




ESTADÍSTICA BÁSICA

Dirección Redes en Salud Pública


2015 – 09 – 16

CONCEPTOS BÁSICOS

Es el conjunto sistemático de procedimientos para la observación, registro, organización, síntesis y análisis e interpretación de los fenómenos y de las leyes que los regulan para poder así predecir o concluir acerca de ellos. Esta definición claramente involucra las dos fases de la estadística: la descriptiva y la inferencial.



Descriptiva (Deductiva). Es la fase de descripción, organización, síntesis y análisis de la información de interés.



Inferencial (Inductiva). Esta fase busca obtener conclusiones sólidas y más profundas que una simple descripción de la información basados en el trabajo con muestras y su posterior generalización de resultados para la toma de decisiones y conclusiones sólidas.

POBLACIÓN

Conjunto total de elementos (datos, personas, objetos, medidas...) que poseen una o más características observables sobre las cuales se buscan conclusiones y decisiones.

Población Colombiana

CONCEPTOS BÁSICOS

MUESTRA

Subconjunto de elementos pertenecientes a la población bajo estudio, seleccionados en forma predefinida

Población de mujeres colombianas menores de 35 años



CONCEPTOS BÁSICOS

PARÁMETRO

Es un número que resume la gran cantidad de datos que puede derivarse del estudio de una variable estadística

Generalmente se describen en letra griega

μ = promedio
 σ = desviación

ESTADÍSTICO – PARÁMETRO



ESTADÍSTICO

Todo valor calculado con base en la muestra.

Generalmente se describen en letra romana

x = promedio
 s = desviación

ERROR DE MUESTREO

ESTUDIO ESTADÍSTICO

PASO 1:
RECOGIDA DE
DATOS

PASO 2:
ORGANIZACIÓN
DE LOS DATOS

PASO 3:
ANÁLISIS FINAL

VARIABLES

CUANTITATIVAS

Cuyas propiedades pueden presentarse en diversos grados o intensidades de carácter numérico

CARACTERÍSTICA OBSERVABLE Y MEDIBLE QUE VARIA ENTRE LOS DIFERENTES INDIVIDUOS DE UNA POBLACIÓN

CUALITATIVAS

No susceptibles de observación medible numéricamente

DISCRETAS: Toman valores enteros, aislados.

No pueden tomar valores intermedios entre dos valores dados

N. de Hijos

CONTINUAS: Toman valores numéricos y entre dos de ellos hay infinitos valores intermedios.

Puede tomar cualquier valor dentro de un recorrido dado

Peso

NOMINALES: Sus valores no se pueden ordenar

Profesión

ORDINALES: Sus valores se pueden ordenar

Grado de Satisfacción

ESCALAS DE MEDIDA

Para categorizar o clasificar variables , se utilizan diferentes tipos de escala, su uso depende de los objetivos del estudio y de la naturaleza de la variable.

ESCALAS NOMINALES

- Se usan para el nivel mas simple de medición cuando los datos se agrupan en categorías.
- Los datos nominales se describen en términos de porcentajes y proporciones
- No se miden, se cuentan sus observaciones
- Se utilizan tablas de contingencia y gráficas de barras para mostrar su información
- Pueden ser dicotómicas o multinomiales

EJEMPOS: GÉNERO, RAZA, RELIGIÓN

ESCALAS DE MEDIDA

ESCALAS ORDINALES

- Las observaciones se clasifican en categorías como las nominales pero algunas observaciones son más o mayores que otras.
- Se describen en términos de porcentajes y proporciones. A veces se pueden resumir en un valor medio
- Se utilizan tablas de contingencia y gráficas de barras para mostrar su información

EJEMPOS: SEVERIDAD DE LA ENFERMEDAD (SEVERO, MODERADO, SANO), NIVEL DEL DAÑO DE UNA PIEZA (LEVE, MODERADO, GRAVE)

ESCALAS DE MEDIDA

ESCALA DE INTERVALO

- Representa un nivel de medición más preciso, matemáticamente hablando, que las anteriores
- Los datos tienen propiedades de dato ordinal y se pueden determinar distancias.
- No hay un cero absoluto o real, el cero es arbitrario.

ESCALA DE RAZÓN

- Los datos tienen propiedades de datos de intervalo y existe un cero real que permite considerar cocientes de mediciones.

TALLER

***“Si oigo algo lo olvido.
Si lo veo lo entiendo.
Si lo hago lo aprendo”.***

Confucio (551-478 A.C)

EJERCICIO 1

DIAGRAMA DE PUNTOS O "DOTPLOT"

Se construye colocando determinado número de puntos sobre un eje horizontal de acuerdo al número de veces que se repite el dato.

