



**FACULTAD DE INGENIERÍA UNAM
DIVISIÓN DE EDUCACIÓN CONTINUA**

CURSOS INSTITUCIONALES

BIOESTADÍSTICA, MÉTODOS Y APLICACIONES

Del 27 de Octubre al 17 de Noviembre de 2004

APUNTES GENERALES

CI - 143

Instructor: Ruth Moreno Meza

ISSSTE

OCTUBRE/NOVIEMBRE DE 2004

DEFINICIÓN DE ESTADÍSTICA

La Estadística es un conjunto de técnicas para alcanzar conocimientos precisos a partir de información incompleta, como un sistema científico para la recolección, análisis interpretación y presentación de información susceptible de ser expuesta en forma numérica.

La Estadística comprende la rama descriptiva, la teoría de la probabilidad y el muestreo

La Estadística descriptiva consiste en organizar, resumir y simplificar, en términos generales, información que a menudo es bastante compleja

La Estadística proporciona instrumentos para la toma de decisiones cuando prevalece condiciones de incertidumbre, estos instrumentos pueden ser de aplicación y utilidad completamente general en cualquier campo de la ciencia: física, biológica, social, etc.

Métodos Estadísticos.

Un método estadístico es cualquier técnica utilizada para obtener, analizar y presentar datos numéricos.

Elementos de la Técnica Estadística

Los elementos de la técnica estadística incluyen:

1. Recopilación y agrupación de datos
2. Clasificación y resumen de los datos
3. Presentación de datos:
 - a) forma directa
 - b) forma tabular
 - c) Mediante gráficas
4. El análisis de los Datos

Características y limitaciones de los Métodos Estadísticos

- Los métodos estadísticos constituyen el único medio para manejar grandes masas de datos numéricos.
- Las técnicas estadísticas solo se pueden aplicar a datos que sean reducibles a una forma cuantitativa.
- Las técnicas estadísticas son objetivas; los resultados esta afectados por interpretaciones subjetivas
- Las técnicas estadísticas son idénticas tanto para las ciencias sociales, como para las ciencias físicas o para las demás ciencias Los métodos se aplican similarmente en estos campos divergentes.
- Antes de analizar los datos numéricos es necesario disponerlos sistemáticamente Los datos pueden disponerse en cierto número de formas diferentes.
- Esta disposición se denomina, técnicamente, una distribución o una serie . A su vez las series se pueden dividir en:

Serie Simple

Como su nombre lo indica esta formada únicamente por un solo conjunto de cifras con una sola clasificación, es decir, dentro de un periodo de tiempo sus avances o retrocesos de las cifras.

Ejemplo: Producto Interno Bruto, Índice Nacional, Deuda Pública, Índice de Inflación

Serie Compuesta.

Puede ser conocida también como serie de datos agrupados o series de clases y frecuencias, esto es que la representación del fenómeno de tipo económico tendrá que agruparse en pequeños sectores para poder realizar su análisis. Comúnmente los grupos se clasifican entre dos límites (límite superior y límite inferior) de tal manera que se forman clases que se cuantifican con la frecuencia con la que se requiere.

Datos Estadísticos

Los datos estadísticos se obtienen mediante un proceso que comprende la observación o medición de conceptos como ingresos anuales de una comunidad, calificaciones de exámenes, cantidad de café por tasa despachada por una maquina vendedora, resistencia a la rotura de fibras de plástico, porcentaje de azúcar en cereales, etc., tales conceptos reciben el nombre de variables, ya que producen valores que tienen a mostrar cierto grado de variabilidad, al efectuarse mediciones sucesivas.

Tipos de Datos

Existen cuatro tipos de datos: *continuos, discretos, nominales y jerarquizados.*

Las variables pueden asumir virtualmente cualquier valor en determinado intervalo de valores se conocen como continuas.

Características tales como altura, peso, longitud, espesor, velocidad, viscosidad y temperatura también quedan dentro de esta categoría. Los datos que se toman a cerca de estas características y otras semejantes se denominan continuos, aun cuando en términos prácticos, los instrumentos de medición presentan ciertas limitaciones de tipo físico que restringen el grado de precisión:

Las variables continuas pueden asumir cualquier valor en un intervalo continuo. Los datos que se obtienen acerca de estas variables reciben el nombre de datos continuos

Como ejemplos de datos continuos tenemos : La cantidad de café que se vende al día, la gasolina que se expende por hora, la velocidad del aire, el tiempo de reacción y la elasticidad de una banda de goma.

