una frecuencia absoluta *fi* entonces la media viene calculada por:

$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{n}$ donde $n = \sum_{i=1}^{n} f_i$ y para el caso continuo:

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

EJEMPLOS

Para el caso discreto: Tomemos el ejemplo anterior del número de personas activas que hay en 50 familias:

Variable	*fi*	x_i *fi*
1	16	16
2	20	40
3	9	27
4	5	20
Total	50	103

Aplicando la relación $\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{n} = \frac{103}{50} = 2.06$

Lo que significa que en promedio 2 personas activas hay en 50 familias

Para el caso continuo: tomando los datos de la tabla que corresponde al tiempo en que una persona se demora en hacer una transacción bancaria tendríamos:

Clase	Límite Inferior	Límite Superior	Punto Medio (*xm*)	Frec. de Clase (*fi*)	*fi xm*
(0 - 1]	0	1	0,5	9	4,5
(1 - 2]	1	2	1,5	11	16,5
(2 - 3]	2	3	2,5	8	20
(3 - 4]	3	4	3,5	6	21
(4 - 5]	4	5	4,5	4	18
(5 - 6]	5	6	5,5	4	22
(6 - 7]	6	7	6,5	3	19,5
(7 - 8]	7	8	7,5	3	22,5
(8 - 9]	8	9	8,5	0	0
(9 -10]	9	10	9,5	2	19
Total				50	163

Aplicando la relación $\bar{x} = \frac{\sum_{i=1}^{n} x_i xm_i}{n} = \frac{163}{50} = 3.26$

Entonces el tiempo medio de transacción bancaria es de 3.26 minutos

Propiedades:

- Si multiplicamos o dividimos todas las observaciones por un mismo número, la media queda multiplicada o dividida por dicho número.
- Si le sumamos a todas las observaciones un mismo número, la media aumentará en dicha cantidad.

Mediana

La mediana es el valor central de la variable, es decir, supuesta la muestra ordenada en orden creciente o decreciente, el valor que divide en dos partes la muestra.

Para calcular la mediana debemos tener en cuenta si la variable es discreta o continua.

Cálculo de la mediana en el caso discreto (serie estadística):

Tendremos en cuenta el tamaño de la muestra.

Si n es Impar, hay un término central, el término $x_{\frac{n+1}{2}}$ que será el valor de la mediana.

Si N es Par, hay dos términos centrales, $x_{\frac{n}{2}}, x_{\frac{n}{2}+1}$ la mediana será la media de esos dos valores

Veamos un ejemplo.

CUANDO n ES PAR	CUANDO n ES IMPAR
1,4,6,7,8,9,12,16,20, 24,25,27 n =12	1,4,6,7,8,9,12,16,20, 24,25,27,30 n=13
Términos Centrales el 6º y 7º 9 y 12	Término Central el 7º , 12
$Me = \frac{9+12}{2} = 10.5$	$Me = 12$

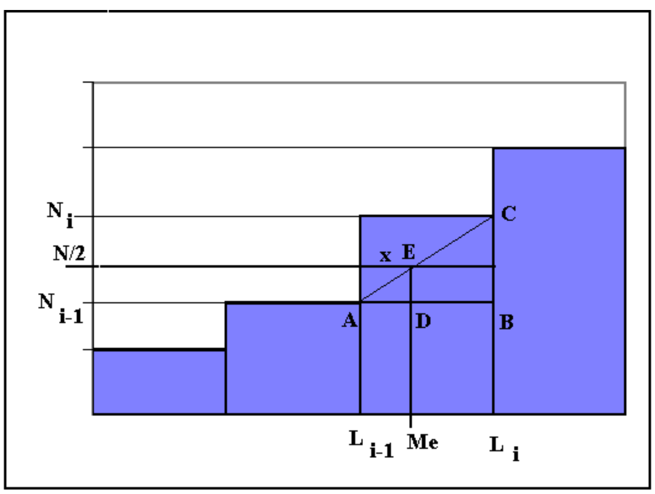
Cálculo de la mediana en el caso discreto (por frecuencias):

Tomando el ejemplo número de personas activas que hay en 50 familias, dividimos el número de datos para 2, es decir n/2 que resulta 25, en la columna de las frecuencias acumuladas (ver tabla) tomamos el valor mayor próximo, es decir 36, cuyo valor de la variable corresponde a 2, entonces $Me = 2$

C***álculo de la mediana en el caso continúo:***

Si la variable es continua, la tabla vendrá en intervalos, por lo que se calcula de la siguiente forma:

Nos vamos a apoyar en un gráfico de un histograma de frecuencias acumuladas.



De donde la mediana vale: $Me = L + \frac{\frac{n}{2} - \sum f}{fm} ai$, donde a_i es la amplitud del intervalo. ***Veámoslo por medio de un ejemplo. (ver tabla anterior para el caso continuo)***

Clase	Límite Inferior	Límite Superior	Punto Medio (*xm*)	Frec. de Clase (*fi*)	Frec. Acum. Abs. *Fi*
(0 - 1]	0	1	0,5	9	9
(1 - 2]	1	2	1,5	11	20
(2 - 3]	2	3	2,5	8	28
(3 - 4]	3	4	3,5	6	34
(4 - 5]	4	5	4,5	4	38
(5 - 6]	5	6	5,5	4	42
(6 - 7]	6	7	6,5	3	45
(7 - 8]	7	8	7,5	3	48
(8 - 9]	8	9	8,5	0	48
(9 -10]	9	10	9,5	2	50
Total				50	

Donde los valores para nuestro problema son:

- L es el valor de la clase mediana (la clase que contiene la mediana de los datos agrupados) y es 2
- n es el número total de datos y es 50
- Σf es la suma de las frecuencias de toda la clase por debajo de la clase mediana y es 20
- fm es la frecuencia de la clase mediana y es 8
- ai es la longitud de la clase mediana y es 1

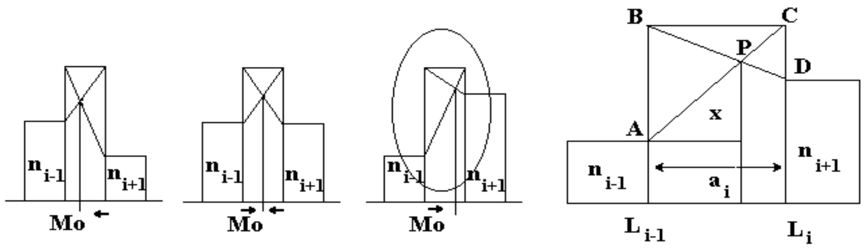
Por tanto, aplicando la fórmula computacional de la mediana tenemos que la *Me* = 2.625.