Útil cuando el conjunto de datos es razonablemente pequeño o presenta pocos valores diferentes

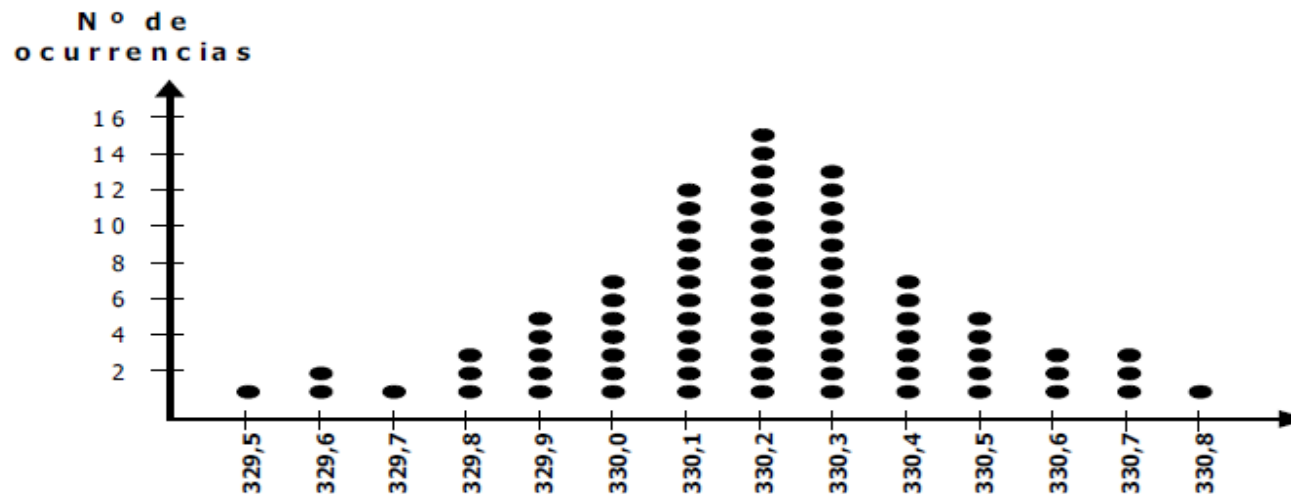


DIAGRAMA DE TALLOS Y HOJAS

Se requiere que los datos estén conformados por al menos dos dígitos.

El último dígito constituye la hoja y el ó los restantes conformarán el tallo.

Para una adecuada descripción de los datos es conveniente trabajar con al menos 4 tallos.

EJEMPLO: Se presenta el porcentaje de algodón en un material usado para la fabricación de camisas

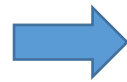


Tabla 1. Datos del porcentaje de algodón

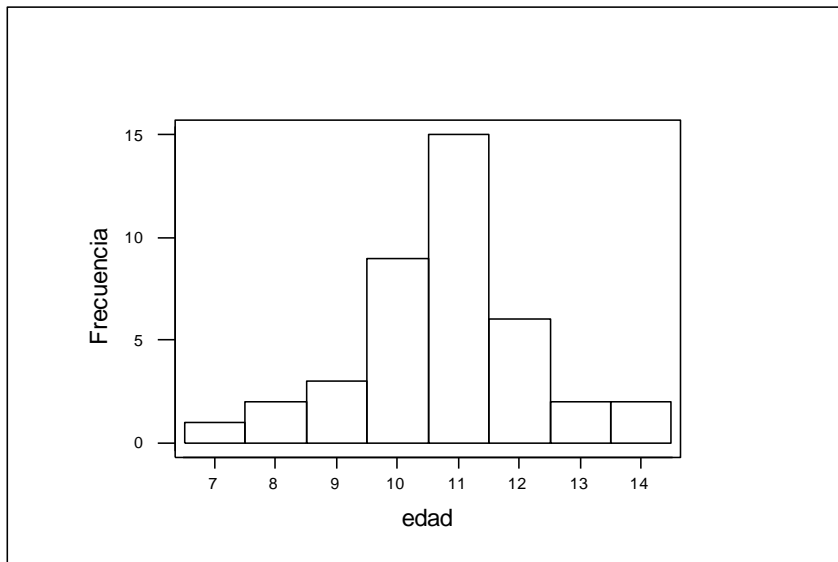
| | | | | | | | |
|------|------|------|------|------|------|------|------|
| 33.1 | 35.3 | 34.2 | 33.6 | 33.6 | 33.1 | 37.6 | 33.6 |
| 34.5 | 34.7 | 33.4 | 32.5 | 35.4 | 34.6 | 37.3 | 34.1 |
| 35.6 | 35.0 | 34.7 | 34.1 | 34.6 | 35.9 | 34.6 | 34.7 |
| 36.3 | 35.4 | 34.6 | 35.1 | 33.8 | 34.7 | 35.5 | 35.7 |
| 35.1 | 36.2 | 35.2 | 36.8 | 37.1 | 33.6 | 32.8 | 36.8 |
| 34.7 | 36.8 | 35.0 | 37.9 | 34.0 | 32.9 | 32.1 | 34.3 |
| 33.6 | 35.1 | 34.9 | 36.4 | 34.1 | 33.5 | 34.5 | 32.7 |
| 32.6 | 33.6 | 33.8 | 34.2 | 34.6 | 34.7 | 35.8 | 37.8 |



