

UDS

ESTADISTICA



**ALEJANDRO
GUZMAN
ARROYO**



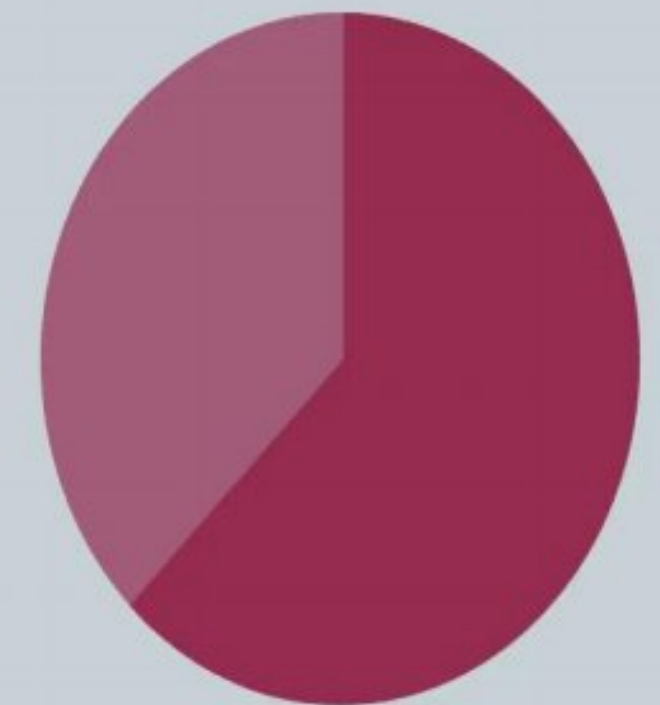
Unidad 1: Introducción a la estadística aplicada a la Psicología

La estadística consiste en métodos, procedimientos y fórmulas que permiten recolectar información para luego analizarla y extraer de ella conclusiones relevantes. Se puede decir que es la Ciencia de los Datos y que su principal objetivo es mejorar la comprensión de los hechos a partir de la información disponible.

La estadística ayuda a obtener conclusiones relevantes para el estudio de todo tipo de agentes como: humanos, animales, plantas, etc. Generalmente lo hace a través de muestras estadísticas.



La estadística es muy importante para la psicología



▶ LA ESTADÍSTICA

Objetivos de la estadística

Los principales objetivos de la estadística son los siguientes:

Conocer las características y hacer inferencias o llegar a conclusiones respecto a una población objetivo. Esto, usualmente a partir del análisis de una muestra. Esto es propio de la estadística inferencial.

Puede permitir establecer relación entre distintas variables, hallando el posible origen de un fenómeno, estudiando los cambios en dicho evento y haciendo proyecciones sobre el mismo, de ser posible.

En base a las conclusiones obtenidas, se pueden tomar decisiones, por ejemplo, si hablamos de un estudio estadístico realizado por el Gobierno para definir una política pública.

En el caso de la estadística descriptiva, permite tener un estado de la cuestión, es decir, conocer las características de una base de datos, por ejemplo, calculando las medidas de tendencia central como la media o la moda.

Sirve de apoyo a otras disciplinas como la economía, en el análisis y proyección de indicadores como la inflación o el Producto Interior Bruto. Asimismo, en el campo de la biología, tenemos la bioestadística que analiza, en otros, datos de salud pública y medioambientales.



Elementos de la estadística

Los principales elementos de la estadística son:



Población: Grupo de individuos que presenta o podría presentar un rasgo característico común que se desea investigar.

Muestra: Es un subgrupo de datos extraídos de una población que debe representar adecuadamente la totalidad del grupo.

Parámetros: Son medidas que ofrecen información sobre el centro de un conjunto de datos (medidas de tendencia central), otras sobre la dispersión o variabilidad (medidas de dispersión) y otras sobre la posición de un valor (medidas de posición como los percentiles).

Experimento: Proceso o actividad llevada a cabo de forma intencional para obtener una serie de datos o para ratificar o refutar una hipótesis.

Variable: La característica o cualidad de una muestra o población a la cual se le puede asignar un valor.

1.2.- Método Científico y Estadística

La ciencia es un estudio que se ocupa de la observación y clasificación de los hechos.

Los científicos deben, entonces ser capaces de observar un suceso o conjunto de eventos como resultado de un plan o diseño.

El método científico es un conjunto de principios y procedimientos para la búsqueda sistemática del conocimiento. El método científico está compuesto por los siguientes pasos:

1. Formula una teoría (problema).
2. Recoger datos para probar la teoría.
3. Analizar los datos.



- Para el Método Científico es esencial el estudio de lo que ya se conoce.

- A partir de este conocimiento, se formulan Hipótesis, las cuales se ponen a prueba partiendo por lo general de situaciones experimentales, siendo esta la parte crucial de todo el proceso ya que cualquier pregunta donde sus respuestas no pueden obtenerse a partir de la investigación planeada, no puede aceptarse como parte del Método Científico.



1.3.- ¿Porque la Estadística en el Grado de la Psicología?

La estadística es importante en psicología porque permite extraer y resumir información útil de las observaciones que se hacen, el psicólogo debe basar sus decisiones en datos limitados y estas son más fáciles de tomar con la ayuda de la estadística, le da mayor claridad y precisión al pensamiento y la investigación psicológica.

Gracias a la estadística, los psicólogos pueden:

Organizar datos: Puede resultar muy abrumador trabajar con una gran cantidad de información. Es por eso que, a la hora de realizar la presentación de los datos, la estadística sea una gran herramienta. Entre las distintas maneras de presentar los datos de manera comprensible se encuentran los gráficos. Existen muchos y con distintas características. Pueden ser de barras, de torta, de líneas o de dispersión, dependiendo que necesitemos presentar. También se pueden utilizar tablas de frecuencia.



- Las áreas de Aplicación de la Psicología para fines de estudio se clasifican en cuatro grupos:
- Social: Conocer las necesidades, gustos o tendencias de una población.
- Clínica: Permite conocer las características individuales de los pacientes y sus padecimientos para un mejor diagnóstico.
- Educativa: Determina las características y las necesidades de una institución escolar para desarrollar programas que les permita mejorar.
- Organizacional: Conocer las características de los empleados para una mejor toma de decisiones en cuanto a su desempeño o habilidades.

1.4.- Conceptos Básicos de Estadística

La Estadística es la parte de las Matemáticas que se encarga del estudio de una determinada característica en una población, recogiendo los datos, organizándolos en tablas, representándolos gráficamente y analizándolos para sacar conclusiones de dicha población.

Estadística descriptiva. Realiza el estudio sobre la población completa, observando una característica de la misma y calculando unos parámetros que den información global de toda la población.

Estadística inferencial. Realiza el estudio descriptivo sobre un subconjunto de la población llamado muestra y, posteriormente, extiende los resultados obtenidos a toda la población.



La estadística es un campo del conocimiento que permite al investigador deducir y evaluar conclusiones acerca de una población a partir de información proporcionada por una muestra.

Específicamente, la estadística trata de teoremas, herramientas, métodos y

técnicas que se pueden usar en:

- a. Recolección, selección y clasificación de datos.
- b. Interpretación y análisis de datos.
- c. Deducción y evolución de conclusiones y de su confiabilidad, basada en datos muestrales.



Los métodos de la estadística fueron desarrollados para el análisis de datos muestrados, así como para propósitos de inferencia sobre la población de la que se selecciono la muestra.

La estadística como ciencia, cubre un extenso campo donde poder aplicarla.

Se agrupa en 2 grandes áreas: estadística descriptiva y estadística inferencial, que desempeñan funciones distintivas, pero complementarias en el análisis.



1.5.- Metodología de Investigación

El término “estadística descriptiva” se refiere al análisis, el resumen y la presentación de los resultados relacionados con un conjunto de datos derivados de una muestra o de toda la población.

La estadística descriptiva comprende tres categorías principales: distribución de frecuencias, medidas de tendencia central y medidas de variabilidad.

Por estadística descriptiva entendemos, por ejemplo, el cálculo de la media y la mediana, dos indicadores muy importantes y sobre todo diferentes. La mediana es un indicador que “no tiene en cuenta los valores extremos, a veces poco frecuentes”, a diferencia de la media, que está muy influida por estos valores extremos.



Importancia de la estadística descriptiva

La estadística descriptiva facilita la visualización de los datos. Permiten presentarlos de forma significativa y comprensible, lo que a su vez da pie a una interpretación simplificada del conjunto de datos en cuestión.

Los datos brutos serían difíciles de analizar, y la determinación de tendencias y patrones puede ser un reto. Además, los datos en bruto dificultan la visualización de lo que muestran los datos.

Además, el uso de la estadística descriptiva permite resumir y presentar un conjunto de datos mediante una combinación de descripciones tabuladas y gráficas. La estadística descriptiva se utiliza para resumir datos cuantitativos complejos.



