

## Guía de Probabilidad y Estadística

Para poder realizar los ejercicios propuestos es necesario conocer las definiciones de los siguientes conceptos:

### **Población**

La **población o Universo** es todo conjunto de personas, cosas, objetos, etc., con ciertas características comunes que puede ser finita o infinita.

### **Muestra**

La **Muestra** es toda porción de elementos tomada de una población, entre más grande sea la muestra, será más representativa de la población.

El **Muestreo** es el procedimiento por el cual se recopila información de los elementos de una muestra, existen varias formas de realizar el muestreo, unas son más sencillas que otras, otras más económicas y otras más representativas.

### **Tipos de variables**

Una **Variable** se puede representar con un símbolo tal como  $x$ ,  $y$ ,  $a$ ,  $b$ , que puede tomar un valor cualquiera. Si esta variable solamente toma un valor, se le llama **Constante**.

### **Variables cualitativas**

Las variables cualitativas son aquellas que describen las cualidades o características de un objeto o evento mediante palabras, frases, letras, etc. Son subjetivas de acuerdo con el observador. Por

**Variables Ordinales** clasifican las observaciones en categorías que exigen ordenación, su variable operacional es una escala ordinal de mayor a menor.

### **Variables cuantitativas**

Las **variables cuantitativas** son aquellas que se pueden representar por una cantidad numérica exacta, por lo general son medidas objetivas, es decir que no depende de quien realice la medición. En ocasiones puede tener un pequeño error dependiendo del instrumento con que se mida.

Las variables cuantitativas se clasifican en:

### **Variables Discretas**

**Variables Discretas** son las que toman solo algunos valores dentro de un intervalo, por ejemplo, valores enteros. Al indicar el número de hijos que tiene una familia, el número de recámaras de una casa, el número de mascotas, etc.

### **Variables Continuas**

**Variables Continuas** Son las que pueden tomar cualquier valor fijo dentro de un intervalo, siempre entre dos valores observables, va a existir un tercer valor intermedio, que también podría tomar la variable continua.



Obtener el Rango, los intervalos, la marca de clase y obtén la tabla de distribución de Frecuencias.

El *Recorrido o Rango* es la diferencia que existe entre el mayor y menor de los datos:

$$Rango = D_{mayor} - D_{menor}$$

Los *intervalos de clase* o *intervalo exacto* son aquellos donde pueden agruparse los datos de una variable estadística. Es conveniente que cuando se seleccione el número de intervalos o de clases NC, éste no sea muy pequeño o muy grande, esto depende de la experiencia del analista que realice la tabla. En ocasiones, se toma el criterio de la raíz cuadrada del número de datos (N), en otras la raíz cúbica o la regla de Storges que está dada por la ecuación:

$$NC = 1 + 3.3n$$

Se define como el punto medio de un intervalo de clase y se representa por (X). La marca de clase también se denomina *punto medio de la clase*, y matemáticamente se determina por la suma de los límites inferior y superior del intervalo de clase, dividida entre dos. Lo anterior se simboliza por la ecuación:

$$Marca\ de\ clase\ (X) = \frac{Límite\ inferior\ (L_i) + límite\ superior\ (L_s)}{2}$$

Dibuja el Diagrama de Sectores (Pastel), la Gráfica de Barras (Histograma) y el Polígono de frecuencias.

### Gráfico Circular, de sectores o Pastel

Es una forma alternativa del gráfico de barras, en su construcción se utiliza una circunferencia dividida en sectores angulares proporcional al valor de la variable. Es conveniente usar las fra para facilitar su trazo



Otra opción es utilizar las frecuencias relativas acumuladas y utilizar el 100% como toda la circunferencia.

### Histograma.

Un histograma es la representación de la información mediante una gráfica en forma de barras, sin espacios entre ellas, esta imagen nos permite apreciar visualmente la distribución de un conjunto de datos. Los histogramas son de utilidad para darnos una idea general de la distribución de una población o muestra.

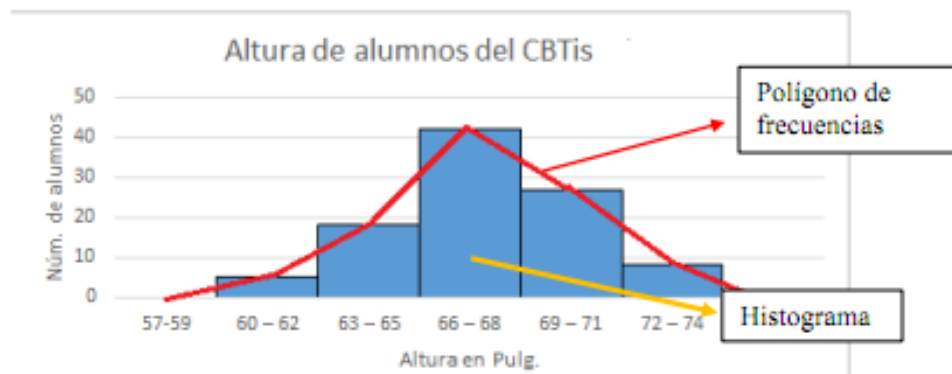
### Gráfico de Barras.

La construcción de este gráfico se basa en la representación de un valor numérico por un rectángulo. Las barras son usualmente verticales, pero también pueden ser horizontales.



### Polígono de Frecuencias

Para complementar el histograma, podemos unir los centros de las barras sobre el eje "x", es decir las marcas de clase, además de prolongar estas marcas a línea inferior y superior. Al conjunto de esta poligonal se le denomina "polígono de frecuencias".



### Media

Un valor que es simbólico y representativo de un conjunto de datos se denomina promedio. Se hace notar que en un conjunto de datos ordenados de acuerdo con su magnitud, el promedio siempre tiende a situarse en el centro de dicho conjunto, razón por la cual los promedios se denominan también medidas de centralización o de tendencia central.

### Media aritmética

Es el promedio más utilizado y por lo general se denomina media. La media aritmética o media de un conjunto de elementos se define como la suma de los valores de estos elementos dividido entre el número total de ellos. Se simboliza  $\bar{X}$  (léase  $\bar{X}$  barra), para **datos no agrupados**, matemáticamente se determina por la ecuación:

$$\bar{X} = \frac{\sum X_i}{N} = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N}$$

Donde:

$\bar{X}$  = Media aritmética o media.

$X$  = Valor de cada elemento.

$N$  = Número total de elementos.

## **Mediana**

Se define como el **valor medio** a aquel que divide a un conjunto de datos que se ordena de acuerdo con su magnitud (de forma ascendente o de forma descendente) en dos partes iguales, es decir, es aquel valor central que deja por debajo igual número de elementos que por arriba de él.

Si el conjunto de elementos ordenados de acuerdo con su magnitud, es **impar**, la mediana será el valor intermedio de dicha sucesión. Si el conjunto de elementos ordenados de acuerdo con su magnitud es **par**, la mediana será la media aritmética de los dos elementos medios.

## **Moda**

Se define como el valor que se presenta con la **mayor frecuencia**, es decir, es el **valor más común** de un conjunto de elementos numéricos dado.

Entre las características de la moda destacan que ésta puede no existir, incluso si existe puede no ser única. Si un conjunto de valores dado presenta una sola moda se denomina **unimodal**. Si presenta dos modas se denomina **bimodal**; si presentan más de dos modas se denomina **multimodal**.