| Tallo Hojas | |
|-------------|-----------------------|
| 32 | 156789 |
| 33 | 114566666688 |
| 34 | 011122355666667777779 |
| 35 | 00111234456789 |
| 36 | 234888 |
| 37 | 13689 |

HISTOGRAMAS

- Permite visualizar la forma de la distribución de una variable a partir de los valores obtenidos en una muestra de tamaño adecuado (generalmente no menor a 50 unidades)
- Para variables discretas se denomina diagrama de barras
- Para variables continuas se conoce como histograma
- En todos los casos un histograma se construye graficando las frecuencias relativas y las frecuencias absolutas



FRECUENCIA ABSOLUTA: Es el número de veces que ocurre un evento determinado

FRECUENCIA RELATIVA: Es el cociente entre la frecuencia absoluta y el tamaño de la muestra (número de eventos posible)

PRESENTACIÓN DE DATOS CUANTITATIVOS

Como construir un histograma?

Paso 1

- Determinar el rango de los datos. (Dato Mayor menos dato menor)

Paso 2

- Obtener los números de clases: regla de Sturges ($1 + 3,3 \log (n)$) o una regla razonable es elegir \sqrt{n} para el número de clases

Paso 3

- Establecer la longitud de clases: rango dividido por el número de clases

Paso 4

- Construir los intervalos de clases: resultan de dividir el rango de los datos en relación con el paso 2 en intervalos diferentes

Paso 5

- Graficar el histograma

EJEMPLO

Datos: número de espectadores en
32 partidos de Colombia (en miles)

| | | | | | | | |
|------|------|------|------|------|------|------|------|
| 42,1 | 51,0 | 30,0 | 35,2 | 29,3 | 10,9 | 16,1 | 51,6 |
| 47,0 | 51,4 | 35,2 | 31,7 | 17,8 | 67,0 | 43,2 | 23,7 |
| 25,2 | 36,1 | 32,3 | 51,7 | 46,0 | 12,2 | 21,1 | 29,0 |
| 14,3 | 47,2 | 31,3 | 35,4 | 29,1 | 23,0 | 10,3 | 34,2 |

Paso 1

- Determinar el rango de los datos. (Dato Mayor menos dato menor)

$$\text{RANGO} = 67,0 - 10,3 = 56,7$$

EJEMPLO

Paso 2

- Obtener los números de clases: regla de Sturges ($1 + 3,3 \log (n)$) o una regla razonable es elegir \sqrt{n} para el número de clases

$$\sqrt{32} = 5,6$$

Paso 3

- Determine los límites de cada clase, de forma tal que el menor valor quede incluido en la primera clase y el mayor valor en la última

$$\text{Longitud de clases} = 56,7 / 5,6 = 10,1 \approx 10$$

EJEMPLO

Paso 4

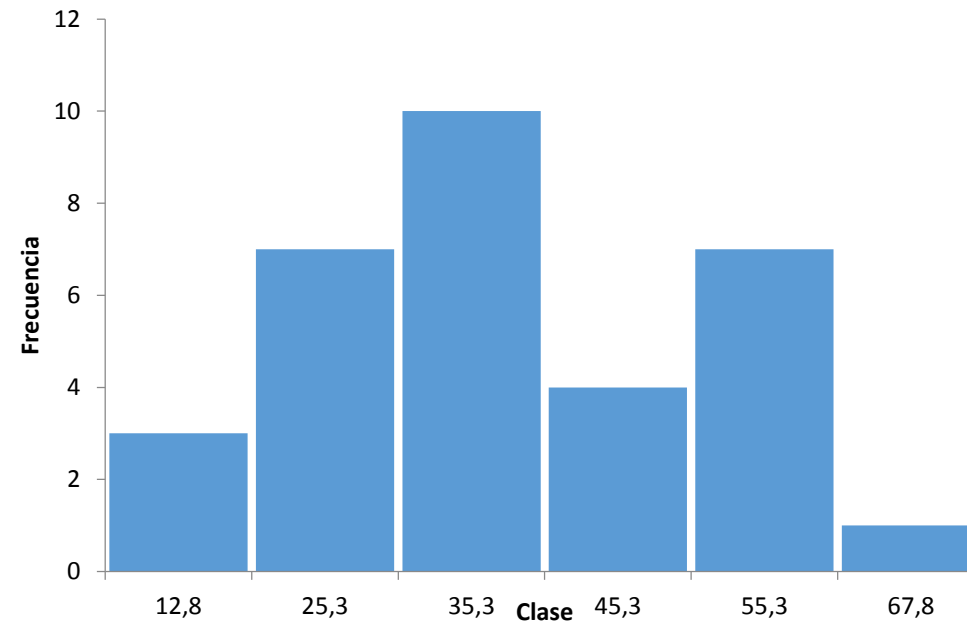
- Construir los intervalos de clases de forma tal que el menor valor quede incluido en la primera clase y el mayor valor en la última clase. Por regla general los límites de clase deben estar definidos con un decimal más que los datos