Moda

La moda es el valor de la variable que tenga mayor frecuencia absoluta, la que más se repite, es la única medida de centralización que tiene sentido estudiar en una variable cualitativa, pues no precisa la realización de ningún cálculo.

Por su propia definición, la moda no es única, pues puede haber dos o más valores de la variable que tengan la misma frecuencia siendo esta máxima. En cuyo caso tendremos una distribución *bimodal* o *multimodal* según el caso.

Por lo tanto, el cálculo de la moda en distribuciones discretas o cualitativas no precisa de una explicación mayor; sin embargo, debemos detenernos un poco en el cálculo de la moda para distribuciones cuantitativas continuas.



EJEMPLO

a) 2, 2, 5, 7, 9, 9, 9, 10, 11, 12 — Moda = 9

b) 3, 5, 8, 10, 12 — No tiene Moda

c) 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, 9 — Modas = 4 y 7 (Bimodal)

Para datos Distribuidos por Intervalos de Clase, la Moda viene dada por la ***Marca de Clase,* de** la Clase que mayor frecuencia tenga.

Aunque la moda no es una medida de tendencia central muy relevante, es la única a la que puede hacerse referencia en el caso de las variables cualitativas o cuantitativas discretas, cuando las mismas se encuentran agrupadas en distribuciones de frecuencia, llamándose Moda a la clase de mayor frecuencia.

Otros autores dan una expresión aproximada para la moda que viene dada por la siguiente expresión:

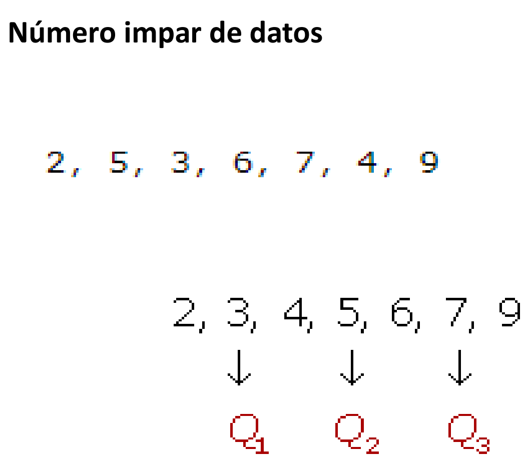
$$Mo = L_{i-1} + \frac{n_{i+1}}{n_{i+1} + n_{i-1}} \cdot a_i$$

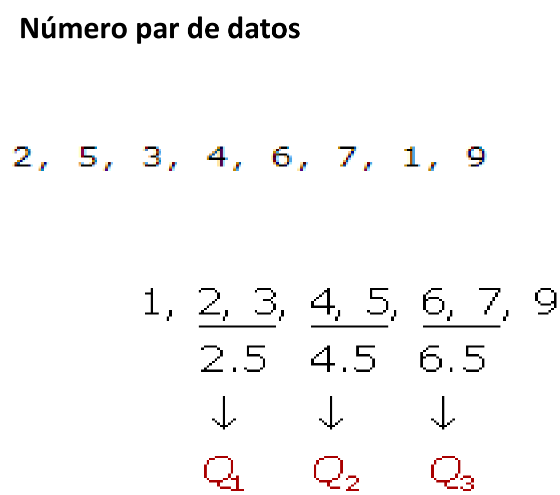
¿Cuál sería el valor de la moda para el caso discreto y continuo?

Cuartiles

Son valores que dividen una muestra de datos en cuatro partes iguales.

Cuartil	**Descripción**
1er cuartil (Q1)	25% de los datos es menor que o igual a este valor.
2do cuartil (Q2)	La mediana. 50% de los datos es menor que o igual a este valor.
3er cuartil (Q3)	75% de los datos es menor que o igual a este valor.
Rango intercuartil	La distancia entre el primer 1er cuartil y el 3er cuartil (Q3-Q1); de esta manera, abarca el 50% central de los datos.





Cuartiles, datos agrupados

En primer lugar buscamos la clase donde se encuentra: $\frac{k*N}{4}$; K=1,2,3 en la **tabla de las frecuencias acumuladas**.

$$Q_k = L_i + \frac{\frac{k \cdot N}{4} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1, 2, 3$$

L_i es el límite inferior de la clase donde se encuentra el cuartil.

N es la suma de las frecuencias absolutas.

F_{i-1} es la **frecuencia acumulada** anterior a la clase del cuartil.

a_i es la amplitud de la clase.

Ejercicio:

Calcular los cuartiles de la distribución de la tabla:

	f_i	F_i
[50, 60)	8	8
[60, 70)	10	18
[70, 80)	16	34
[80, 90)	14	48
[90, 100)	10	58
[100, 110)	5	63
[110, 120)	2	65
Σ	65	

$$\frac{65 \cdot 1}{4} = 16.25$$

$$Q_1 = 60 + \frac{16.25 - 8}{10} \cdot 10 = 68.25$$

Deciles

Los deciles son los nueve valores que dividen la serie de datos en diez partes iguales.

Los deciles dan los valores correspondientes al 10%, al 20%... y al 90% de los datos.