La estadística inferencial

Es una parte de la estadística que comprende los métodos y procedimientos que por medio de la inducción determina propiedades de una población estadística, a partir de una parte de esta. Su objetivo es obtener conclusiones útiles para hacer deducciones sobre una totalidad, basándose en la información numérica de la muestra.

Estudio de la estadística inferencial

Toma de muestras o muestreo cuantitativo, que se refiere a la forma adecuada de considerar una muestra que permita obtener conclusiones estadísticamente válidas y significativas.

Estimación de parámetros o variables estadísticas, que permite estimar valores poblacionales a partir de muestras de mucho menor tamaño.

Contraste de hipótesis, que permite decidir si dos muestras son estadísticamente diferentes, si un determinado procedimiento tiene un efecto estadístico significativo, etc.

Diseño experimental.

Inferencia bayesiana.

Métodos no paramétricos.



La población en investigación es un conjunto completo de elementos que poseen un parámetro común entre sí.

Es importante mencionar que todos somos conscientes de lo que la palabra “población” significa en nuestra vida cotidiana. A menudo se utiliza para describir la población humana o el número total de personas que viven en un área geográfica de algún país o estado.



Una muestra es la parte más pequeña del total, es decir, un subconjunto de toda la población. Cuando se realizan encuestas, la muestra son los miembros de la población que son invitados a participar en la encuesta.

Dicho de manera sencilla, una muestra es un subgrupo o subconjunto dentro de la población, que puede ser estudiado para investigar las características o el comportamiento de los datos de población.

Las muestras de datos se crean utilizando varios métodos de investigación como el muestreo probabilístico y el muestreo no probabilístico. Los métodos de muestreo varían según los tipos de investigación y la calidad de la información requerida.

En estadística, un parámetro es un número que resume la gran cantidad de datos que pueden derivarse del estudio de una variable estadística. El cálculo de este número está bien definido, usualmente mediante una fórmula aritmética obtenida a partir de datos de la población.

Los parámetros estadísticos son una consecuencia inevitable del propósito esencial de la estadística: crear un modelo de la realidad.



Estadísticos:

estudio de los datos cuantitativos de la población, de los recursos naturales e industriales, del tráfico o de cualquier otra manifestación de las sociedades humanas.

Conjunto de datos estadísticos.

Rama de la matemática que utiliza grandes conjuntos de datos numéricos para obtener inferencias basadas en el cálculo de probabilidades.

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.



La organización de los datos constituye la primera etapa de su tratamiento, puesto que facilita los cálculos posteriores. Para ir analizando los datos podemos organizarlos en una tabla de frecuencias, o sea una tabla que divide los valores en grupos similares llamados clases llamada distribución de frecuencias.

Es una agrupación de datos en clases mutuamente excluyentes que indican el número de observaciones en cada clase (frecuencias).



2.2.-Representación Gráfica de Una Distribucion de Frecuencias

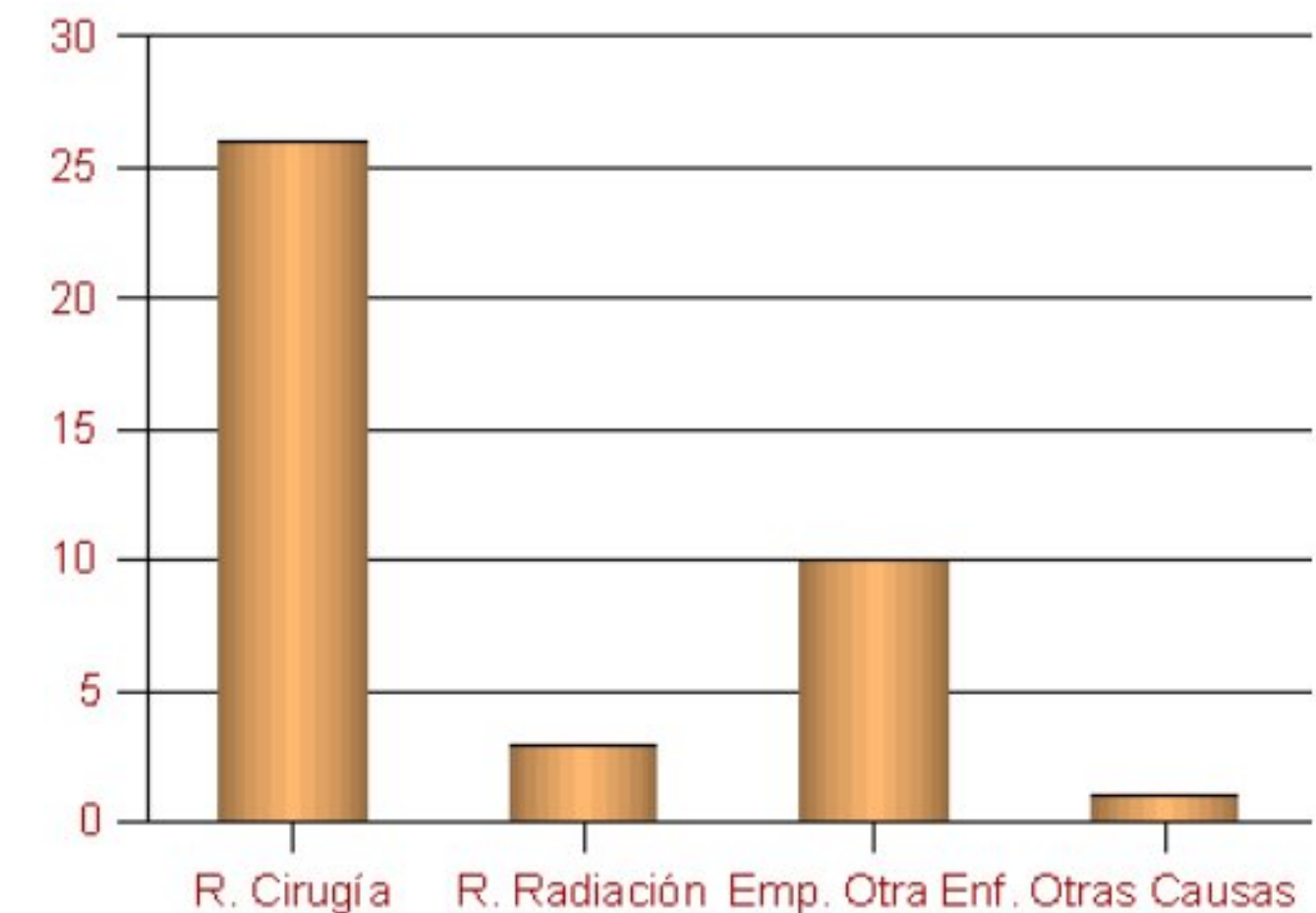
La representación gráfica de una distribución de frecuencias depende del tipo de datos que la constituya.

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:

Esta 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.

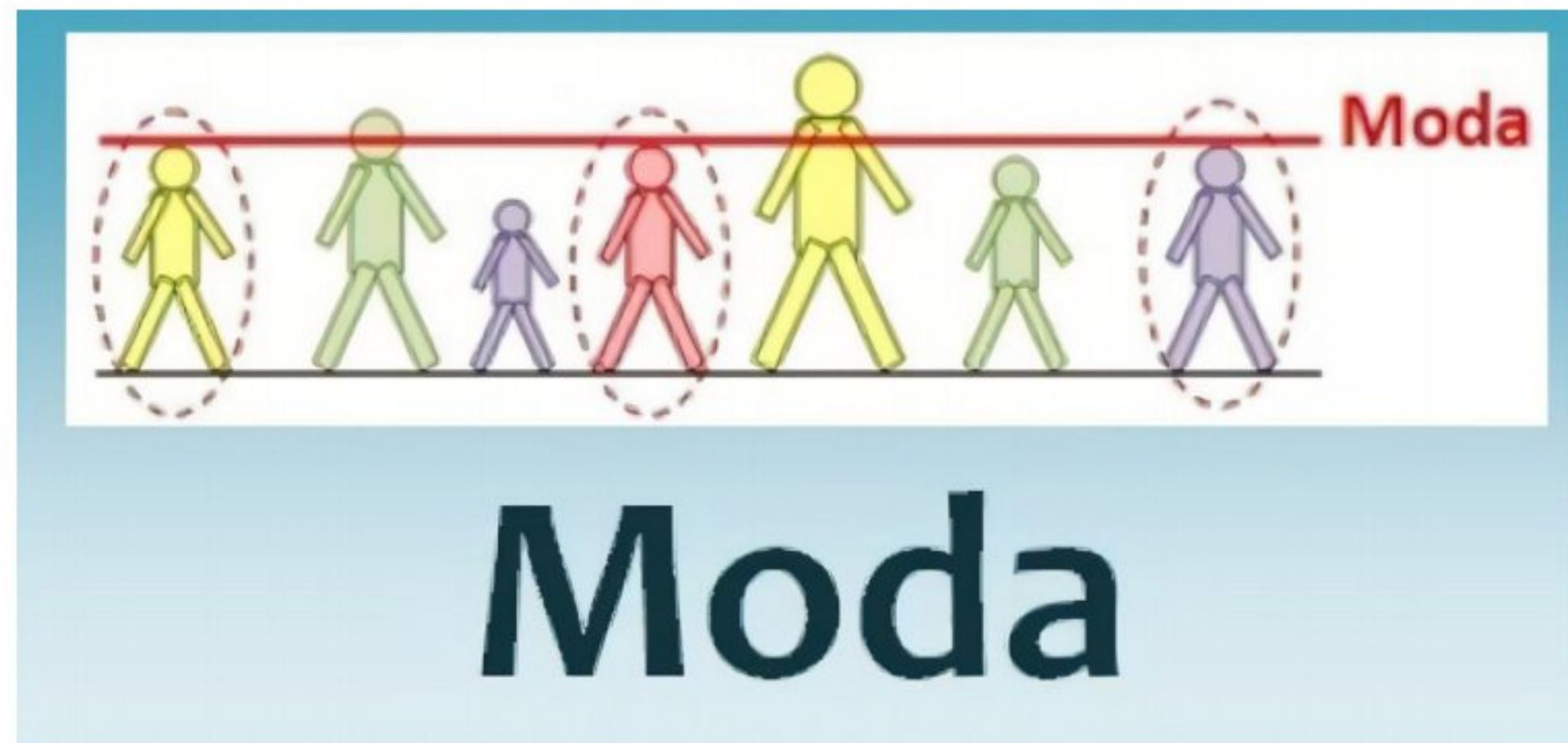
Esto proporciona un valor añadido a la agrupación de datos.