| LIMITE INFERIOR | LIMITE SUPERIOR | FRECUENCIA |
|-----------------|-----------------|------------|
| 5,3 | 20,3 | 4 |
| 20,3 | 30,3 | 6 |
| 30,3 | 40,3 | 10 |
| 40,3 | 50,3 | 4 |
| 50,3 | 60,3 | 7 |
| 60,3 | 75,3 | 1 |

Paso 5

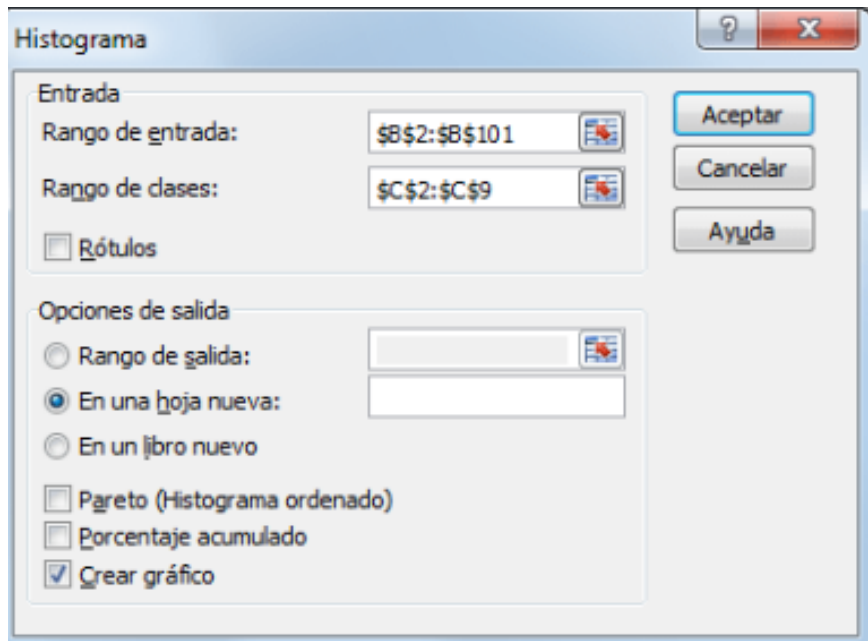
- Graficar el histograma

Histograma



PRESENTACIÓN DE DATOS CUANTITATIVOS

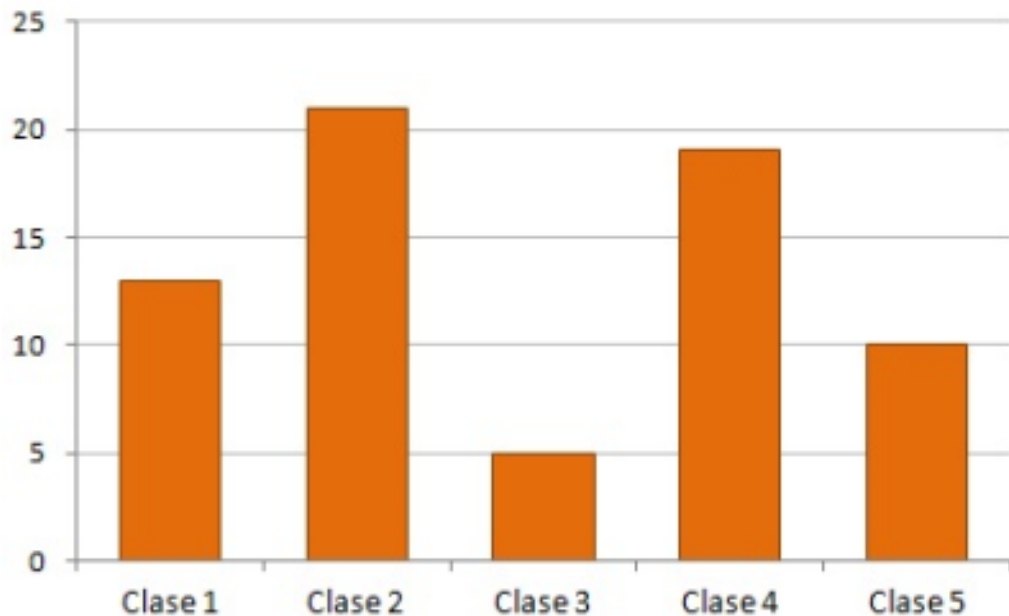
Como construir un histograma en excel?