D5 coincide con la mediana

En primer lugar buscamos la clase donde se encuentra $\frac{k*N}{10}$; K=1,2,..9 en la tabla de las frecuencias acumuladas.

$$D_k = L_i + \frac{\frac{k \cdot N}{10} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1,2,..9$$

	f_i	F_i
[50, 60)	8	8
[60, 70)	10	18
[70, 80)	16	34
[80, 90)	14	48
[90, 100)	10	58
[100, 110)	5	63
[110, 120)	2	65
Σ	65	

$$\frac{65 \cdot 4}{10} = 26$$

$$D_4 = 70 + \frac{26 - 18}{16} \cdot 10 = 75$$

Percenti

Los percentiles son los 99 valores que dividen la serie de datos en 100 partes iguales.

Los percentiles dan los valores correspondientes al 1%, al 2%... y al 99% de los datos.

P50 coincide con la mediana.

En primer lugar buscamos la clase donde se encuentra $\frac{k*N}{100}$; K=1,2,.9.9 en la tabla de las frecuencias acumuladas.

$$P_k = L_i + \frac{\frac{k \cdot N}{100} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1, 2, ..99$$

Medidas de Dispersión

Rango o Amplitud

Es la primera medida que vamos a estudiar, se define como la diferencia existente entre el valor mayor y el menor de la distribución, Lo notaremos como R. Realmente no es una medida muy significativa en la mayoría de los casos, pero indudablemente es muy fácil de calcular.

Es la primera medida que vamos a estudiar, se define como la diferencia existente entre el valor mayor y el menor de la distribución, Lo notaremos como R. Realmente no es una medida muy significativa en la mayoría de los casos, pero indudablemente es muy fácil de calcular.

Varianza

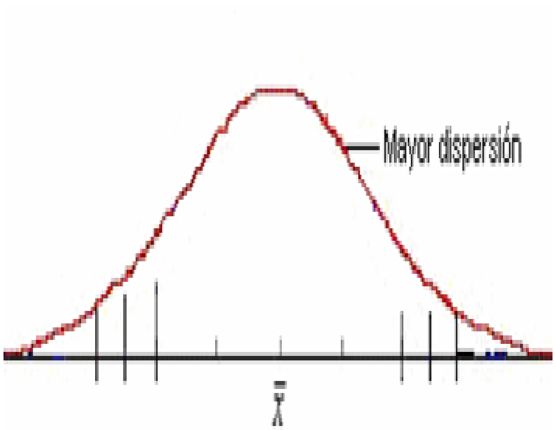
La varianza mide la dispersión dentro de un conjunto de datos. Si el valor de la varianza es pequeño, significa que los valores del conjunto están bastante agrupados. Si, por el contrario, el resultado de la varianza es mayor, quiere decir que los elementos dentro del conjunto que se analiza están dispersos. La varianza se representa con la letra griega Sigma (σ) elevada al cuadrado, o sea (σ^2), y se calcula según representa la fórmula:

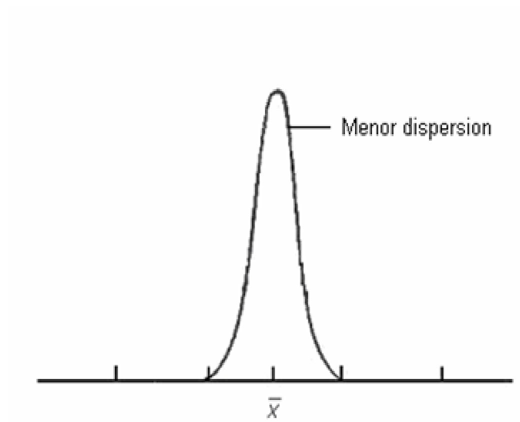
$$\sigma^2 = \frac{\Sigma (Xi-\mu)^2}{n}$$

O lo que es lo mismo:

$$\sigma^2 = \frac{(Xi-\mu)^2 + (Xi-\mu)^2 + (Xi-\mu)^2 + \ldots\ldots\ldots\ldots (Xn-\mu)^2}{n}$$

Dado el nivel de abstracción que implica el concepto de varianza, así como la dificultad para su comprensión, la varianza generalmente se calcula como punto de partida para conocer y cuantificar la desviación estándar.





Es la media de los cuadrados de las desviaciones, y la denotaremos por S_x^2

$$S_x^2 = \sigma_x^2 = \sum_{i=1}^{n} \frac{d_i^2 \bullet n_i}{N} = \sum_{i=1}^{n} \frac{\left(x_i - \bar{x}\right)^2 \bullet n_i}{N}$$

Aunque también es posible calcularlo como:

$$S_x^2 = \sigma_x^2 = \sum_{i=1}^{n} \frac{x_i^2 \bullet n_i}{N} - \bar{x}^2$$

Este estadístico tiene el inconveniente de ser poco significativo, pues se mide en el cuadrado de la unidad de la variable, por ejemplo, si la variable viene dada en cm. La varianza vendrá en cm

Desviación

Es la diferencia que se observa entre el valor de la variable y la media aritmética. La denotaremos por di.

No es una medida, son muchas medidas, pues cada valor de la variable lleva asociada su correspondiente desviación, por lo que precisaremos una medida que resuma dicha información.

La primera solución puede ser calcular la media de todas las desviaciones, es decir, si consideramos como muestra la de todas las desviaciones y calculamos su media. Pero esta solución es mala pues como veremos siempre va a ser 0.

$$\overline{D} = \sum_{i=1}^{n} \frac{d_i \cdot n_i}{N} = \sum_{i=1}^{n} \frac{(x_i - \overline{x}) \cdot n_i}{N} = \sum_{i=1}^{n} \frac{x_i \cdot n_i}{N} - \sum_{i=1}^{n} \frac{n_i}{N} \cdot \overline{x} = \overline{x} - \overline{x} = 0$$

Luego por lo tanto esta primera idea no es válida, pues las desviaciones positivas se contrarrestan con las negativas.