Una variable discreta es la que se puede asumir solo ciertos valores, por lo regular enteros. Los datos discretos surgen al contar el numero de conceptos que poseen cierta característica.

Ejemplos de datos discretos son el numero de clientes por día, la cantidad de alumnos de un salón de clase, los defectos de un auto, los accidentes de trabajo en una fabrica, las paradas de un autobús, el numero de anotaciones en un partido de fútbol, etc.

Las variables discretas adquieren valores enteros Los datos discretos son resultado de contar un numero de conceptos y objetos.

Tanto los datos discretos como los continuos, se conocen como datos cuantitativos ya que son inherentemente numéricos.

Una variable discreta puede tomar valores observados únicamente en puntos aislados a lo largo de una escala de valores.

Una variable continua puede adoptar un valor en cualquier punto fraccionario.

Datos nominales se obtienen cuando se define las categorías y se cuenta el numero de observaciones que quedan en cada una

Datos jerarquizados consta de valores relativamente asignados para denotar orden, primero, segundo, cuarto, etc y así sucesivamente.

¿Que es una Variable Aleatoria?

La variable que tiene resultados o valores que tienden a variar de observación en observación debido a los factores relacionados con el azar, recibe el nombre de variable aleatoria.

Una Variable aleatoria es una función valorada numéricamente, cuyo valor está regida por factores en los que interviene el azar.

Las variable aleatorias pueden ser discretas o continuas.

Una variable aleatoria se considera *discreta* si los valores que asume se pueden contar

Una variable aleatoria se considera *continua* si puede asumir cualquier valor dentro de un determinado intervalo

¿Que es una Muestra?

La parte de elementos que se examina.

¿Que es una Población?

El grupo total a partir del cual se selecciono la muestra se conoce como población o universo.

Notación y Uso del Sigma.

En muchos procedimientos estadísticos es necesario obtener la suma de un conjunto de números, la letra griega Σ se utiliza para denotar una suma. De ahí que, si alguna variable x tiene los valores 1,5,6 y 9, la sumatoria será:

$$\Sigma x = 1+5+6+9 = 21.$$

Ejemplo 1

Si los valores de x son 2, 4, 5 y 9 encuentre Σx , Σx^2 , y $(\Sigma x)^2$.

Solución

$$\Sigma x = 2+4+5+9 = 20$$

$$\Sigma x^2 = 2^2 + 4^2 + 5^2 + 9^2 = 4 + 16 + 25 + 81 = 126.$$

$$(\Sigma x)^2 = 20^2 = 400$$

Si sólo se van a sumar algunos de los valores, se utilizan subíndices para indicar dichos valores de este modo:

$$\Sigma x_i$$

indica la suma de los valores de la variable x , empezando con el primero ($i = 1$) y terminando con el quinto ($i = 5$):

$$\Sigma x_i = x + x + x + x + x$$

$\Sigma_{i=1}^n x_i$ significa que n observaciones (todas) han de ser sumadas y a menudo esto se abrevia con los símbolos Σx_i o Σx

Ejemplo 2

Utilizando los datos que se indican, calcule :

- a) $\sum_{i=1}^1 x_i$
- b) $\sum_{i=2}^2 x_i$
- c) $\sum_{i=7}^7 x_i$
- d) $\sum x_i$

DATOS

<u>i</u>	<u>x_i</u>
1	8
2	2
3	3
4	6
5	7
6	8
7	9
8	4
9	5
10	4
11	1
<hr/>	
	57

Al invertir el proceso, podemos utilizar este método para abreviar la expresión de los datos que intentamos sumar:

- 1 $x_1 + x_2 + x_3$ se convierte en $\sum_{i=1}^3 x_i$.
- 2 $x_8 + x_9 + x_{10} + x_{11}$ se convierte en $\sum_{i=8}^{11} x_i$.

DISTRIBUCION DE FRECUENCIAS

Una distribución de frecuencias es una tabla en la cual se agrupan en clases valores posibles de una variable y donde se registra el número de valores observados correspondiente a cada clase.

