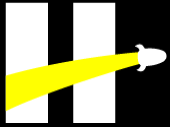


The word "HENRY" is written in a bold, black, sans-serif font. A bright yellow beam of light originates from the left edge of the frame and points directly at the letter 'R'. The beam is wider on the left and tapers as it approaches the 'R', where it ends in a small, white, circular glow.

HENRY

Estadística Descriptiva



Estadística

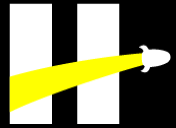
El término estadística se refiere a datos numéricos, tales como promedios, medianas, porcentajes y números índices que ayudan a entender una gran variedad situaciones.

En un sentido amplio, la estadística se define como el arte y la ciencia de reunir datos, analizarlos, presentarlos e interpretarlos. Esto ayuda a las personas que deben tomar decisiones una mejor comprensión del entorno, permitiéndoles así tomar mejores decisiones con base en mejor información.



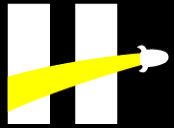
Estadística Descriptiva

La mayor parte de la información estadística en periódicos, revistas, informes de empresas y otras publicaciones consta de datos que se resumen y presentan en una forma fácil de leer y de entender. A estos resúmenes de datos, que pueden ser tabulares, gráficos o numéricos se les conoce como estadística descriptiva. Mediante estadísticos se pueden describir y comprender un entorno.



Estadística Inferencial

Una de las principales contribuciones de la estadística es emplear datos de una muestra para hacer estimaciones y probar hipótesis acerca de las características de una población mediante un proceso al que se le conoce como inferencia estadística.



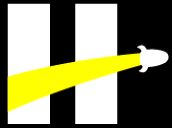
Población y muestra

Cuando se examina un grupo entero o universo completo de observaciones, se lo denomina población. Cuando se examina una pequeña parte del grupo, lo llamamos muestra.

El concepto de población en Estadística va más allá de la clásica definición que se da en la Demografía.

En la actividad estadística una población puede estar constituida por elementos de cualquier tipo, no solamente por seres humanos.

Por ejemplo, se puede hablar de la población de viviendas de un barrio; de la población de comprobantes contables de una empresa; de la población de alumnos en Henry, etc.



Distribución de frecuencias

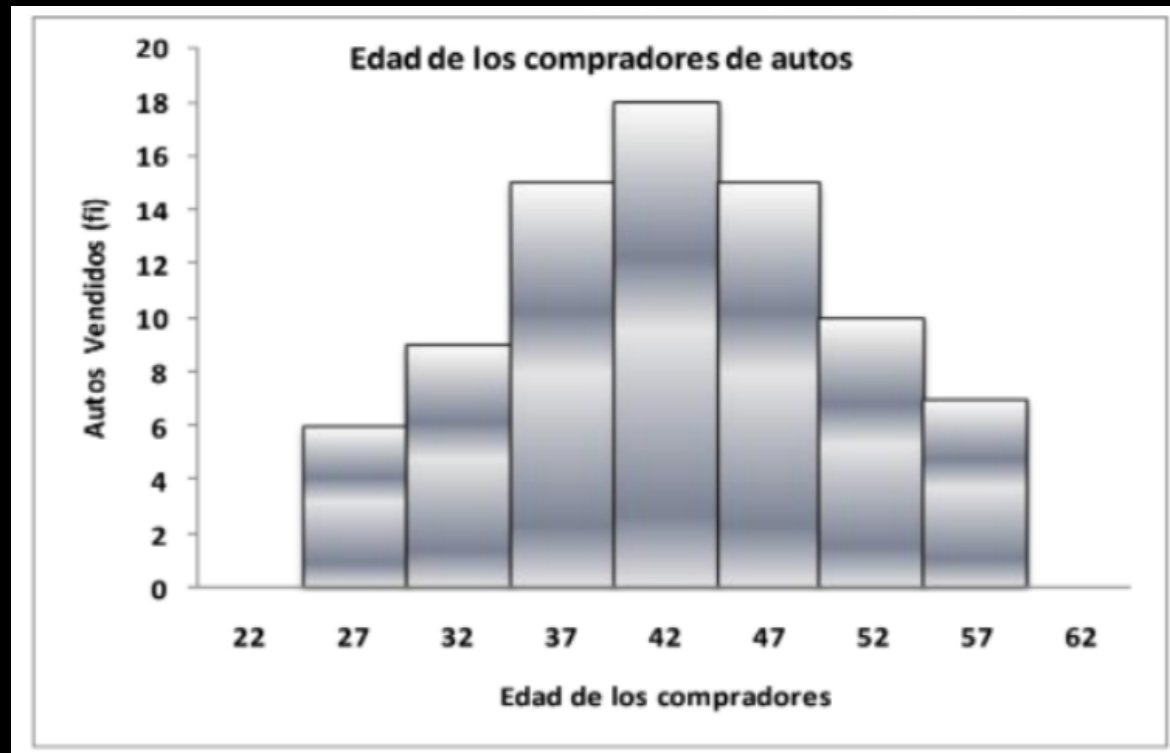
La distribución de frecuencias es una forma de presentación de los datos que facilita su tratamiento conjunto y permite una comprensión diferente de ellos. Es una tabla de datos en base a observaciones (frecuencias). La frecuencia es el número de casos que pertenecen a un valor determinado.

