

# ESTADISTICA UNIDAD 2



**Josefa Solis**

## Unidad 2: Organización Y Representación Gráfica

En una distribución de frecuencias, las observaciones se clasifican de acuerdo a categorías de valores. En otras palabras, una distribución de frecuencias nos indica cuántas veces se observó cada categoría. En otras palabras, una distribución proporciona datos sobre el número de observaciones que corresponde a cada categoría. En el caso de las variables discretas, esto es más intuitivo, pero también se aplica a las variables continuas.



## 2.2.-Representación Gráfica de Una Distribucion de Frecuencias

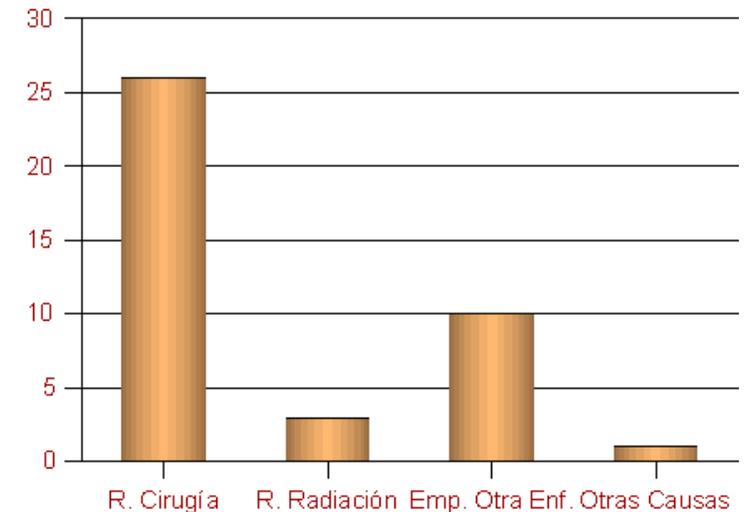
La representación gráfica de una distribución de frecuencias depende del tipo de datos que la constituya.

Datos correspondientes a un carácter cualitativo

La representación gráfica de este tipo de datos está basada en la proporcionalidad de las áreas a las frecuencias absolutas o relativas. Veremos dos tipos de representaciones:

Diagrama de sectores:

Está representación gráfica consiste en dividir un círculo en tantos sectores circulares como modalidades presente el carácter cualitativo, asignando un ángulo central a cada sector circular proporcional a la frecuencia absoluta  $n_i$ , consiguiendo de esta manera un sector con área proporcional también a  $n_i$ .



## 2.3.- Propiedades

Es un valor de la variable que se encuentra hacia el centro de la distribución de frecuencias. A este valor se le denomina promedio y es un valor que sintetiza a todos los valores de la distribución.

Variabilidad

Sesgo O Asimetría

Curtosis O Apuntamiento

Distribución de frecuencias a la agrupación de datos en categorías mutuamente excluyentes que indican el número de observaciones en cada categoría.

Esto proporciona un valor añadido a la agrupación de datos.

## 2.4.- Estadísticos de Posición Grupal

Se les denomina aquí como estadísticos de posición grupal porque sintetizan la información de todos los datos de una variable, esto es, para todo el grupo de casos a partir de los que esos datos han sido recogidos.

La moda de una variable  $X$  ( $MoX$ ) es un estadístico de tendencia central que se obtiene como el valor que más se repite en el conjunto de datos correspondiente a una variable, esto es, el que tiene la frecuencia absoluta ( $n_i$ ) más alta en su distribución de frecuencias.