Los datos organizados en una distribución de frecuencias se llaman datos agrupados.

Intervalos de Clase

El intervalo de clase identifica el rango de valores incluidos dentro de una clase y puede determinarse restando del límite exacto de clase superior de la clase el límite exacto de clase inferior.

Punto Medio

De una clase denominado a menudo marca de clase se determina localizando la mitad entre los límites de clase declarados o los límites de clase verdaderos. Se evalúa sumando los límites inferior y superior y dividiendo el total entre dos.

Medidas de Tendencia Central

La medida de tendencia central es un número que representa el valor central de un conjunto de valores. Estas medidas son denominadas habitualmente "promedios" y las que mayormente se emplean son: la mediana, la moda, La media aritmética (simple o ponderada)= la media geométrica y la media armónica.

El promedio es el termino comúnmente usado para definir que es una medida de tendencia central. Esto es que si el promedio es un valor representativo de un conjunto de datos, tendremos que el promedio podría determinar la localización del punto de tendencia central.

Las medidas de tendencia central que se estudiarán para datos simples y compuestos son:

1. Media Aritmética
2. Mediana
3. Moda
4. Otras medidas como déciles, cuartiles y perciles.

DATOS SIMPLES

La Media Aritmética

Conocida también como promedio esta representada por la suma de los valores individuales correspondientes a un fenómeno, dividida por el número de tales valores.

La Media es el más sencillo de los promedios de tal forma que para su calculo no es necesario ni siquiera el ordenamiento de datos.

La media aritmética es lo que viene a la mente de la mayoría de las personas cuando se menciona la palabra promedio. La media aritmética se calcula al sumas los valores de un conjunto y al dividir el producto de esta suma entre el número de valores del mismo. De este modo, la media de los valores 70, 80, 120 es.

Mediana

La segunda medida de tendencia central de un conjunto de números es la mediana. Su característica principal es que divide a un conjunto ordenado en dos grupos iguales; la mitad de los números tendrá valores que son menores que la mediana, y la otra mitad alcanzara valores mayores que está. Para encontrar la mediana, primeramente es necesario ordenar los valores (generalmente) de menor a mayor. Posteriormente se deberá separar la mitad de los valores para obtener la mediana.

El procedimiento para obtener la mediana es como sigue:

1. Ordenar o clasificar los valores
2. Contar para saber si existe un número de valores par o impar

3. En caso de que se tenga un número impar de valores, la mediana es valor intermedio. En cambio para un número par de valores, la mediana es el promedio de los valores intermedios.

La mediana de un conjunto de números es mayor que la mitad de los valores y menor que la otra mitad de los mismos.

Características de la Mediana

1. La Mediana es un promedio de posición
2. La Mediana esta afectada por el número de elementos y no por el tamaño de los valores extremos.
3. La suma de las desviaciones respecto a la mediana prescindiendo del signo serán menores que la suma de las desviaciones respecto a cualquier otro punto.
4. La mediana es un valor muy adecuado cuando se utiliza para describir distribuciones cuyos valores centrales están muy próximos
5. Un valor seleccionado al azar se situara por encima o por debajo de la mediana con idéntica probabilidad.

Por ello algunas veces, la mediana recibe el nombre de valor probable.

Ventajas

1. La Md se calcula fácilmente
2. Su valor no queda distorsionado por elementos inusuales
3. Algunas veces es un valor mas representativo de una serie que otros promedios gracias a su independencia de los valores inusuales
4. La Md puede calcularse aun cuando los intervalos de clase de la distribución sean limites abiertos.

Desventajas

1. La Md no es tan familiar como la X.
2. Los elementos deben disponerse según su tamaño antes de que pueda calcularse la Md.
3. Presenta un error tipo mayor que el de la X
4. La Md no puede manipularse algebraicamente.

El promedio de las Md de los subgrupos por ejemplo no es la Md del grupo.

Moda

La Medida de Tendencia central mas sencilla de determinar es la Moda. Se determina en una serie simple de acuerdo a la mayor cantidad de veces en que se repite algún elemento, por ejemplo, si se tiene la siguiente serie simple:

5 7 9 4 5 8 7 5 2 9

Entonces la Moda estará determinada por el No. 5

Si en la serie no se repite ninguno de sus elementos se dice que la serie no tiene moda.