2.4.- Estadísticos de Posición Grupal

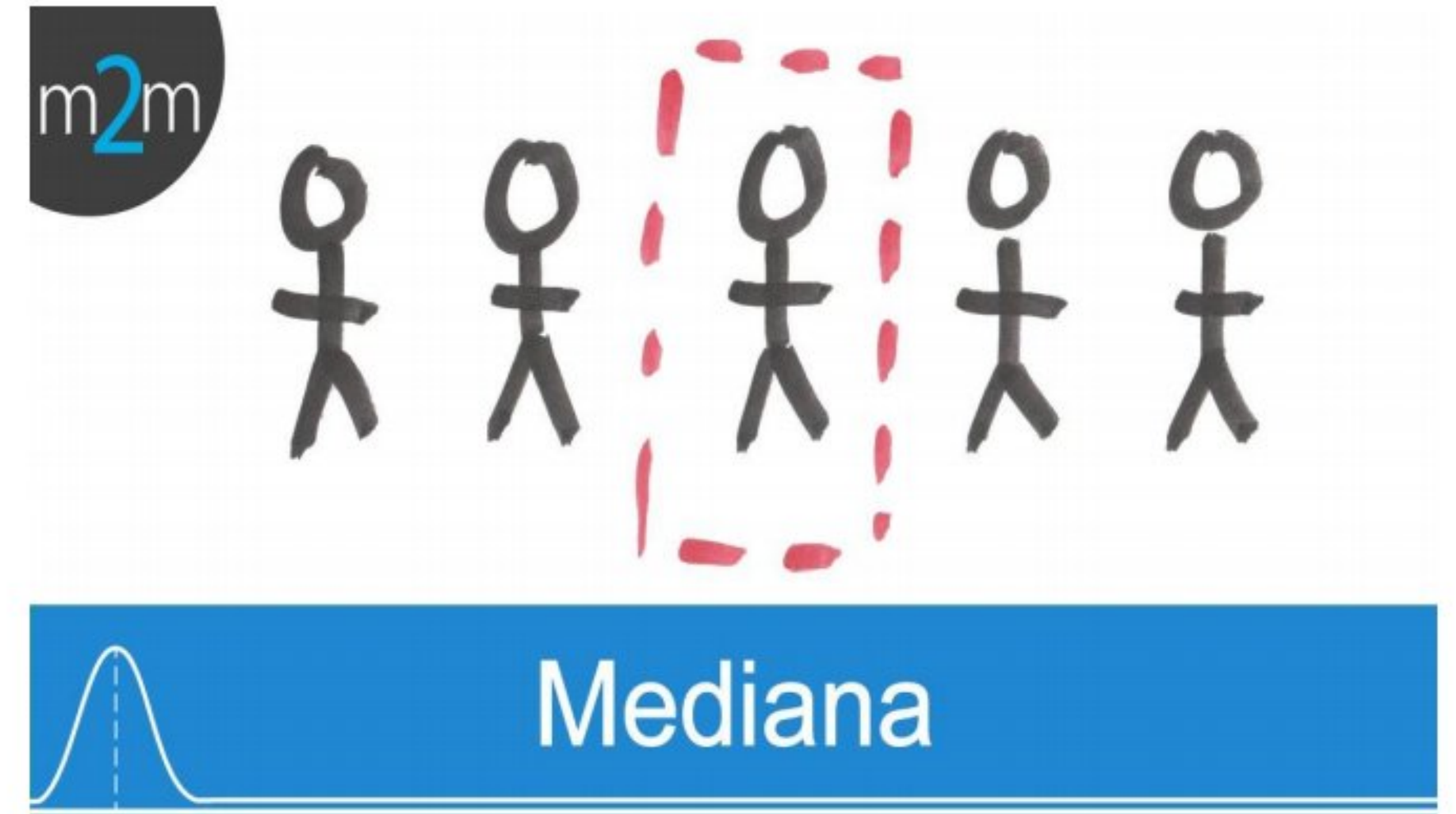
Se les denomina aquí como estadísticos de posición grupal porque sintetizan la información de todos los datos de una variable, esto es, para todo el grupo de casos a partir de los que esos datos han sido recogidos.

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.



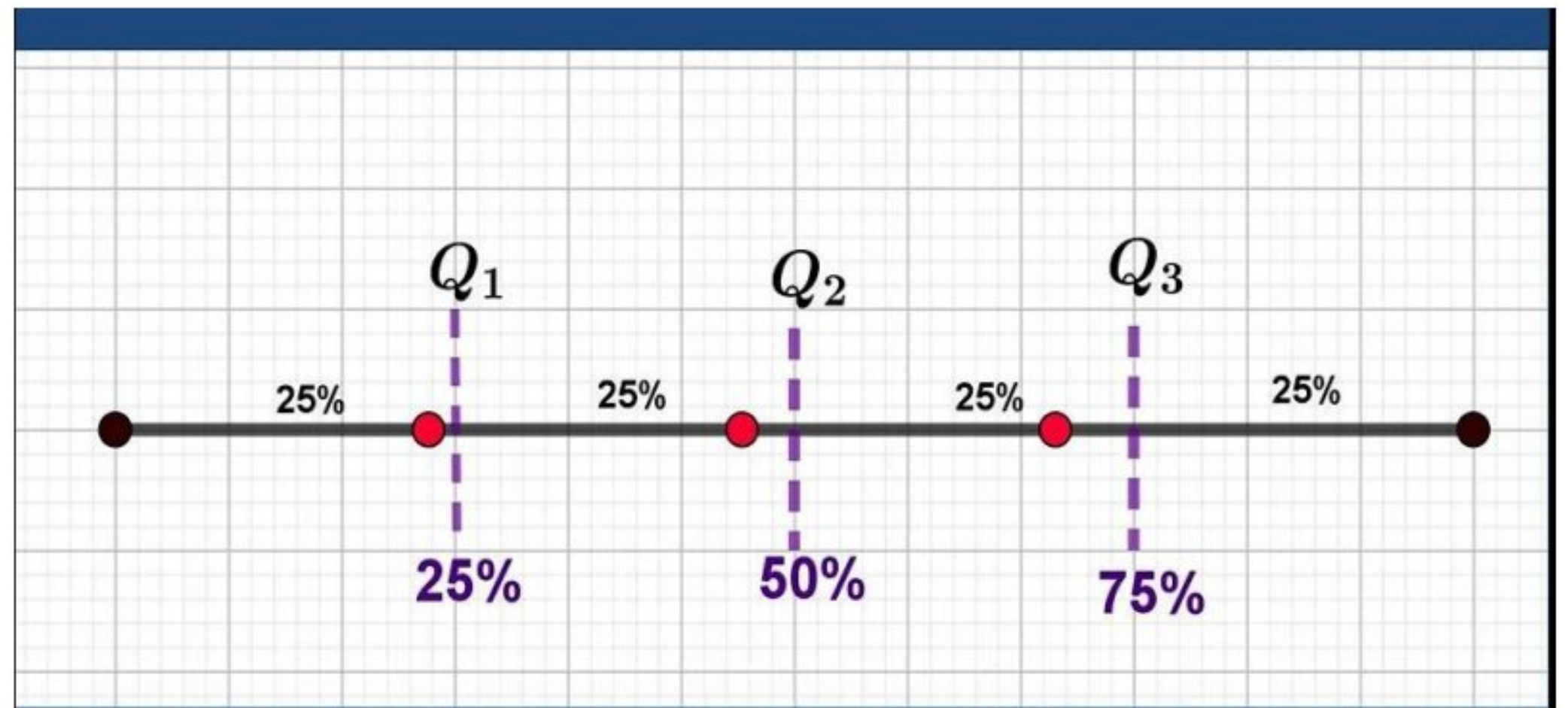
El Mínimo y El Máximo

- 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.