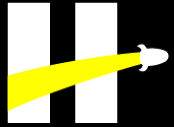
Edades de los compradores de automóviles								
Edades	Nº de autos f_i	Verdadero Limite \underline{VL}_i	Punto Medio x_i	Frecuencia Acumulada Menor que $F_i^{(-)}$	Frecuencia Acumulada Mayor que $F_i^{(+)}$	Frecuencia Relativa h_i	Frecuencia Relativa Acumulada Menor que $H_i^{(-)}$	Frecuencia Relativa Acumulada Mayor que $H_i^{(+)}$
25 - 29	6	24,5	27	6	80	7,50%	7,50%	100,00%
30 - 34	9	29,5	32	15	74	11,25%	18,75%	92,50%
35 - 39	15	34,5	37	30	65	18,75%	37,50%	81,25%
40 - 44	18	39,5	42	48	50	22,50%	60,00%	62,50%
45 - 49	15	44,5	47	63	32	18,75%	78,75%	40,00%
50 - 54	10	49,5	52	73	17	12,50%	91,25%	21,25%
55 - 59	7	54,5	57	80	7	8,75%	100,00%	8,75%
	80					100,00%		



Histograma

El Histograma es un gráfico de la distribución de frecuencias, que se construye con rectángulos de superficie proporcional al producto de la amplitud por la frecuencia absoluta (o relativa) de cada uno de los intervalos de clase.