Cuando en alguna serie se repite de igual manera dos o mas elementos la serie deja de ser unimodal (una moda) para convertirse en bimodal (dos modas), Trimodal (tres modas) o hasta convertirse en polimodal (muchas modas).

MEDIDAS DE TENDENCIA CENTRAL PARA DATOS AGRUPADOS.

La Formula para la serie de datos Agrupados son:

$$\text{Media Aritmética } X = \frac{\sum fx}{n}$$

Donde:

\sum sumatoria

fx es la multiplicación de la frecuencia por el punto medio.

n es el numero de datos

$$\text{Mediana } Md = Li + \left(\frac{N + \sum Fa}{fmd} \right) C$$

Donde:

Li : Limite real inferior de la clase Md

$N = \sum f$ = numero total de elementos

$\sum fa$ = Sumatoria de las frecuencias acumuladas hasta antes de la clase mediana.

fmd = frecuencia de la clase mediana

C = Intervalo de clase.

Moda

$$Mo = Li + \left(\frac{d1}{d1 + d2} \right) C$$

Donde:

Li : Limite inferior de la clase modal

$d1$ = frecuencia de la clase modal menos la frecuencia de la clase anterior

$d2$ = frecuencia de la clase modal menos la frecuencia de la siguiente clase.

C = Intervalo.

Medidas de Dispersión

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

Una medida de dispersión permite apreciar cuan dispersas están dos o más distribuciones.

Medidas para datos No Agrupados

a) Amplitud o Rango

La medida de dispersión más sencilla es la amplitud total. Se trata de la diferencia entre los valores mayor (o más alto) y menor (o más bajo) de un conjunto de datos.

$$\text{Amplitud Total} = \text{Valor más alto} - \text{Valor más bajo}$$

Ejemplo

1.- Las capacidades de varios recipientes metálicos son : 38,20,37,64 y 27 litros respectivamente. ¿Cuál es la amplitud total de esos valores?

b) Desviación Media

Un defecto importante de la amplitud total es que se basa solo en dos valores, el mayor y el menor, no toma en consideración todos los datos. La desviación media sí lo hace. Denominada también como desviación promedio, mide el promedio en donde los valores de una población, o muestra, varían con respecto a su media.

Desviación Media: Promedio de los valores absolutos de las desviaciones con respecto a la Media Aritmética

Su fórmula es:

$$DM = \frac{\sum |x - \bar{x}|}{n}$$

Donde:

X es el valor de cada observación

\bar{x} es la media aritmética de los valores

n es el número de observaciones en la muestra

| | este símbolo indica el valor absoluto. En otras palabras no se toman en cuenta los signos de las observaciones respecto a la media.

¿Porque no se consideran tales signos de las desviaciones? Si no se hiciera esto, las desviaciones positivas y negativas se compensarían exactamente y la desviación media siempre sería cero. Tal medida(cero) es un dato estadístico inútil. Debido a que se toman desviaciones absolutas, la desviación media suele denominarse *desviación media absoluta(D.M.A.)*

Ejemplos para serie simple

1.-Los pesos de una muestra de cajas listas para embarcarse a Francia son (en kilogramos): 103,97,101,106 y 103.

a) ¿Cual es la desviación media?

Pesos en Kg.	X-X (desviaciones absolutas)
97	5
101	1
103	1
103	1
106	4

2.- Calcule la Desviación Media del siguiente grupo de datos.

X	X - X
2	2 - 9.75 = 7.75
3	3 - 9.75 = 6.75
5	5 - 9.75 = 4.75
7	7 - 9.75 = 2.75
10	10 - 9.75 = 0.25
13	13 - 9.75 = 3.25
18	18 - 9.75 = 8.25
20	20 - 9.75 = 10.25

c) Desviación Estándar

La Desviación Estándar es una forma especial de desviación promedio respecto a la media.

Su formula:

$$\sigma = \sqrt{\frac{\sum |x - \bar{x}|}{n}}$$

1.- Ejercicio: Con los datos anteriores, determine la Desviación Estándar.