The screenshot shows the 'Histograma' dialog box in Microsoft Excel. It is divided into two main sections: 'Entrada' and 'Opciones de salida'. In the 'Entrada' section, 'Rango de entrada' is set to '\$B\$2:\$B\$101' and 'Rango de clases' is set to '\$C\$2:\$C\$9'. There are three buttons on the right: 'Aceptar', 'Cancelar', and 'Ayuda'. In the 'Opciones de salida' section, there are three radio buttons: 'Rango de salida', 'En una hoja nueva' (which is selected), and 'En un libro nuevo'. There are also four checkboxes: 'Rótulos' (unchecked), 'Pareto (Histograma ordenado)' (unchecked), 'Porcentaje acumulado' (unchecked), and 'Crear gráfico' (checked).

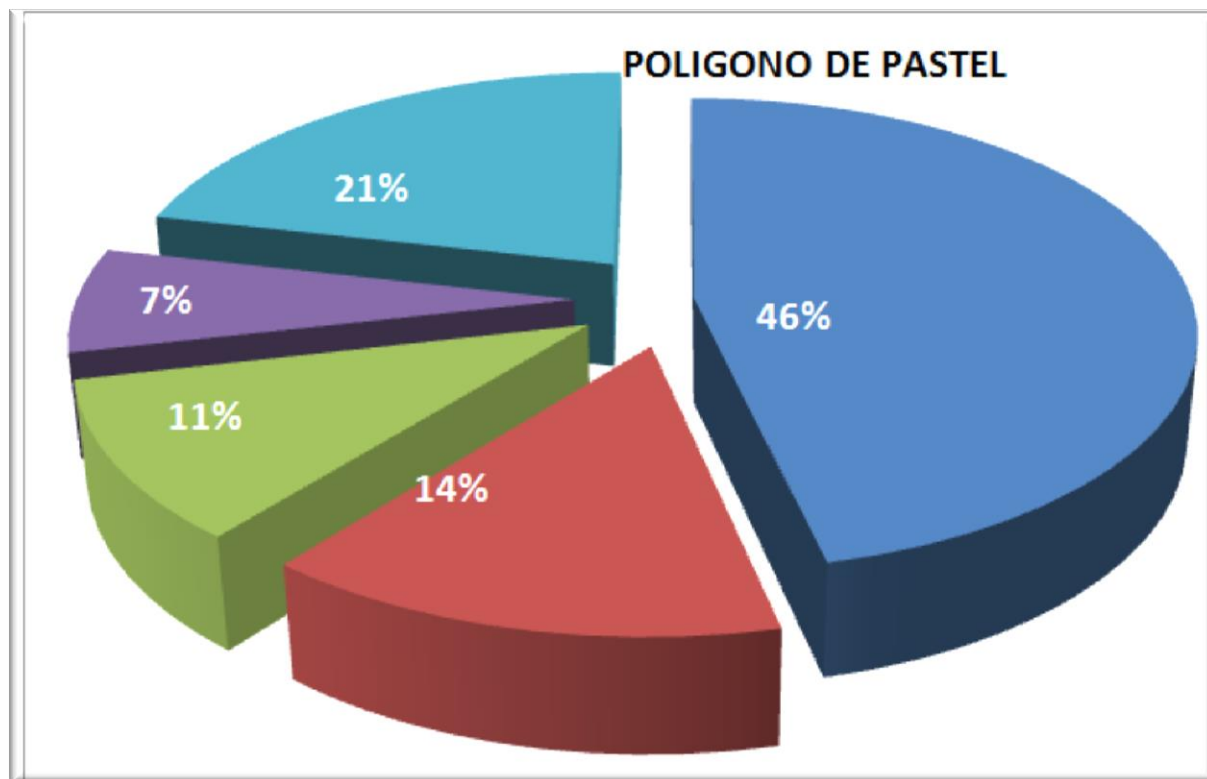
RANGO DE CLASES: INTERVALO DIVIDIDO EN DOS