Para resolver este problema, tenemos dos caminos:

Tomar el valor absoluto de las desviaciones. ***Desviación media***

Elevar al cuadrado las desviaciones. ***Varianza.***

Desviación Media

Es la media de los valores absolutos de las desviaciones, y la denotaremos por d_m.

$$d_m = \sum_{i=1}^{n} \frac{|d_i| \cdot n_i}{N} = \sum_{i=1}^{n} \frac{|x_i - \overline{x}| \cdot n_i}{N}$$

Desviación Típica:

Es la raíz cuadrada de la varianza, se denota por S_x

$$S_x = \sigma_x = \sqrt{\sum_{i=1}^{n} \frac{d_i^2 \cdot n_i}{N}} = \sqrt{\sum_{i=1}^{n} \frac{x_i^2 \cdot n_i}{N} - \overline{x}^2}$$

Este estadístico se mide en la misma unidad que la variable por lo que se puede interpretar mejor.

Actividades de Aprendizaje

En un grupo de estudiantes se considera el número de ensayos que necesita cada uno para memorizar una lista de seis pares de palabras.

Los resultados fueron: 5 8 3 9 6 7 10 6 7 4 6 9 5 6 7 9 4 6 8 7

a) Construya la tabla de frecuencias.

b) Calcule la moda, la media, la mediana y el tercer cuartil de las observaciones dadas. Obtenga la frecuencia del conjunto de los resultados superiores a 5.

c) Calcule la varianza y el desvío estándar.

d) Un grupo de 20 actores fue sometido a la misma experiencia que los estudiantes mencionados arriba.

Para ellos resultó una media de 4,8 y un desvío de 1,8. En base a los resúmenes estadísticos adecuados señale: d1) cuál es el grupo de mejor desempeño en la experiencia realizada. Justifique su respuesta.

d2) en cuál grupo los integrantes son más parecidos entre sí en relación a la cantidad de ensayos necesarios para memorizar la lista de seis pares de palabras. Justifique su respuesta.

LA MUESTRA

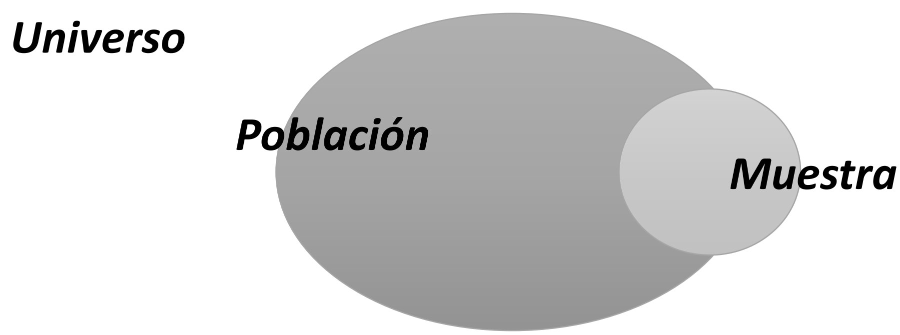
Definición de términos

Población (o universo).- es cualquier colección finita o infinita de elementos o sujetos.

Población es finita.- se refiere a un número limitado de elementos, ejemplo: todos los habitantes del país.

Población es infinita.- hace referencia cuando no se puede contabilizar todos sus elementos que existen, como por ejemplo, las estrellas, insectos en el mundo.

Muestra.- es un subconjunto de la población, que se obtiene para averiguar las propiedades o características.



Ventajas y limitaciones del uso de muestras

Ventajas:

- Costo reducido,
- Mayor rapidez,
- Mayor exactitud,
- Mayores posibilidades.

Limitaciones:

- No es recomendable emplear muestras en poblaciones pequeñas.
- La teoría del muestreo es compleja.

Características de una buena muestra

- Debe ser adecuada en cantidad y en calidad.

- La calidad involucra el concepto de representatividad de la muestra.
- Debe reunir las características principales de la población en relación a la variable de estudio.

Tipos de muestreo

Marco muestral.- es una lista de elementos que componen el universo en estudio y de la cuál se extrae la muestra. Los elementos pueden ser personas, empresas, hogares, vehículos, otros.
Cada uno de estos elementos presentes en el marco muestral se conoce como unidades muestrales
La proporción existente entre el tamaño de la muestra y el tamaño del marco muestral se conoce como fracción muestral

Tipos:

- **Probabilístico**
- **No probabilístico**

Muestreo Probabilisrtico

Características

Los elementos de mi población tienen una probabilidad mayor de cero de ser seleccionados en la muestra.
Se conoce de forma precisa dicha probabilidad para cada elemento, lo que se conoce como probabilidad de inclusión.

En la muestra, el cumplimiento de estas características hace posible obtener resultados no sesgados.

Muestreo NoProbabilisrtico

Características

Disponer de un marco muestral es poco habitual en estudios de mercado.

Lograr que la población tenga una probabilidad no nula de ser seleccionados es un requisito igualmente exigente, más aún conocer la probabilidad de inclusión exacta de cada unidad muestral.

Las técnicas aplicadas en este tipo de muestreo son: el muestreo por conveniencia, secuencial, por cuotas, discrecional y por bola de nieve.

Definición del tamaño de muestra para poblaciones finitas y no finitas (estimación de proporciones)

POBLACIÓN FINITA

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

POBLACIÓN INFINITA

$$n = \frac{Z^2 * p * q}{e^2}$$

Dónde:

n: es el resultado del tamaño de muestra.

N: tamaño de la población.

e: es el error muestral, es decir el error que se asume al equivocarse.

P: es la probabilidad de éxito o el evento de ocurrencia, por lo general se recomienda que sea del 50% o 0.50. Cuando se desconoce la varianza se utiliza la equivalencia de “p*q”.

Q: es la probabilidad de fracaso o el evento de no ocurrencia, por lo general se recomienda que sea del 50% o 0.50.

Z: valor estimado a partir del nivel de confianza de la muestra, es decir, en qué medida son confiables los datos de la muestra, es recomendable utilizar el 90%, 95% y 99%, sin embrago se utiliza con mayor frecuencia el 95%.