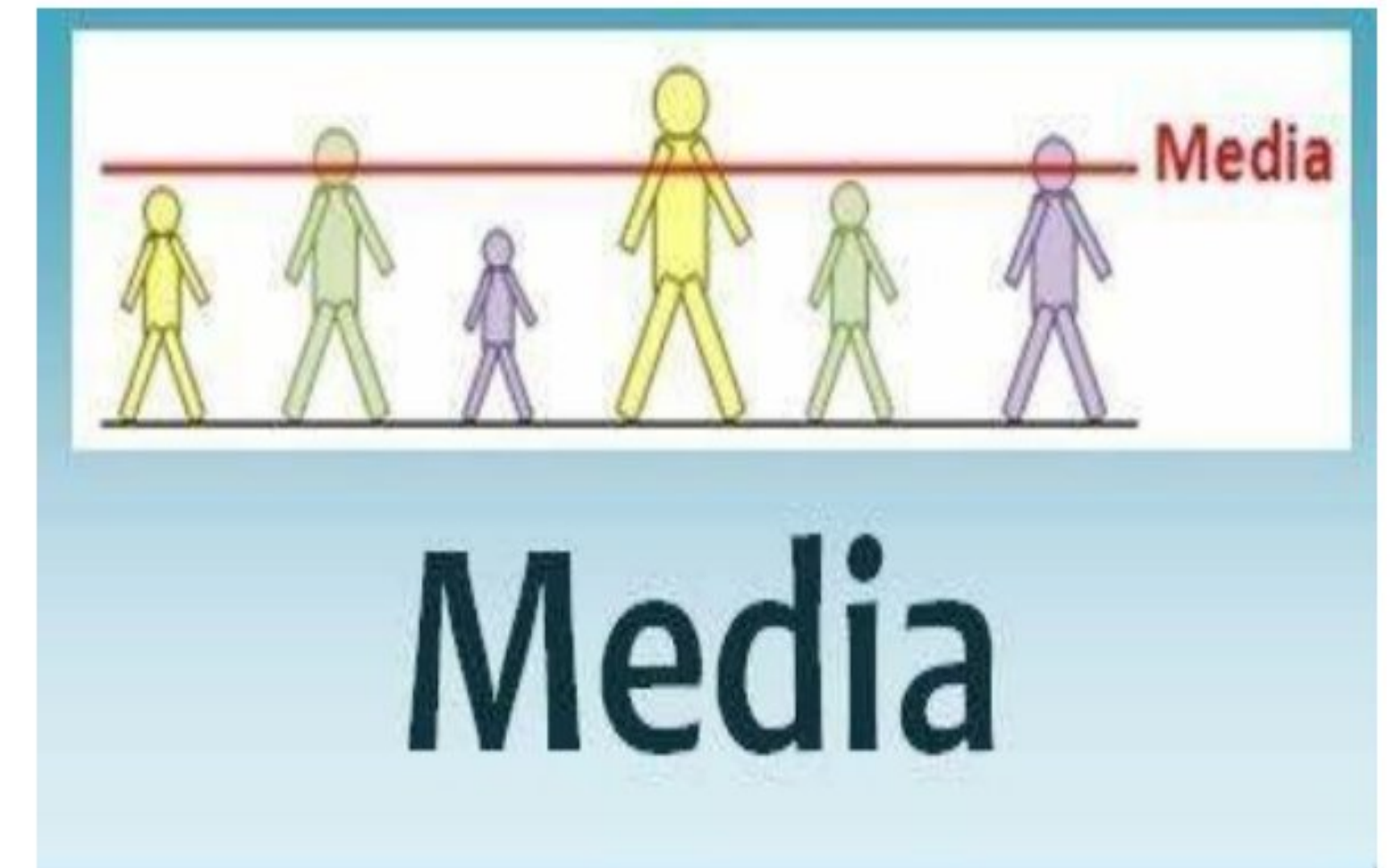


2.4.3.- Variables Cuantitativas: La Media y sus Alternativas Robustas

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

La dispersión estadística es el grado en que una distribución de datos se aleja, o se acerca, en valor absoluto a la media aritmética, como estadístico de posición central. Por tanto, las medidas de dispersión van a acompañar siempre a esa media o promedio.

Es la medida de dispersión que mejor proporciona la variación de los datos con respecto a la media aritmética. Su valor se relaciona directamente con la dispersión de los datos, a mayor dispersión de ellos, mayor desviación típica y a menor dispersión menor desviación típica.

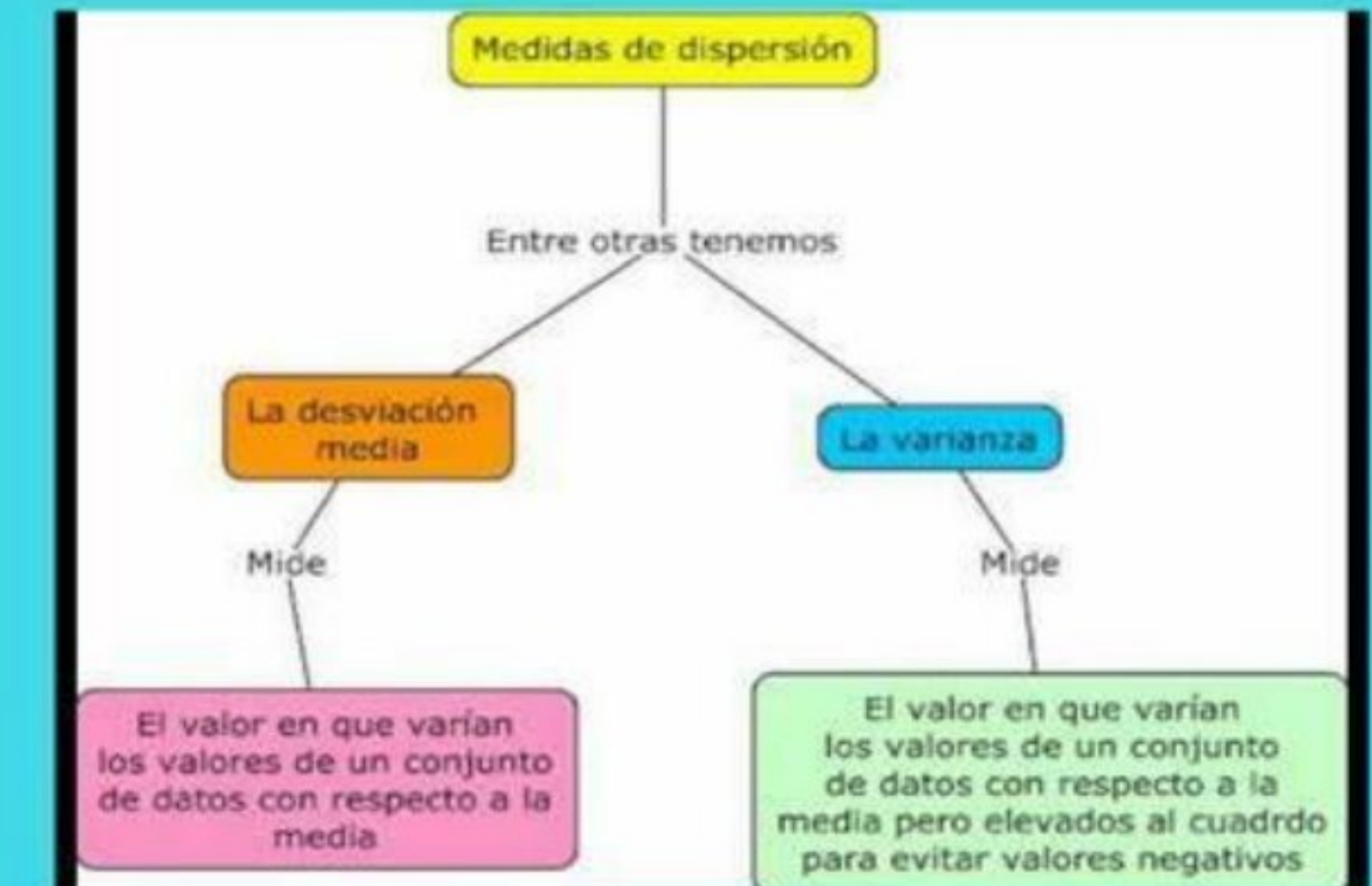
Las medidas de dispersión consisten en números que nos otorgan información acerca de la variabilidad de los datos. Es decir, se encargan de mostrar qué tan juntos o separados se encuentran los datos de una distribución.