PRESENTACIÓN DE DATOS CUALITATIVOS



- Se basa en las tablas de frecuencia (conteo del número de elementos o individuos que tienen determinada característica).
- Categóricas cualitativas barras horizontales
- Categóricas cuantitativas barras verticales
- PARETO (Diagrama de Barras Ordenado)

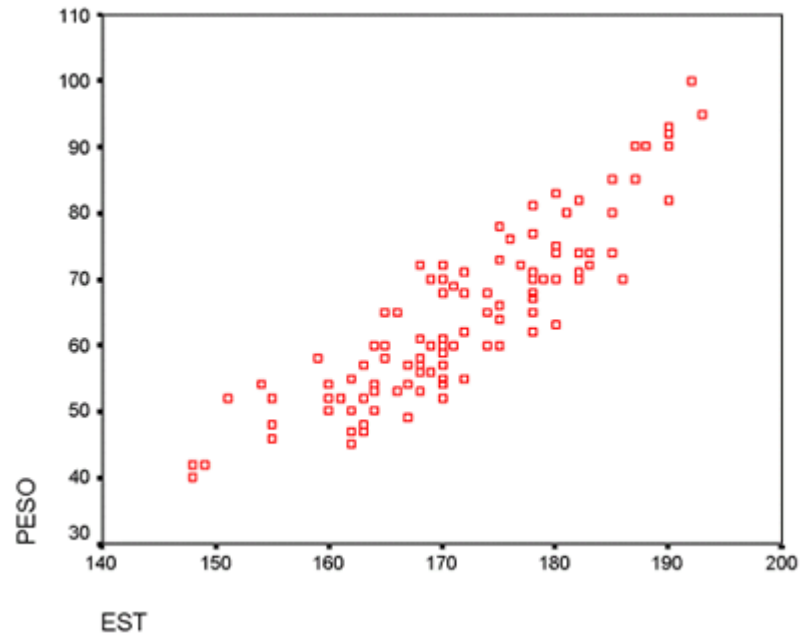
DIAGRAMA DE BARRAS



- Permiten visualizar mejor la proporción en que aparece una característica respecto del total.

DIAGRAMAS DE PASTEL

DESCRIPCIÓN DE DOS VARIABLES



- Consiste en un gráfico en el plano cartesiano que muestra la relación entre dos variables.

DIAGRAMA DE DISPERSIÓN

MEDIDAS DE TENDENCIA CENTRAL

Propiedades

1. La suma de las desviaciones respecto de la media es igual a cero
2. Si se tiene la media de un conjunto de datos y a cada observación se multiplica por una constante b y se le suma una constante a , entonces la nueva media de los datos se obtiene multiplicando la media de los datos originales por b y sumándole a .

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

La Media Aritmética

MEDIDAS DE TENDENCIA CENTRAL

- Poca variabilidad de una muestra a otra
- No debe emplearse con datos ordinales
- Es sensible a valores externos

La Media Aritmética

MEDIDAS DE TENDENCIA CENTRAL

- La Mediana es el valor que divide un conjunto de datos ordenado en dos partes porcentualmente iguales

Datos Cuantitativos ordenados de menor a mayor

| |
|-----------|
| x |
| $x_{(1)}$ |
| $x_{(2)}$ |
| \vdots |
| $x_{(n)}$ |

Mediana

$$M_E = x_{(k)} \quad \text{Si } n \text{ es impar}$$

$$M_E = \frac{x_{(k)} + x_{(k+1)}}{2} \quad \text{Si } n \text{ es par}$$

$$x_{(k)} = \text{dato del centro}$$

- Es menos sensible a valores extremos
- Puede usarse con observaciones ordinales

Mediana

MEDIDAS DE TENDENCIA CENTRAL – (ejemplo)

| DATOS | |
|-------|------|
| 7,79 | 7,73 |
| 7,60 | 7,29 |
| 7,80 | 9,04 |
| 9,88 | 8,21 |
| 6,88 | |

$$\bar{X} = \frac{7,79 + 7,60 + 7,80 + 9,88 + 7,73 + 7,29 + 9,04 + 8,21 + 6,88}{9} = 8,02$$

| DATOS ORDENADOS |
|--------------------|
| 6,88 |
| 7,29 |
| 7,60 |
| 7,73 |
| 7,79 |
| 7,80 |
| 8,21 |
| 9,04 |
| 9,88 |

MEDIANA = 7,79

MEDIDAS DE TENDENCIA CENTRAL

- Se define como el valor que se presenta con mayor frecuencia
- Para el caso de un conjunto de datos se puede presentar una moda (unimodal) o dos modas (bimodal)
- Es muy usada para datos cualitativos

Moda

DESCRIPTORES CUANTITATIVOS



TENGA EN CUENTA:

- 1) LA ESCALA DE MEDICIÓN : ordinal o numérica**
- 2) LA FORMA DE DISTRIBUCIÓN DE RESULTADOS:
simetría, curtosis...**

MEDIA: Para datos numéricos o distribuciones simétricas

MEDIANA: Para datos ordinales o numéricos, si hay datos anómalos

MODA: Para datos cualitativos o distribuciones bimodales

MEDIDAS DE TENDENCIA CENTRAL

- **Media Ponderada** = Promediar observaciones con diferentes importancias o pesos
- **Media Geométrica** = Promediar porcentajes, índices y cifras relativas
Determinar el incremento porcentual promedio en ventas , producción u otras actividades o series económicas de un periodo a otro
- **Media Armónica** = Generalmente se utiliza para promediar variaciones con respecto al tiempo

MEDIDAS DE DISPERSIÓN

Determinan nivel de concentración de un conjunto de datos

Más variación = Heterogeneidad

Menos variación = Homogeneidad

- **ABSOLUTAS: Varianza, Desviación Estándar, Rango**
- **RELATIVAS : Coeficiente de Variación**

MEDIDAS DE DISPERSIÓN

- Es la medida de dispersión de los datos más simple de calcular.
- Se expresa en la mismas unidades de la muestra.
- Su principal inconveniente es que solo depende de los valores extremos de la muestra.
- Es poco confiable en presencia de datos “anómalos”

$$\text{Rango} = X_{\max} - X_{\min}$$

RANGO

MEDIDAS DE DISPERSIÓN

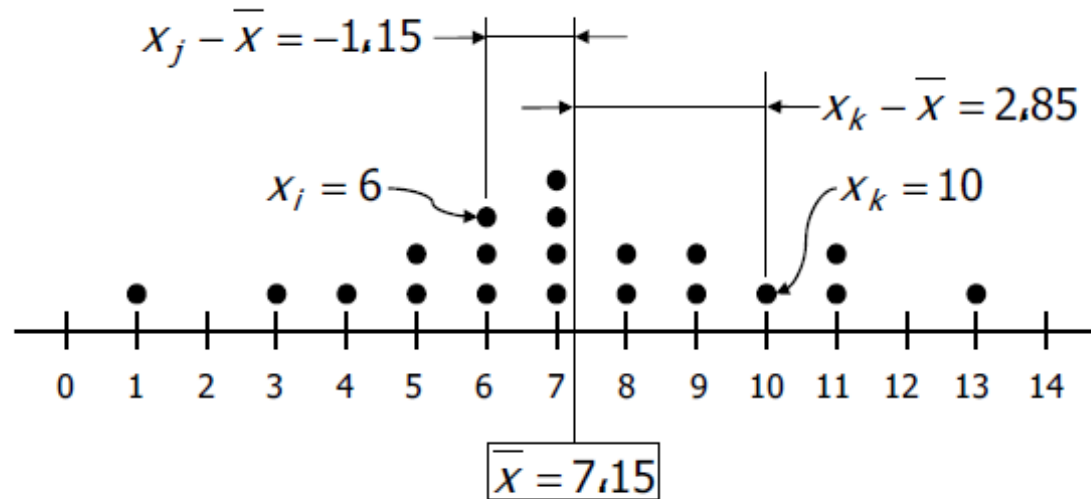
Se define como el promedio (poblacional) ó el "casi promedio" (muestral) de los cuadrados de las desviaciones de los datos con respecto a la media muestral.

Propiedades de la varianza

- La varianza de una constante es cero
- Si se tiene la varianza de un conjunto de datos y cada observación se multiplica por una constante b , entonces la nueva varianza de los datos se obtiene multiplicando la varianza de los datos originales por b^2 .

VARIANZA

MEDIDAS DE DISPERSIÓN



$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Esta dada en unidades de la muestra al cuadrado

VARIANZA

MEDIDAS DE DISPERSIÓN

La desviación estándar está definida como la raíz de la Varianza

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

DESVIACIÓN ESTÁNDAR

MEDIDAS DE DISPERSIÓN

Se utiliza ante la presencia de datos atípicos y se define como la mediana del valor absoluto de las diferencias de los datos respecto a su mediana

$$MEDA = \text{Mediana}|Mediana - x_i|$$

MEdiana de las Desviaciones Absolutas (MEDA)

MEDIDAS DE DISPERSIÓN

Es utilizado para comparar la variabilidad entre dos grupos de datos que tienen distinta media o referidos a distintos sistemas de unidades de medida. Por ejemplo, kilogramos y centímetros.