Definición del tamaño de muestra para poblaciones finitas y no finitas (estimación de promedios)

POBLACIÓN FINITA

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

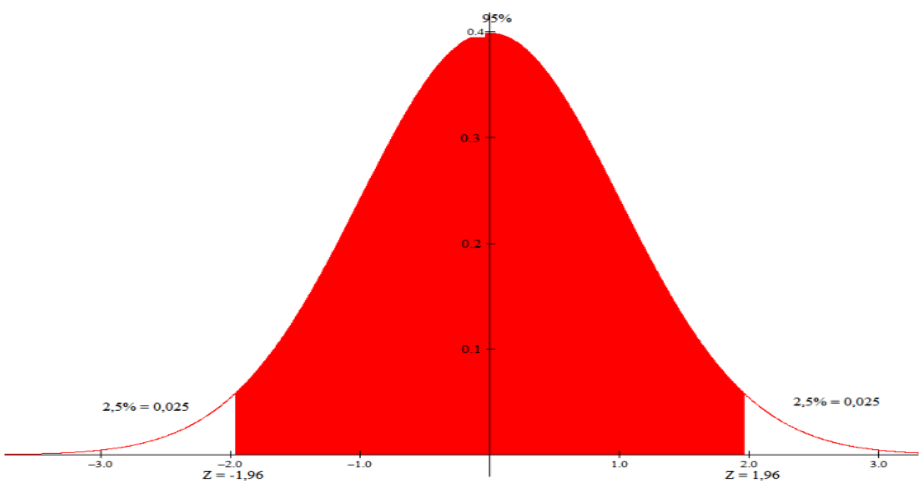
POBLACIÓN INFINITA

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

EJEMPLO:

Calcular el tamaño de la muestra de una población de 500 elementos con un nivel de confianza del 95%

Realizando el gráfico que representa el 95% de confianza se obtiene



Se tiene N=500, para el 95% de confianza Z = 1,96, y como no se tiene los demás valores se tomará e = 0,05.

REEMPLAZANDO: n= 217

NÚMEROS ÍNDICES

El número índice es un valor relativo que se expresa en porcentaje o cociente, medido un período dado en relación a un periodo base.
Estos índices, indica variación relativa en precios, cantidad o valor**.**
Entre los índices simples y complejos de mayor interés son:

- índice de precios
- Índice de cantidades
- Índice de valores

Clasificación de los números índices.

- Índice Suimple
- Índice completo

Índice Complejos, hace referencia a varios productos o conceptos.
Estos se clasifican en:

- **Sin ponderar:** existen varias metodologías: Sauerbeck (media aritmetica), Media Geométrica, Media armónica, Bradstreet – Dutot (media agregativa)
- **Ponderados:** Laspeyres, Paasche, Edgeworth, Fisher

Índice simple

Considera los siguientes datos como los referentes a los precios, volumen (cantidad) de ventas e ingreso (monto o valor) en el caso de una comercializadora
líder de productos electrónicos.

AÑO	PRECIO	VOLUMEN DE VENTAS (CANTIDAD)	INGRESO
2000	3000	60	180000
2001	3300	63	207900
2002	3900	60	234000
2003	4500	66	297000
2004	4500	72	324000
2005	4800	75	360000
2006	4950	66	326700

La siguiente tabla muestra los índices ya calculados:

AÑO	PRECIO		CANTIDAD		INGRESOS	
	DÓLARES	ÍNDICE	UNIDADES	ÍNDICE	DÓLARES	ÍNDICE
2000	3000	100	60	100	180000	100
2001	3300	110	63	105	207900	115,5
2002	3900	130	60	100	234000	130
2003	4500	150	66	110	297000	165
2004	4500	150	72	120	324000	180
2005	4800	160	75	125	360000	200
2006	4950	165	66	110	326700	181,5

a) Como puede apreciarse los precios de los artículos aumentaron un 50% entre 2000 y 2004, la cantidad vendida aumentó un 20% entre esos mismos años y el ingreso o valor de las operaciones aumentó un 80%.

Índices complejos No ponderados

La tabla siguiente muestra los precios de un conjunto de artículos en el año 2004 y en el año 2007:

ARTÍCULO	PRECIO (2004)	PRECIO (2007)
1	0.80/ KG	1.20/ KG
2	0.10/ PIEZAS	0.08/ PIEZAS
3	1.00/ METRO2	2.00/ METRO
4	0.10/ PIEZAS	0.25/ PIEZAS

Obtener

a) Los índices simples de precios

b) El promedio simple de índices de precios e interpreta el resultado

c) El índice agregado simple e interpreta el resultado

Solución

a) Los índices simples de precios

ARTÍCULO	PRECIO (2004)	PRECIO (2007)	ÍNDICE
1	0.80/ KG	1.20/ KG	150.0
2	0.10/ PIEZAS	0.08/ PIEZAS	80.0
3	1.00/ METRO2	2.00/ METRO	200.0
4	0.10/ PIEZAS	0.25/ PIEZAS	250.0
TOTAL	2.00	3.53	

b) El promedio simple de índices de precios e interpreta el resultado

$$P=\frac{\Sigma Pi}{n}=\frac{150+80+200+250}{4}=170$$

Lo cual indica que en promedio los índices de precios de los artículos aumentó 70% del 2004 al 2007.

c) El índice agregado simple se obtiene con las simples sumas de los precios de los artículos en cada año, aplicando su fórmula:

$$P=\frac{\Sigma Pn}{\Sigma P0} * 100 = \frac{3.53}{2} * 100=176.5$$

Lo cual indica que el grupo agregado de preciso ha aumentado de 2004 a 2007 un 76.5% en tan solo 3 años.