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 $b \neq 0$.

se denomina variación a cada una de las tuplas que pueden formarse tomando elementos de un conjunto. En combinatoria de conjuntos finitos frecuentemente se necesita conocer número de variaciones de un conjunto de m elementos tomados en tuplas de n elementos

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.

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

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.

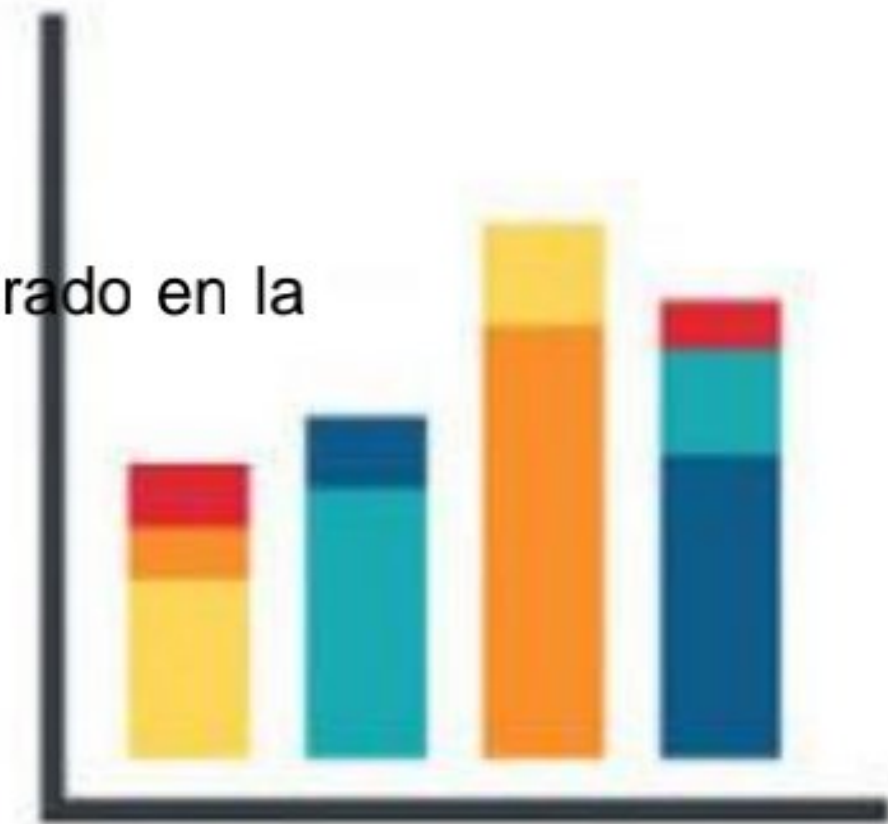
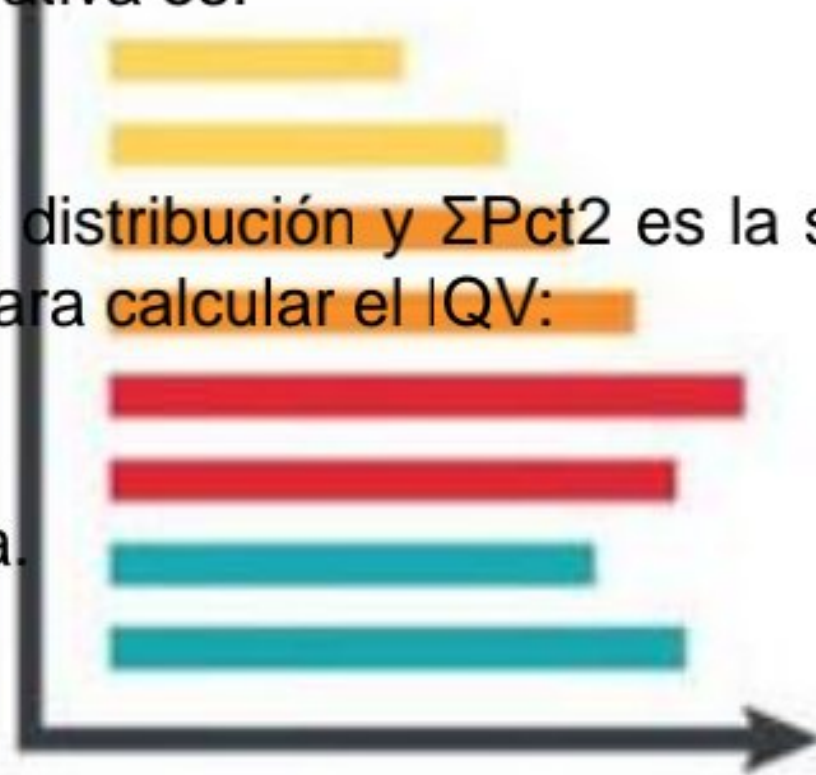
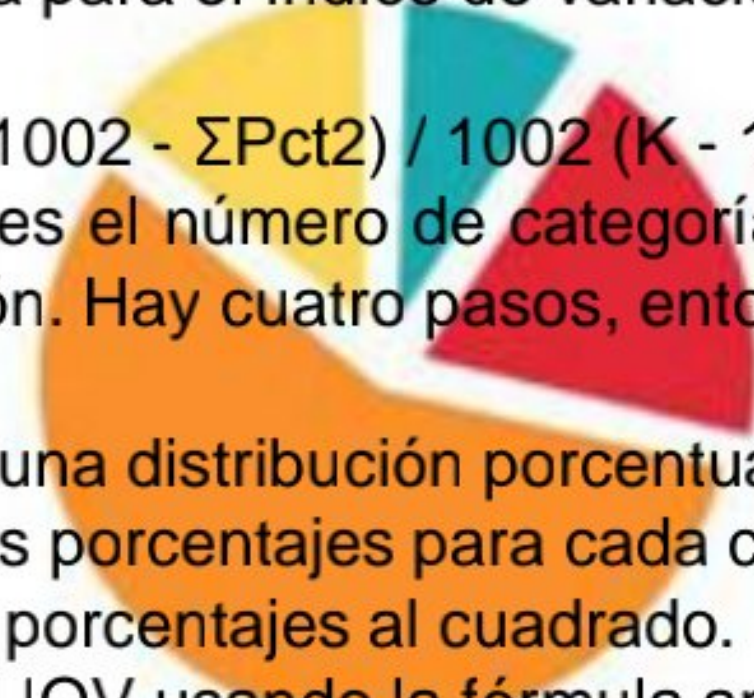


La fórmula para el índice de variación cualitativa es:

$$IQV = K (1002 - \sum Pct^2) / 1002 (K - 1)$$

Donde K es el número de categorías en la distribución y $\sum Pct^2$ es la suma de todos los porcentajes al cuadrado en la distribución. Hay cuatro pasos, entonces, para calcular el IQV:

- Construir una distribución porcentual.
- Cuadra los porcentajes para cada categoría.
- Suma los porcentajes al cuadrado.
- Calcule el IQV usando la fórmula anterior.



2.5.2.- El Rango y El Rango Intercuartil

$$Rango = (Max) - (Min)$$

Rango

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.

$$Q = \frac{Q_3 - Q_1}{2}$$

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: $RQ = Q_3 - Q_1$. A la mitad del rango intercuartil se le conoce como desviación cuartil (DQ), es afectada muy poco por cuentas extremas.

$$IQR = Q_3 - Q_1$$

Rango intercuartílico

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.