$$CV_{Muestral} = \frac{s}{\bar{x}} \times 100\%$$

COEFICIENTE DE VARIACIÓN

MEDIDAS DE DISPERSIÓN

Es utilizado para comparar la variabilidad entre dos grupos de datos que tienen distinta media o referidos a distintos sistemas de unidades de medida. Por ejemplo, kilogramos y centímetros.

$$CV_{Muestral} = \frac{s}{\bar{x}} \times 100\%$$

COEFICIENTE DE VARIACIÓN

MEDIDAS DE FORMA

Coefficiente de Asimetría

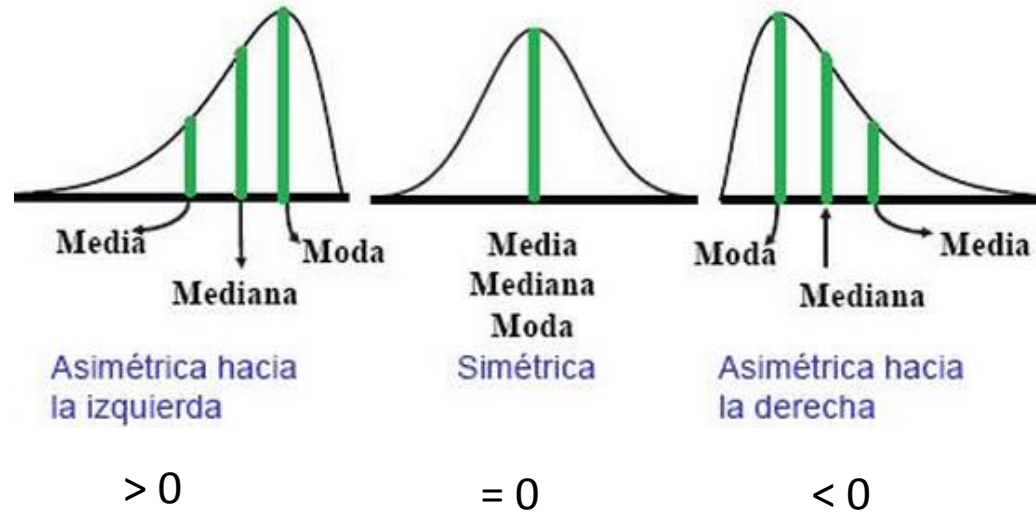
$$CA = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n \cdot s^3}$$

- Si $CA=0$ si la distribución es simétrica alrededor de la media.
- Si $CA<0$ si la distribución es asimétrica a la izquierda
- Si $CA>0$ si la distribución es asimétrica a la derecha

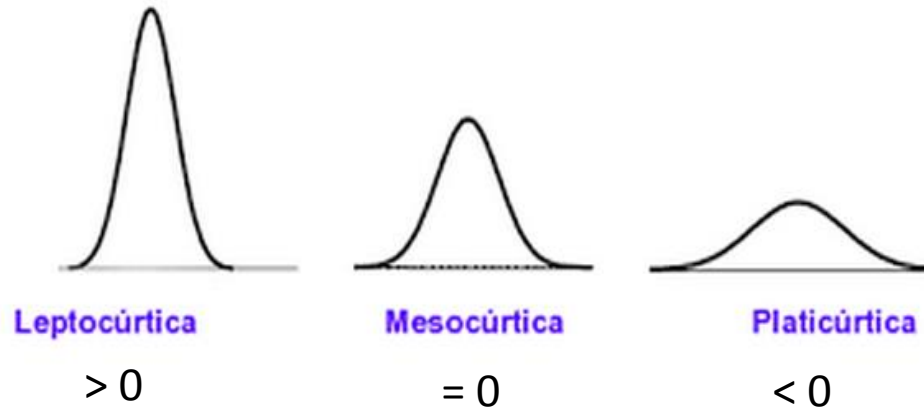
Coefficiente de Apuntamiento

$$CAp = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n \cdot s^4}$$

- Si $CAp=0$ la distribución se dice normal (similar a la distribución normal de Gauss) y recibe el nombre de *mesocúrtica*.
- Si $CAp>0$, la distribución es más puntiaguda que la anterior y se llama *leptocúrtica*, (mayor concentración de los datos en torno a la media).
- Si $CAp<0$ la distribución es más plana y se llama *platicúrtica*.



ASIMETRIA



CURTOSIS



GRACIAS

Dirección de Redes
Subdirección de Calidad

Olga Avila
oavila@ins.gov.co

Instituto Nacional de Salud
Correo electrónico: contactenos@ins.gov.co
Teléfono: (1) 220 7700 Ext. 1703 – 1704
fax 220 7700 Ext. 1283 – 1269
Bogotá, COLOMBIA
www.ins.gov.co
Línea gratuita nacional: 01 8000 113 400