Índices complejos ponderados

Nombre	Fórmula	donde:	Ventajas	Desventajas
Indice de Laspeyres	$I(L)=\frac{\Sigma p_n q_0}{\Sigma p_0 q_0} * 100$	Los precios se ponderan con las cantidades del año base	Requiere datos solo del periodo base Compara mejor con respecto al tiempo Sus cambios reflejan más los cambios de precios	No refleja cambios en los patrones de compra al paso del tiempo Podría ponderar más los artículos cuyos precios aumentan
Indice de Paasche	$I(P)=\frac{\Sigma p_n q_n}{\Sigma p_0 q_n} * 100$	Los precios se ponderan con las cantidades del año dado	Refleja los hábitos actuales de compra	Requiere datos de cantidad de cada año Como se basa en las cantidades sus cambios o no pueden ser atribuidos a los precios Pondera más los artículos cuyos precios han bajado Requiere que los precios se recalculen cada año
Indice ideal de Fisher	$IIF=\sqrt{I(L) * I(P)}$	Es la media geométrica de los dos anteriores	Combina las mejores características de los dos anteriores.	Requiere de determinar un nuevo grupo de cantidades cada año
Promedio ponderado de relativos de precios	$I_p=\frac{\Sigma[(p_0 q_0)\ (p_n / p_0 * 100)]}{\Sigma p_0 q_0}$		Los índices simples se ponderan con una cifra de valor pq Estos pueden ser del periodo base o del periodo dado	

En la siguiente tabla están los precios de 6 distintos artículos tanto en el año 2000 como en al año 2006:

ARTÍCULO	PRECIO EN 2000	CANTIDAD EN 2000	PRECIO EN 2006	CANTIDAD EN 2006
NARANJAS (KG)	**1,46**	**30**	**1,56**	**40**
AZUCAR (KG)	**4,4**	**12**	**4,62**	**12**
ACEITE (LTRO)	**1,58**	**40**	**1,7**	**41**
HUEVOS (DOCENA)	**1,85**	**26**	**1,84**	**20**
LECHE (LTRO)	**0,88**	**102**	**1,01**	**130**
PAN (KG)	**0,77**	**50**	**0,89**	**55**
TOTAL	**10,94**		**11,62**	

Determinar índices ponderados según:

a) Laspeyres

b) Paasche

Desarrollo

Determinar índices ponderados según:

a) Laspeyres

ARTÍCULO	PRECIO EN 2000	CANTIDAD EN 2000	PRECIO EN 2006	CANTIDAD EN 2006	PRECIO 2000*CANTIDAD 2000	PRECIO 2006*CANTIDAD 2000
NARANJAS (KG)	**1,46**	**30**	**1,56**	**40**	**43,8**	**46,8**
AZUCAR (KG)	**4,4**	**12**	**4,62**	**12**	**52,8**	**55,44**
ACEITE (LTRO)	**1,58**	**40**	**1,7**	**41**	**63,2**	**68**
HUEVOS (DOCENA)	**1,85**	**26**	**1,84**	**20**	**48,1**	**47,84**
LECHE (LTRO)	**0,88**	**102**	**1,01**	**130**	**89,76**	**103,02**
PAN (KG)	**0,77**	**50**	**0,89**	**55**	**38,5**	**44,5**
TOTAL	**10,94**		**11,62**		**336,16**	**365,6**

$$I(L) = \frac{\sum p_n q_0}{\sum p_0 q_0} * 100 \qquad \frac{365,60}{336,16}*100=108.8$$

Indica que el precio de este grupo artículos aumentó 8.8% del año 2000 al 2006.

Determinar índices ponderados según:

b) Paasche

ARTÍCULO	PRECIO EN 2000	CANTIDAD EN 2000	PRECIO EN 2006	CANTIDAD EN 2006	PRECIO 2006*CANTIDAD 2006	PRECIO 2000*CANTIDAD 2006
NARANJAS (KG)	**1,46**	**30**	**1,56**	**40**	**62,4**	**58,4**
AZUCAR (KG)	**4,4**	**12**	**4,62**	**12**	**55,44**	**52,8**
ACEITE (LTRO)	**1,58**	**40**	**1,7**	**41**	**69,7**	**64,78**
HUEVOS (DOCENA)	**1,85**	**26**	**1,84**	**20**	**36,8**	**37**
LECHE (LTRO)	**0,88**	**102**	**1,01**	**130**	**131,3**	**114,4**
PAN (KG)	**0,77**	**50**	**0,89**	**55**	**48,95**	**42,35**
TOTAL	**10,94**		**11,62**		**404,59**	**369,73**

$$I(P) = \frac{\sum p_n q_n}{\sum p_0 q_n} * 100 \qquad \frac{404{,}59}{369{,}73} * 100 = 109.4$$

Indica que entre 2000 y 2006 ha habido un incremento en el precio de este grupo de artículos de un 9.4%, es decir, que en 2006 cuesta 9.4% más comprar estos artículos que lo que costaban en 2000.

Determinar índices ponderados según:

c) Fisher

Para determinar el índice ideal de Fisher, ya solo se aplica la fórmula:

$$IIF = \sqrt{I(L) * I(P)}$$

$$IIF = \sqrt{(108.8) * (109.4)} = 109.1$$

Indica que los precios de estos artículos se han incrementado en un 9.1%

Deflactación

Deflactar, significa eliminar de un valor monetario los efectos producidos por los cambios en los precios (inflación o deflación).
Por lo tanto, deflactar consiste en convertir una magnitud medida en términos nominales en otra expresada en términos reales.

El índice de Precios al Consumidor (IPC), es un indicador mensual, nacional que mide los cambios en el tiempo del nivel general de los precios, correspondientes al consumo final de bienes y servicios de los hogares de estratos de ingreso.

EJERCICIO

Con la siguiente información obtenga el índice de precios del consumidor y deflactor del consumo.

PRODUCTOS	2010			2011			IPC	
	CANTIDAD PRODUCIDA	CANTIDAD VENDIDA	PRECIOS	CANTIDAD PRODUCIDA	CANTIDAD VENDIDA	PRECIOS	PRECIO 2011*CANTIDAD VENDIDA 2010	PRECIO 2010*CANTIDADES VENDIDAS 2010
ARROZ (KG)	100	80	10	110	80	15	1200	800
AZUCAR (KG)	120	130	6	170	175	7	910	780
TOTAL							2110	1580