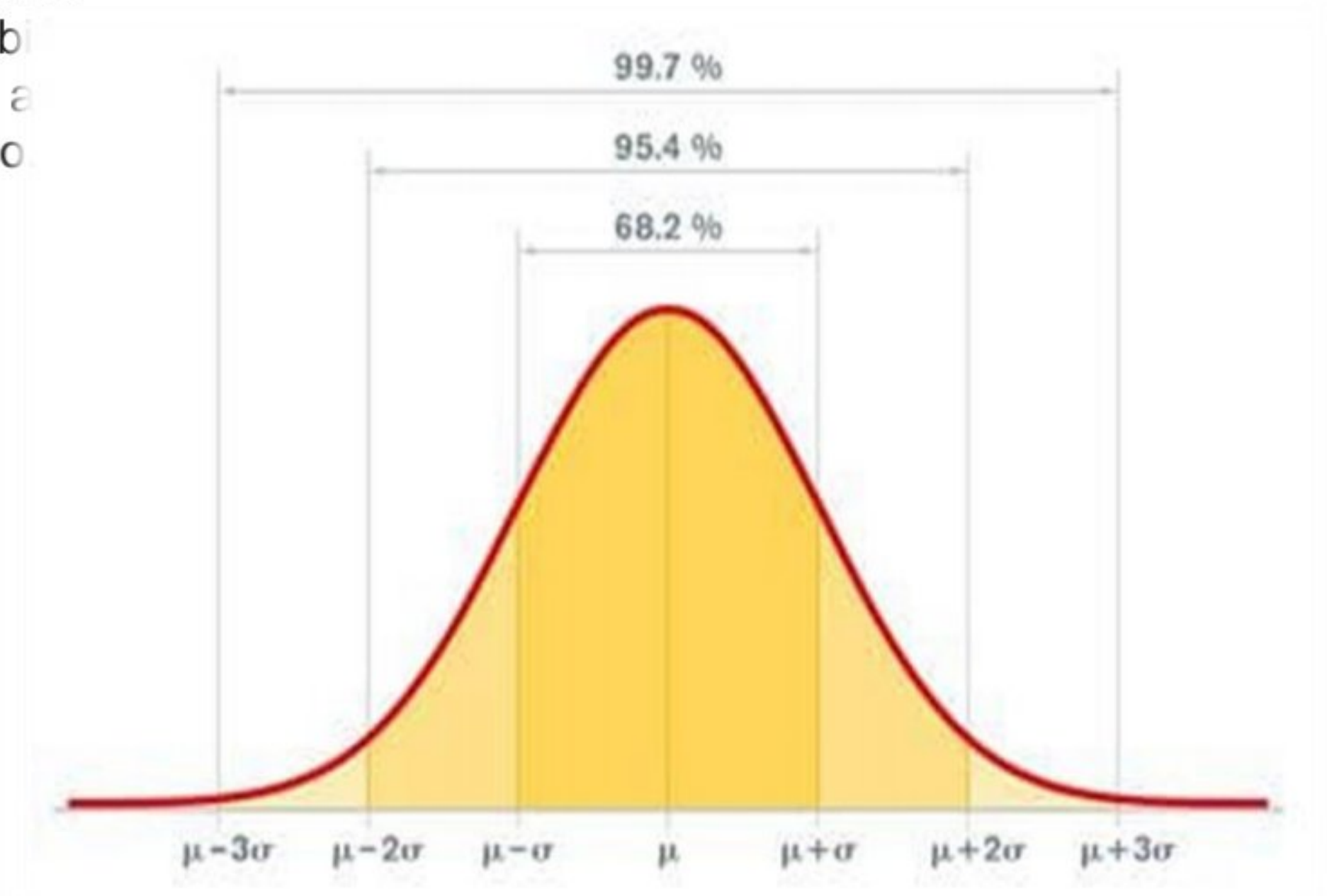
La varianza será siempre un valor positivo o cero, en el caso de que las puntuaciones sean iguales. 2 Si a todos los valores de la variable se les suma un número la varianza no varía. 3 Si todos los valores de la variable se multiplican por un número la varianza queda multiplicada por el cuadrado de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



Desviación Típica

La desviación típica (también conocida como desviación estándar y desvío típico y representada de manera abreviada por la letra griega minúscula sigma σ 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.¹

Una desviación estándar baja indica que la mayor parte de los datos de una muestra tienden a estar agrupados cerca de su media (también denominada el valor esperado), mientras que una desviación estándar alta indica que los datos se extienden sobre un rango de valores más amplio.

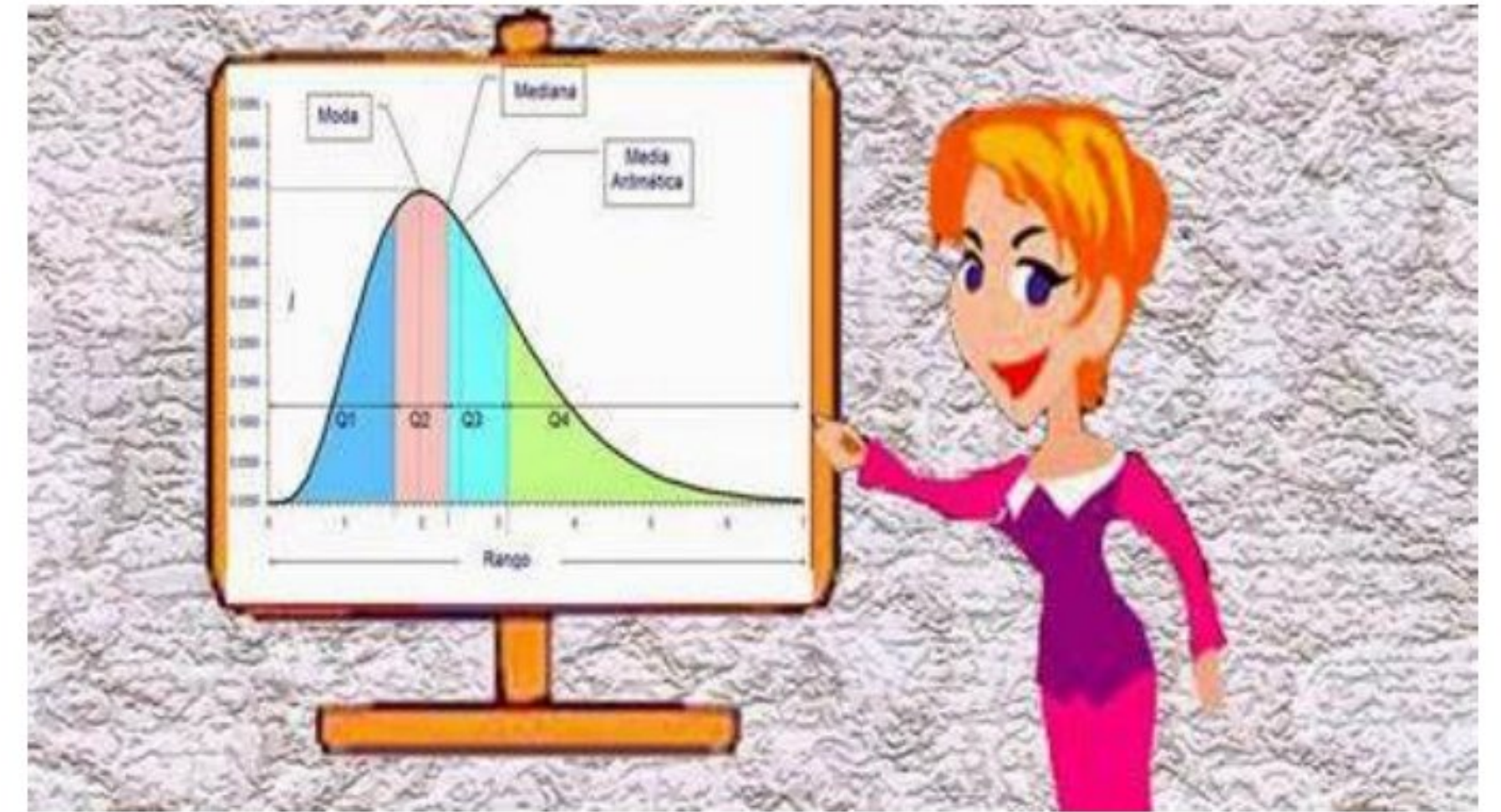


Coeficiente de Variación

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.").

Su fórmula expresa la desviación estándar como porcentaje de la media aritmética, mostrando una interpretación relativa del grado de variabilidad.

Por otro lado presenta problemas ya que a diferencia de la desviación típica este coeficiente es fuertemente sensible ante cambios de origen en la variable. Por ello 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.