X - X	(X-X) ²
7.75	
6.75	
4.75	
2.75	
0.25	
3.25	
8.25	
10.25	
Σ	

e) Coeficiente de Variación

Si la Dispersión absoluta es la Desviación Estándar (S), la Dispersión Relativa recibe el nombre de Coeficiente de Variación.

Su formula es :

$$cv = \frac{\sigma}{\bar{x}}$$

Medidas de Dispersión para Datos Agrupados

a) Desviación Media Absoluta: Para determinar la desviación media absoluta, se requiere una serie de pasos:

- 1.- Determinar la Media Aritmética o Mediana de la serie
- 2.- Determinar la diferencia entre el $PM(X)$ y el valor de la Media o Mediana
- 3.- Obtener el producto de la frecuencia correspondiente a cada diferencia.
- 4.- Determinar la suma de los valores absolutos del punto anterior
- 5.- Obtener el promedio de la suma del punto anterior.

La formula para datos agrupados es:

$$DM = \frac{\sum fx |x - x|}{n}$$

b) Desviación Estándar

Su formula para datos agrupados es:

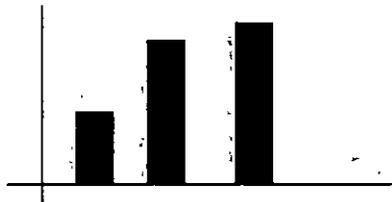
$$\sigma = \sqrt{\frac{\sum fx(x - x)^2}{n}}$$

REPRESENTACION GRAFICA

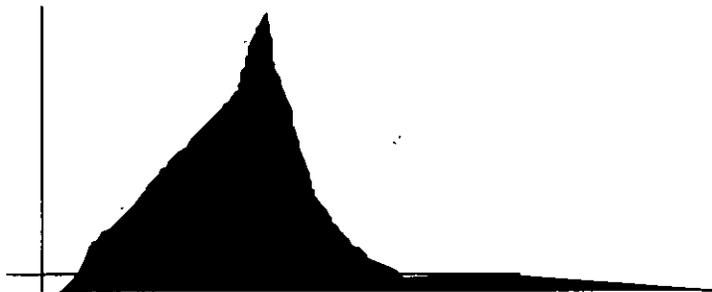
Una representación gráfica de los datos ayuda a que se distingan en forma inmediata algunas de las características mas importantes de las observaciones.

HISTOGRAMA

Un Histograma es una gráfica de barras de una distribución de frecuencias.

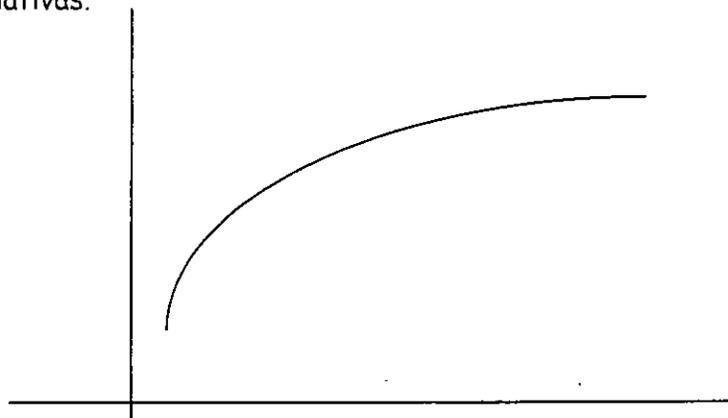


Un Polígono de frecuencias es una gráfica de líneas de una distribución de frecuencias.



Una curva de frecuencias es un polígono de frecuencias

Una Ojiva es una gráfica construida con segmentos de líneas rectas que unen los puntos obtenidos al colocar en el eje horizontal a los límites superiores reales de clase y en el vertical a las frecuencias acumuladas absolutas o relativas.



Una distribución de frecuencias acumuladas identifica el número acumulado de observaciones incluidas bajo el límite exacto superior de cada clase de la distribución.

Las frecuencias acumuladas de una clase pueden determinarse sumando las frecuencias observadas de esa clase a las frecuencias acumuladas de la clase anterior.

Una distribución de frecuencias relativas es aquella cuyo número de observaciones asociadas con cada clase se ha convertido en una frecuencia relativa dividiendo entre el número total de observaciones de toda la distribución.