$$IPC=\frac{\Sigma Pn*Q0}{\Sigma P0*Q0}*100$$

$$IPC=\frac{2110}{1970}*100 = 133,5$$

EJERCICIO

Deflactor del consumo

PRODUCTOS	2010			2011			IPC	
	CANTIDAD PRODUCIDA	CANTIDAD VENDIDA	PRECIOS	CANTIDAD PRODUCIDA	CANTIDAD VENDIDA	PRECIOS	PRECIO 2011*CANTIDAD VENDIDA 2011	PRECIO 2010*CANTIDADES VENDIDAS 2011
ARROZ (KG)	100	80	10	110	80	15	1200	800
AZUCAR (KG)	120	130	6	170	175	7	1225	1050
TOTAL							2425	1850

$$\text{Deflactor}=\frac{\Sigma Pn*Qn}{\Sigma P0*Qn}*100$$

$$\text{Deflactor}=\frac{2425}{1850}*100 = 131{,}08$$

COEFICIENTE DE REGRESIÓN

Expresándolo en forma simple, la regresión lineal es una técnica que permite cuantificar la relación que puede ser observada cuando se grafica un diagrama de puntos dispersos correspondientes a dos variables, cuya tendencia general es rectilínea (Figura la); relación que cabe compendiar mediante una ecuación “del mejor ajuste” de la forma:

y = a + bx

En esta ecuación, “y” representa los valores de la coordenada a lo largo del eje vertical en el gráfico (ordenada); en tanto que “x” indica la magnitud de la coordenada sobre el eje horizontal (abscisa). El valor de “a” (que puede ser negativo, positivo o igual a cero) es llamado el intercepto; en tanto que el valor de “b” (el cual puede ser negativo o positivo) se denomina la pendiente o coeficiente de regresión.

Serie de datos para el cálculo de una regresión (“a” y “b”) y del coeficiente de correlación (“r”)

Número	**Valores de x**	**Valores de y**	**Número**	**Valores de x**	**Valores de y**
1	9,0	0,50	7	6,7	1,00
2	9,4	0,50	8	8,4	0,50
3	7,4	1,23	9	8,0	0,50
4	9,7	1,00	10	10,0	0,50
5	10,4	0,30	11	9,2	0,50
6	5,0	1,50	12	6,2	1,00
			13	7,7	0,50

El procedimiento para obtener valores de “a” y “b” para una serie de pares de datos de “x” y de “y” es como sigue:

Paso 1 Calcule, para cada par de valores de “x” e “y”, las cantidades “x^2”, “y^2”, y “x.y”.

Paso 2 Obtenga las sumas (∑) de estos valores para todos los pares de datos de “x” e “y”, así como las sumas del total de los valores de “x” e “y”. Los resultados de los Pasos 1 y 2 aparecerán en forma similar a la siguiente:

Número de pares de datos	**x**	**x^2**	**y**	**y^2**	**x.y**
1	…	…	…	…	…
2	…	…	…	…	…
3	…	…	…	…	…
.					
.					
.					
n	…	…	…	…	…
Monto de las sumas	$\sum x$	$\sum x^2$	$\sum y$	$\sum y^2$	$\sum x\cdot y$

Paso 3 Estime la pendiente (b) por medio de la relación:

Paso 4 Estime el intercepto (a) por medio de la relación:

A partir de esos valores de “a” y de “b” obtenidos mediante las Ecuaciones 2 y 3, es posible trazar a lo largo de los puntos dispersos de un gráfico la línea recta mejor ajustada a los mismos, y verificar visualmente si tales puntos están bien “expresados” por la línea.

Coeficiente De Correlación Lineal

El análisis de correlación se encuentra estrechamente vinculado con el análisis de regresión y ambos pueden ser considerados de hecho como dos aspectos de un mismo problema.

La correlación entre dos variables es - otra vez puesto en los términos más simples - el grado de asociación entre las mismas. Este es expresado por un único valor llamado coeficiente de correlación (r), el cual puede tener valores que oscilan entre -1 y +1. Cuando “r” es negativo, ello significa que una variable (ya sea “x” o “y”) tiende a decrecer cuando la otra aumenta (se trata entonces de una “correlación negativa”, correspondiente a un valor negativo de “b” en el análisis de regresión). Cuando “r” es positivo, en cambio, esto significa que una variable se incrementa al hacerse mayor la otra (lo cual corresponde a un valor positivo de “b” en el análisis de regresión).

Los valores de “r” pueden calcularse fácilmente en base a una serie de pares de datos de “x” e “y”, utilizando la misma tabla y montos que se indican en el Paso 2 de la sección “regresión” de este capítulo. De este modo “r” puede ser obtenido - indirectamente - a partir de la relación:

$$r^2 = \frac{\Sigma xy - \left[\frac{(\Sigma x)(\Sigma y)}{n}\right]^2}{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]\cdot\left[\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right]}$$

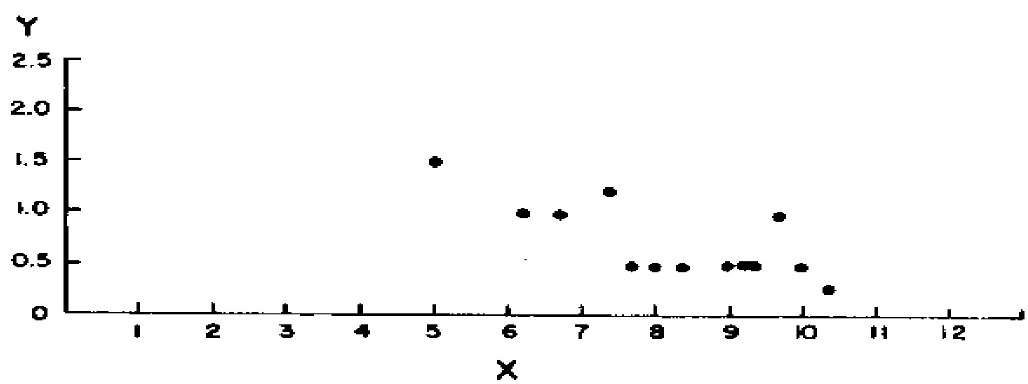


Figura 1 Diagrama de puntos dispersos correspondientes a pares de valores de “x” y de “y”. Nótese que “y” tiende a decrecer con el aumento de “x”, lo cual sugiere coeficientes de regresión y de correlación negativos.