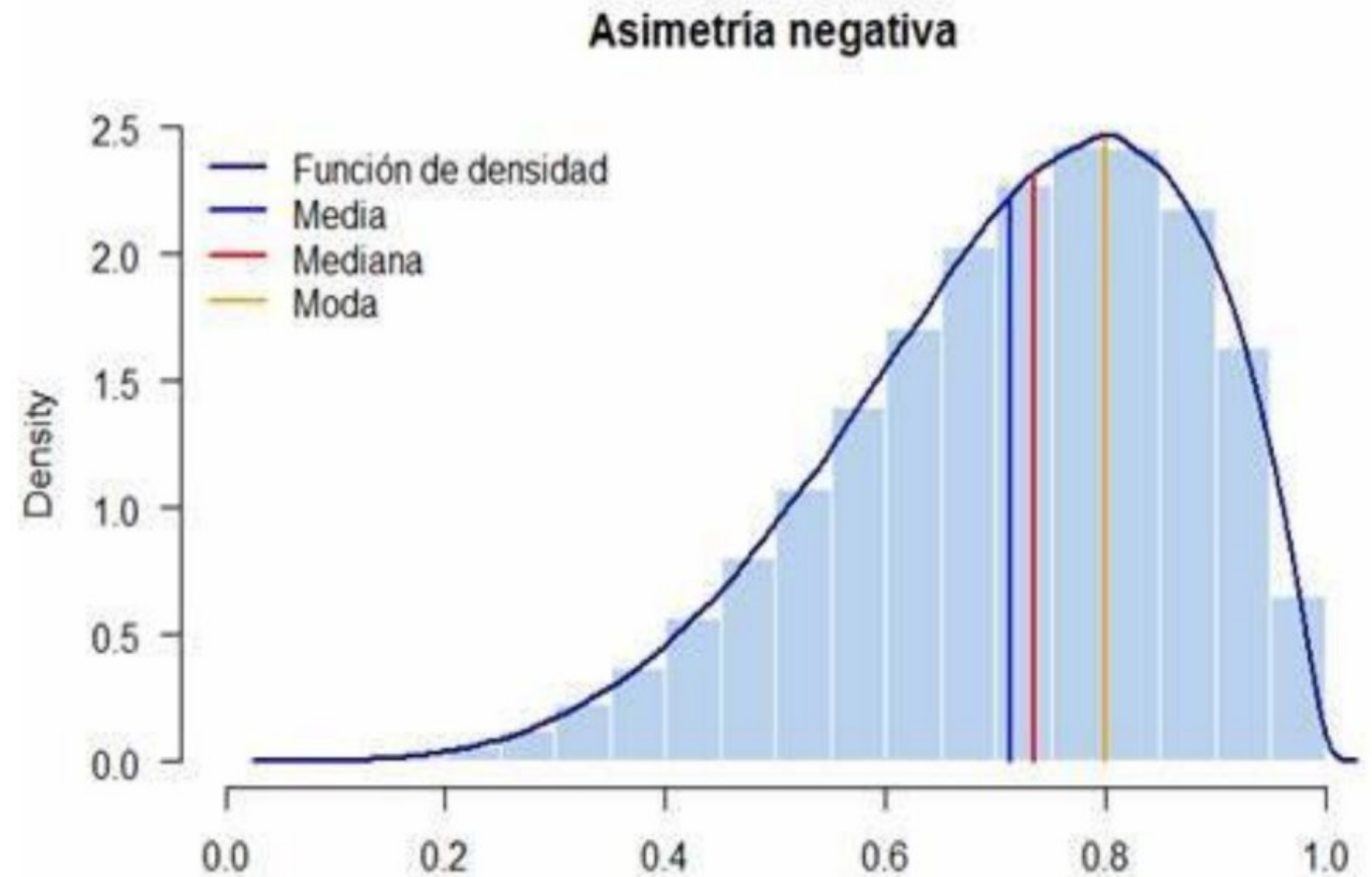


Unidad 3: Estadísticos de Forma de la Distribucion

3.1.-Asimetría

La asimetría es una propiedad de determinados cuerpos, dibujos, funciones matemáticas y otros tipos de elementos en los que, al aplicarles una regla de transformación efectiva, se observan cambios respecto al elemento original.

Los indicadores gracias a los cuales es posible identificar el grado de simetría o de asimetría presente en una distribución de probabilidad (una función que asigna a una serie de eventos determinados la probabilidad de que ocurran) de una variable aleatoria (también conocida con el nombre de variable estocástica, es una función que asigna sucesos a números reales) sin necesidad de representarla gráficamente se denominan medidas de asimetría.



Apuntamiento

La curtosis es una medida estadística que determina el grado de concentración que presentan los valores de una variable alrededor de la zona central de la distribución de frecuencias. También es conocida como medida de apuntamiento.

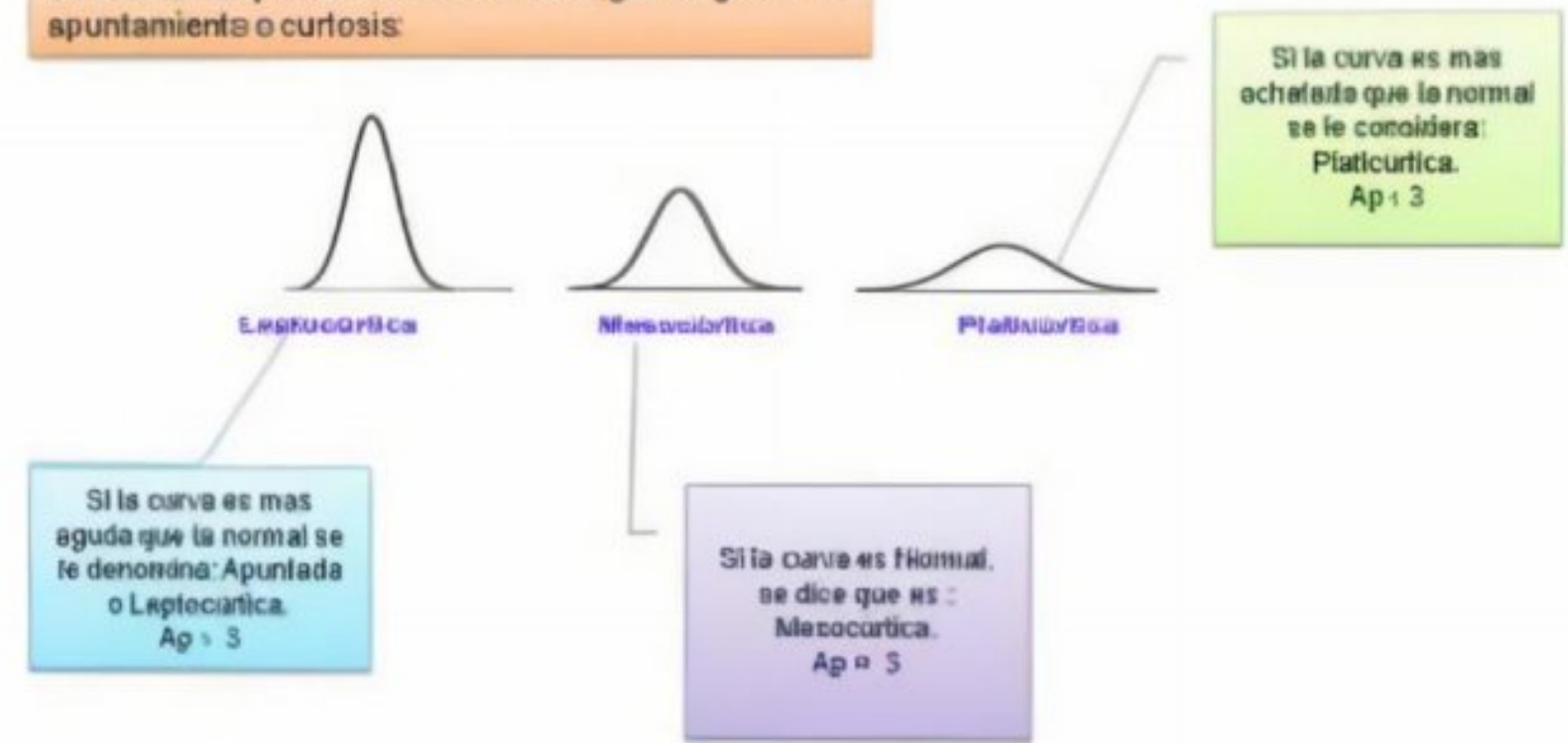
Según su concepción clásica, una curtosis grande implica una mayor concentración de valores de la variable tanto muy cerca de la media de la distribución (pico) como muy lejos de ella (colas), al tiempo que existe una relativamente menor frecuencia de valores intermedios. Esto explica una forma de la distribución de frecuencias/probabilidad con colas más gruesas, con un centro más apuntado y una menor proporción de valores intermedios entre el pico y colas.

Una mayor curtosis no implica una mayor varianza, ni viceversa.