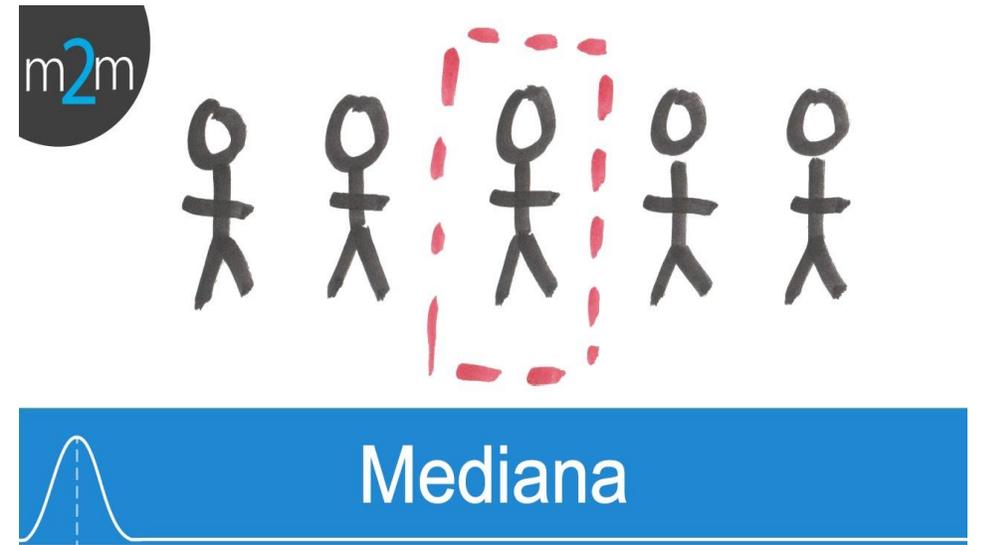


## La Mediana.

- La Mediana puede ser aplicada también en variables cuantitativas, si bien, no es el índice estadístico que mejor resume la tendencia central de este tipo de variables.
- La mediana de una variable  $X$  ( $MdnX$ ) es el valor del dato que, tras ordenar todos los datos de la variable, deja al mismo número de ellos por debajo que por encima. En el caso de un número par de datos, se obtiene el promedio de los valores de los datos centrales.

## El Mínimo y El Máximo

- El mínimo es, de las modalidades que adopta una variable, aquella cuyo valor sea más bajo en la escala de medida de la variable; complementariamente, el máximo será aquella cuyo valor sea más alto. Ambos valores permiten hacerse una idea de entre qué valores se localizan los datos de una variable.



## Los Cuantiles

- Se define como cuantil  $k$  ( $C_k$ ) al valor de la variable tal que un  $k\%$  de los sujetos tienen un valor inferior o igual a ese valor ( $k$  puede ser cualquier número entre 0 y 100, entero o decimal).
- La mediana es un caso particular de cuantil, en concreto, el cuantil 50 ( $C_{50}$ ).
- El cálculo de un determinado cuantil  $k$  de una variable resulta relativamente sencillo de obtener a partir de la distribución de frecuencias de dicha variable si nos fijamos en la columna de los porcentajes acumulados: el  $C_k$  corresponderá al valor de la variable cuyo porcentaje acumulado sea igual a  $k$  o, en su caso, el porcentaje acumulado mayor a  $k$  que sea más bajo.

### **2.4.3.- Variables Cuantitativas: La Media y sus Alternativas Robustas**

La media es el valor promedio de un conjunto de datos numéricos, calculada como la suma del conjunto de valores dividida entre el número total de valores.

La estadística robusta es una aproximación alternativa a los métodos estadísticos clásicos. El objeto es producir estimadores que no sean afectados por variaciones pequeñas respecto a las hipótesis de los modelos.

Las estadísticas robustas intentan proporcionar métodos que emulan a los métodos clásicos, pero que no son afectados indebidamente por valores atípicos u otras pequeñas discrepancias respecto de las asunciones del modelo.

## 2.5.- Estadísticos de Dispersión

### Medidas de dispersión (ejemplo)

En 3 máquinas de empaque de cajas de cereal, que deben llenar cajas de **350 gramos**, se desea conocer si su rendimiento es óptimo. Tienen un ritmo de producción de 1000 cajas cada 8 horas diarias. Se ha tomado una muestra de X observaciones de cada máquina, cuyas distribuciones son las siguientes :



- La máquina A es la de menor/peor producción, sólo llena cajas a 275 grs, en lugar de los 350 grs que dice el empaque.
- La máquina B llena a 362.5 grs, pero su **dispersión** es tan amplia que muchas cajas contienen menos de lo indicado.
- La máquina C llena a 362.5 grs pero la **dispersión** es estrecha, por lo que pocas cajas contienen menos de lo indicado.



## 2.5.1.- La Razón de Variación

*Razón es el cociente entre dos números o dos cantidades comparables entre sí, expresado como fracción. Los términos de una razón se llaman: antecedente y consecuente. El antecedente es el dividendo y el consecuente es el divisor. No hay que confundir razón con fracción. Si es una fracción, entonces a y b son números enteros con  $b \neq 0$ .*

### Coeficiente de variación.

- ❖ Comparar las variabilidades de dos conjuntos de datos.
- ❖ Representa la diferencia entre la magnitud tanto en la media como en la desviación estándar.

$$CV = \frac{S}{\bar{X}} (100)$$

## Índice de Variación Cualitativa

El índice de variación cualitativa (IQV) es una medida de variabilidad para variables nominales, como carrera etnia o género. Este tipo de variables divide a las personas por categorías que no se pueden clasificar, a diferencia de una medida variable de ingresos o educación, que se puede medir de mayor a menor.



## El Rango Intercuartil

El Rango es el intervalo entre el valor máximo y el valor mínimo; por ello, comparte unidades con los datos. Permite obtener una idea de la dispersión de los datos, cuanto mayor es el rango, aún más dispersos están los datos (sin considerar la afectación de los valores extremos). El rango, también es llamado amplitud o recorrido de medida.