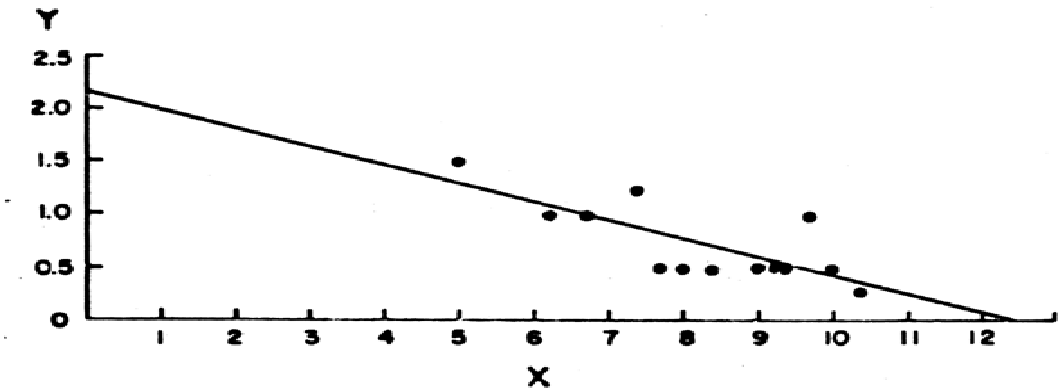


Figura 2 Los mismos datos que en 1a Fig. 1a, pero ajustados en base a la regresión y = 2,16 - 0,173x, con r = 0,75; la cual proporciona el valor del "coeficiente de determinación" (r^2). Entonces, lo único necesario es calcular

$$|r| = \sqrt{r^2}$$

es decir, tomar la raíz indicada del coeficiente de determinación a los fines de obtener el valor absoluto de "r", y luego agregar el signo (+ o -) de acuerdo a que la correlación sea positiva o negativa (lo cual puede ser establecido visualmente a partir del gráfico, o bien en base al cálculo del valor de "b" de la correspondiente regresión y utilizando para "r" el mismo signo).

Cuando se calculan los valores de "r" se querrá saber, sin embargo, hasta qué punto la correlación identificada pudiera haber surgido únicamente por casualidad. Esto puede ser establecido verificando si el valor estimado de "r" es "significativo", es decir si el valor absoluto de "r" es mayor o igual que un valor "crítico" de "r" indicado en las tablas estadísticas.

Coeficiente de determinación (R^2).- Es una medida que indica de manera porcentual el cambio de la variable dependiente respecto a la independiente.

Series de tiempo.- Nos referimos a datos estadísticos que se recopilan, observan o registran en intervalos de tiempo regulares (diario, semanal, semestral, anual, entre otros).

Componentes:

- **Tendencia secular:** es por lo común el resultado de factores a largo plazo

- **Variación estacional**: representa la variabilidad en los datos debida a influencias de las estaciones.
- **Variación cíclica:** Es la secuencias alternas de puntos abajo y arriba de la línea de tendencia que duran más de un año.
- **Variación Irregular:** Esta se debe a factores a corto plazo, imprevisibles y no recurrentes que afectan a la serie de tiempo.

EJERCICIOS PRÁCTICOS

Ejercicio: Calcule “a”, “b” y “r” a partir de los datos presentados en la Tabla 1 y verifique, por medio de la Tabla del Apéndice 1, hasta qué punto el valor estimado de “r” es significativo para valores de P = 0,01 y de P = 0,05

B. Base de Consulta

TÍTULO	AUTOR	EDICIÓN	AÑO	IDIOMA	EDITORIAL
Estadística 2	Espinoza, Alfredo	Guayaquil-Ecuador	2017	Español	
Estadística Descriptiva	Espinoza, Alfredo.	Guayaquil-Ecuador	2000	Español	
Estadística 1	Brito, Jorge	Guayaquil-Ecuador	2005	Español	
WEBGRAFÍA					
http://www.mat.uda.cl/hsalinas/cursos/2010/eyp2/Clase1.pdf					
https://www.vitutor.net/2/11/cuartiles_percentiles.html					
http://www.dgeec.gov.py/convocatoria/document/etapas_invest_estadistica.pdf					
Suárez, Mario, (2011), Interaprendizaje de Estadística Básica; Tapia, Fausto Ibarra, Ecuador.					
http://www.smo.edu.mx/colegiados/apoyos/muestreo.pdf					

C. Base práctica con ilustraciones

ESTRATEGIAS DE APRENDIZAJE

ESTRATEGIA DE APRENDIZAJE 1: Análisis y Planeación
Descripción: Discusión sobre las lecturas, artículos y videos. Observación atenta y detallada de las éticas que emiten los niños y las personas que están en su contexto para lograr la respuesta de los demás.
Ambiente(s) requerido: Aula amplia con buena iluminación.
Material (es) requerido: Infocus.
Docente: Con conocimiento de la materia.

5. ACTIVIDADES

- Controles de lectura
- Exposiciones
- Presentación del Trabajo final

Se presenta evidencia física y digital con el fin de evidenciar en el portafolio de cada aprendiz su resultado de aprendizaje. Este será evaluable y socializable

6. EVIDENCIAS Y EVALUACIÓN

Tipo de Evidencia	**Descripción (de la evidencia)**
De conocimiento:	Ensayo expositivo grupal de lecturas Definición del tema de investigación
Desempeño:	Trabajo grupal presentación de resolución de ejercicios en clase.
De Producto:	Trabajo de realizado
Criterios de Evaluación	Tareas, Taller, Participación en Clase, Exposición, Evaluación

(Mínimo 5 Actividades por asignatura)	Final.	
Elaborado por: **(Docente)**	**Revisado Por:** **(Coordinador)**	**Reportado Por:** **(Vicerrector)**

INSTITUTO TECNOLÓGICO
SUPERIOR JAPÓN

AMOR AL CONOCIMIENTO

POMASQUI-
c/Marieta Veintimilla E5-471 y Sta. Teresa 4ta transversal

Tlfs: 022356-368 - 0986915506

www.itsjapon.edu.ec