APUNTAMIENTO O CURTOSIS (A_p)

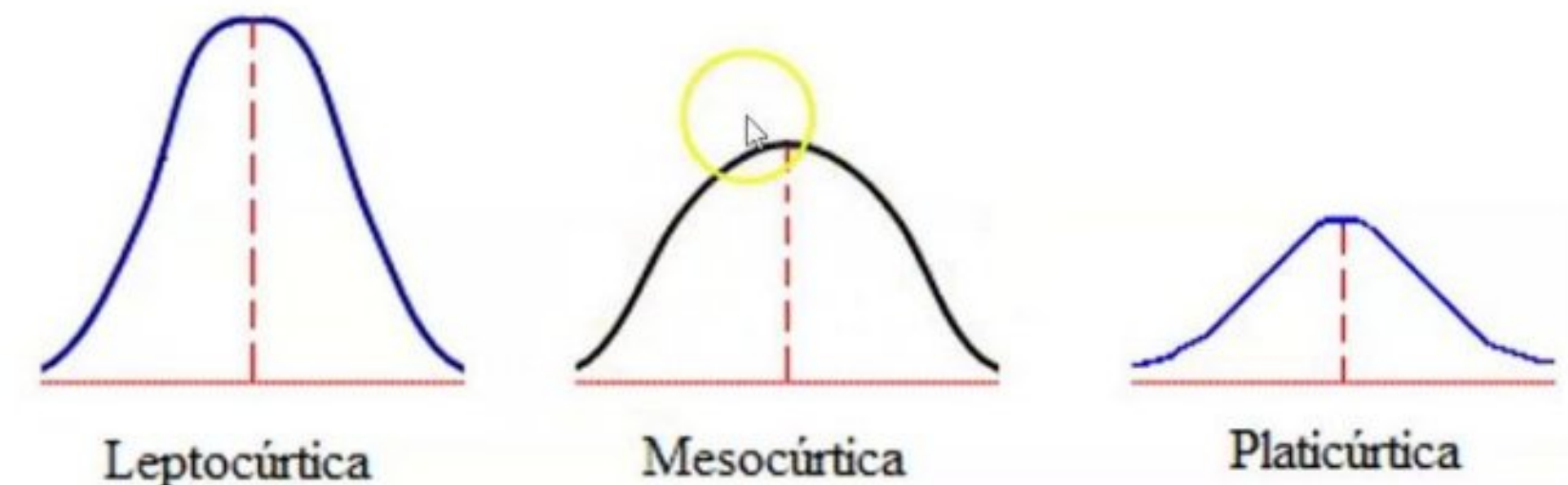
Mide la mayor o menor cantidad de datos que se agrupan en torno a la moda.

Se definen 3 tipos de distribuciones según su grado de apuntamiento o curtosis:



Medidas de apuntamiento (Curtosis)

Representa la medida de concentración de datos alrededor de la media.



3.3.- Estadísticos de Posición Individual

3.1.1.- Porcentajes Acumulados



El porcentaje acumulado (%a) de un valor concreto de una variable es el porcentaje de casos que obtienen un valor inferior o igual a ese en la variable en cuestión, información que puede obtenerse directamente a partir de la distribución de frecuencias correspondiente a esa variable.

- A estos porcentajes acumulados se les suele denominar más habitualmente, aunque de un modo equívoco, percentiles (término ya utilizado en el contexto de los estadísticos de posición grupal). En referencia al porcentaje acumulado de una determinada puntuación es la de rango centil, si bien, su uso no está muy extendido.

3.1.2.- Puntuaciones Típicas

Una alternativa al problema planteado en la utilización de las puntuaciones diferenciales son las puntuaciones típicas (estándar o z), una transformación de las puntuaciones directas que tiene en cuenta tanto la tendencia central (media) como la dispersión (desviación típica) de la distribución de frecuencias de la variable.

La puntuación z correspondiente a un determinado valor expresa el número de desviaciones típicas que ese valor dista de la media del conjunto de las observaciones.

11.1. Puntuaciones típicas o normalizadas

Para poder comparar dos datos correspondientes a dos distribuciones distintas, hay que tipificar —o normalizar— dichos valores, es decir, calcular los valores $z = \frac{x - \bar{x}}{\sigma}$, después, comparar los resultados.

Las puntuaciones típicas o normalizadas, también llamadas puntuaciones z, tienen las siguientes propiedades:

- Si se transforma una distribución en puntuaciones típicas, no varía la forma de la distribución original.
- La media aritmética de las puntuaciones normalizadas es cero, es decir, $\bar{z} = 0$
- La desviación típica de las puntuaciones típicas es la unidad, es decir, $\sigma_z = 1$

3.1.3.- Las Escalas Derivadas



Una dificultad con las puntuaciones típicas se puede plantear a la hora de comunicar resultados debido a las posiciones decimales y valores negativos inherentes a las mismas. Es por ello que se han propuesto algunas transformaciones lineales de las puntuaciones típicas que pretenden hacerlas más intuitivamente interpretables.

- Todas estas escalas derivadas de la escala de las puntuaciones típicas se basan en una transformación genérica del tipo:

$$D_i = a \cdot z_i + b,$$

3.4.- Organización Y Representación Grafica de Datos Multivariados

3.4.1.- La Distribucion Conjunta Multivariada

En el caso de tener más variables, se denominaría multivariada. El nombre largo de distribución conjunta es distribución de probabilidad conjunta. El nombre se abrevia dado que ya se conoce que esas distribuciones son de probabilidad.

Una distribución conjunta es la distribución de probabilidad de la intersección de las realizaciones de dos o más variables aleatorias cualesquiera. En otras palabras, una distribución conjunta es la distribución de probabilidad que forman dos o más variables aleatorias cuando sus realizaciones se producen simultáneamente.



3.4.2.- La Tabla de Contingencia

Una tabla de contingencia es una herramienta utilizada en la rama de la estadística, la cual consiste en crear al menos dos filas y dos columnas para representar datos categóricos en términos de conteos de frecuencia.

Esta herramienta, que también se conoce como tabla cruzada o como tabla de dos vías, tiene el objetivo de representar en un resumen, la relación entre diferentes variables categóricas.

La tabla de contingencia es una de las herramientas analíticas más útiles y un pilar de la industria de la investigación de mercados.

La tabla permite medir la interacción entre dos variables para conocer una serie de información “oculta” de gran utilidad para comprender con mayor claridad los resultados de una investigación.



TABLAS DE CONTINGENCIA

Estadística bidimensional y Probabilidad

		Itinerario (y)		
		Ciencias	Letras	
Sexo (x)	Chicas	13	6	19
	Chicos	9	12	21
		22	18	40



Ventajas de realizar una tabla de contingencia:

Entre los principales beneficios de realizar una tabla de contingencia se encuentran los siguientes:

Facilita la lectura de los datos recolectados, ya que permite agruparlos cuando aún se encuentran sin procesar, lo que disminuye el margen de error al realizar un informe de investigación.

Gracias a la tabla de contingencia es posible realizar gráficas que permitan visualizar la información fácilmente para su comprensión.

A diferencia de otros métodos estadísticos de análisis de datos, la tabla de contingencia permite ahorrar tiempo durante la correlación de variables.

Las tablas ofrecen resultados claros y precisos que permiten tomar mejores decisiones y crear estrategias basadas en datos.

Organización de datos



Tablas de contingencia

Tabla 1. Estado civil y sexo de los encuestados

Estado Civil \ Sexo	Sexo		Total
	Masculino	Femenino	
Soltero	2	1	3
Casado	2	2	4
Divorciado	2	1	3
Total	6	4	10

Fuente: Datos imaginarios



3.4.3.- Representación Graficas

3.4.3.1.- Dos Variables Categóricas

Las variables categóricas también se denominan variables cualitativas o variables de atributos. Los valores de una variable categórica son categorías o grupos mutuamente excluyentes. Los datos categóricos pueden tener o no tener un orden lógico.

Se puede hacer estadística con las variables categóricas, a pesar de no ser variables numérica o cuantitativas. Por ejemplo, para conocer la tendencia o valor más probable de una variable categórica se toma la moda. La moda es, en este caso, el resultado o valor más repetido de una variable categórica

Gráficos para variables cualitativas o categóricas



Diagrama circular



Gráfico de barras

Estas son algunas de las representaciones más utilizadas para variables categóricas.

LAS VARIABLES CATEGÓRICAS Ejemplo:

Al conjunto constituido por un grupo de individuos que comparten una característica ("estudiantes de enfermería") le asignamos el símbolo "1", o "I", o "A1"; también al conjunto constituido por individuos varones le asignamos "1" y al de los individuos pertenecientes al sexo femenino, "0".

En todo caso, la idea es evitar asignar el mismo número a dos conjuntos.

3.4.3.2.- Dos Variables Cuantitativas

Lo cualitativo se centra en la calidad o cualidad de las cosas, mientras que lo cuantitativo se refiere a la cantidad o a aquello que es susceptible de ser representado numéricamente.

Las variables cuantitativas son aquellas variables estadísticas que otorgan, como resultado, un valor numérico.





Tipos de variables

VARIABLES CUALITATIVAS ó CATEGÓRICAS

Expresa una cualidad, característica o atributo que solo se pueden clasificar o categorizar mediante el conteo.

Dependiendo del número de categorías pueden ser dicotómicas o politómicas

Ejemplos:

Sexo, Estado Civil, Dolor, Grupo Sanguíneo, Métodos anticonceptivos, Profesionales de la Salud.

VARIABLES CUANTITATIVAS ó NUMÉRICAS

Son variables que se expresan numéricamente. Son susceptibles de medirse

Ejemplos:

Número de hijos, Número de camas, Peso, Talla, Edad, Glucosa, Colesterol, etc.