Se le llama Rango Intercuartil, a la diferencia entre el tercer y el primer cuartil de una distribución. Es una medida de la dispersión estadística.

A diferencia del rango, se trata de un estadístico robusto.

Es una medida de variabilidad adecuada cuando la medida de posición central empleada ha sido la mediana. Se define como la diferencia entre el tercer cuartil (Q3) y el primer cuartil (Q1), es decir:  $RQ = Q3 - Q1$ . A la mitad del rango intercuartil se le conoce como desviación cuartil (DQ), es afectada muy poco por cuentas extremas.

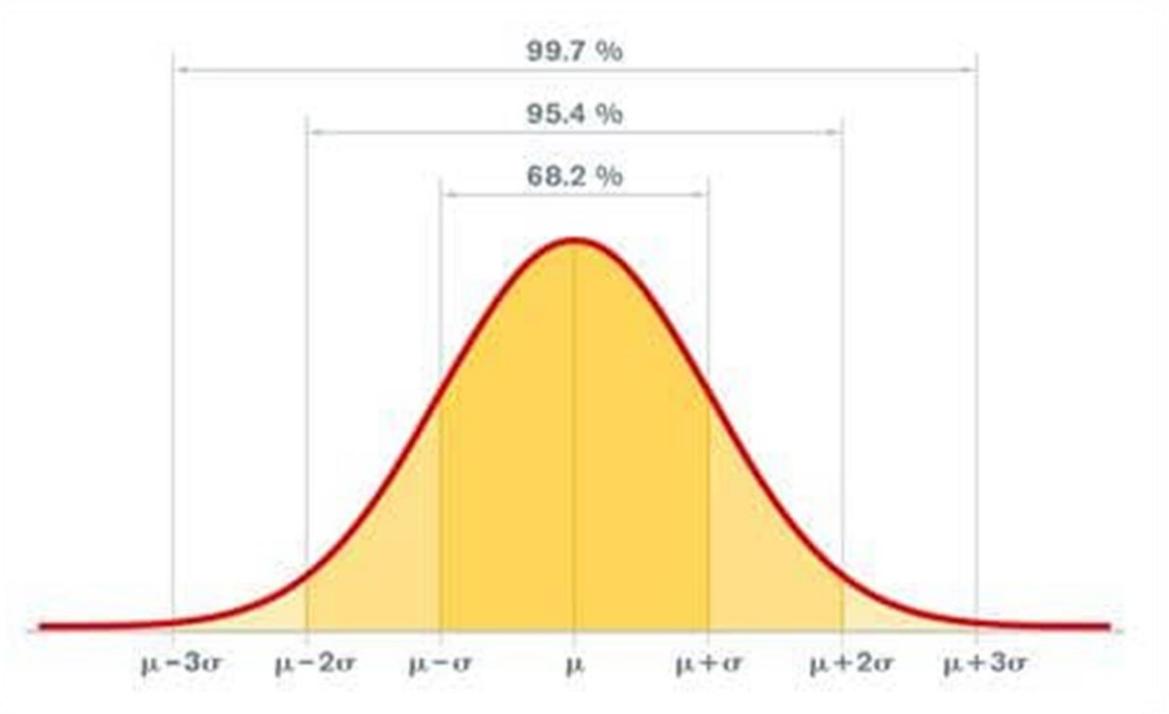
### 2.5.3.- La Varianza

*La varianza de una muestra o de un conjunto de valores, es la sumatoria de las desviaciones al cuadrado con respecto al promedio o a la media, todo esto dividido entre el número total de observaciones menos 1. De manera muy general se puede decir que la varianza es la desviación estándar elevada al cuadrado.*

*La varianza será siempre un valor positivo o cero, en el caso de que las puntuaciones sean iguales. 2 Si a todos los valores de la variable se les suma un número la varianza no varía. 3 Si todos los valores de la variable se multiplican por un número la varianza queda multiplicada por el cuadrado de dicho número.*



La desviación típica (también conocida como desviación estándar y desvío típico y representada de manera abreviada por la letra griega minúscula sigma  $\sigma$  o la letra latina s, así como por las siglas SD (de standard deviation, en algunos textos traducidos del inglés)) es una medida que se utiliza para cuantificar la variación o la dispersión de un conjunto de datos numéricos.



## Coeficiente de Variación

***En estadística, cuando se desea hacer referencia a la relación entre el tamaño de la media y la variabilidad de la variable, se utiliza el coeficiente de variación (suele representarse por las siglas "C.V.").***

***Es importante que todos los valores sean positivos y su media dé, por tanto, un valor positivo. A mayor valor del coeficiente de variación mayor heterogeneidad de los valores de la variable; y a menor C.V., mayor homogeneidad en los valores de la variable.***