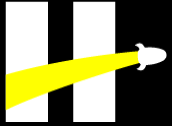




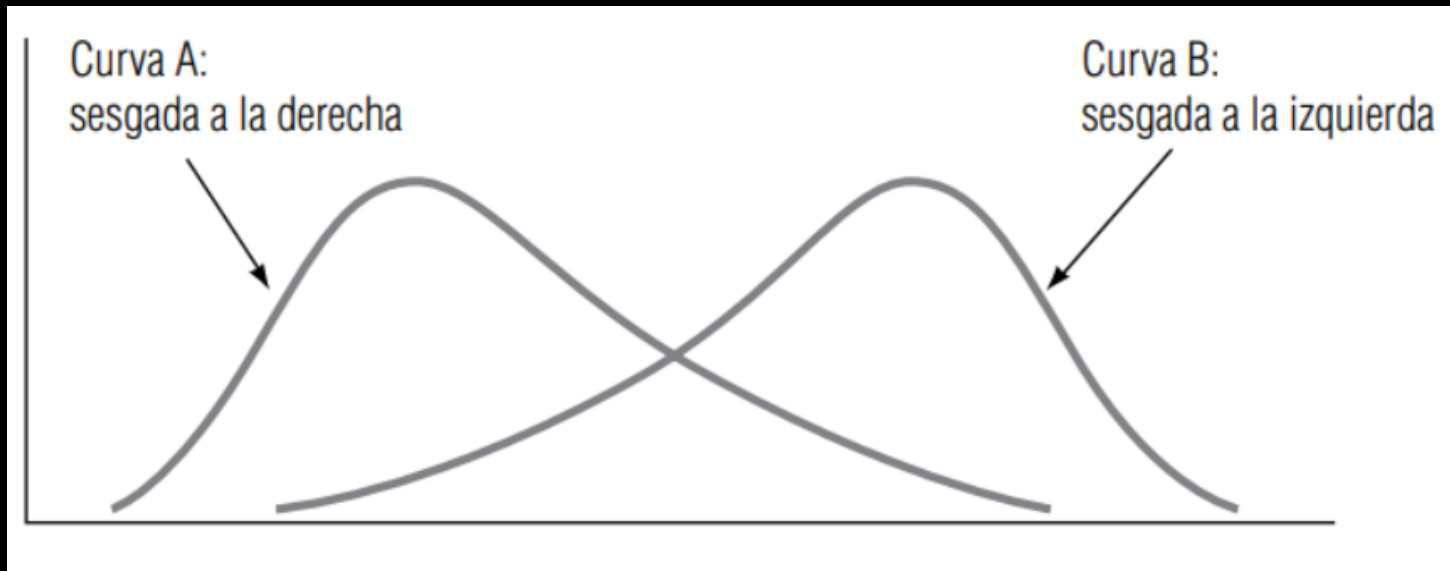
Medidas de tendencia central

La tendencia central se refiere al punto medio de una distribución.

El sesgo se produce cuando al trazar una línea vertical que pase por el punto más alto de la curva dividirá su área en dos partes que no son iguales. Cuando se da el caso de que cada parte es una imagen de espejo de la otra, esta curva se denomina simétrica. Si la curva esta sesgada hacia la derecha, se considera positivamente sesgada y si el sesgo se pronuncia hacia la izquierda, se denomina negativamente sesgada.



Sesgos





Media aritmética

Es la suma de los valores de todas las observaciones, dividido la cantidad de elementos de la muestra.

Media aritmética de la población

$$\mu = \frac{\sum x}{N}$$

Suma de los valores de todas las observaciones

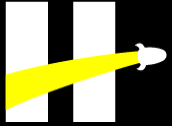
Número de elementos de la población

Media aritmética de la muestra

$$\bar{x} = \frac{\sum x}{n}$$

Suma de los valores de todas las observaciones

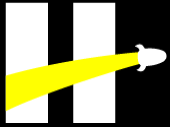
Número de elementos de la muestra



Ventajas - Desventajas

- Un solo número que representa a un conjunto de datos completo.
- Concepto familiar.
- Es única debido a que cada conjunto de datos posee una y sólo una media.
- Es útil para la comparación de medias de varios conjuntos de datos.

- Puede verse afectada por valores extremos.
- Para grandes conjuntos de datos que no están agrupados en frecuencias relativas resulta un cálculo tedioso.
- Cuando existen valores de clase extremos abiertos ("60 años o más", "18 años o menos", etc.) no se puede calcular.

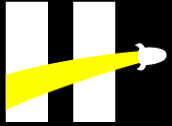


La Mediana

Es un solo valor del conjunto de datos que mide la observación central del conjunto. Es decir que esta sola observación es el elemento que está más al centro del conjunto de números, la mitad de los elementos están por arriba de este punto y la otra mitad está por debajo. Entendiendo al conjunto de datos como la cantidad total de observaciones.

Mediana

$$\text{Mediana} = \left(\frac{\overset{\text{Número de elementos del arreglo}}{\nearrow} n + 1}{2} \right)\text{-ésimo término del arreglo de datos}$$



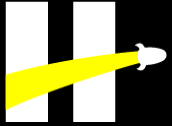
Ventajas - Desventajas

- No se ve afectada por valores extremos.
- Es fácil de entender y se puede calcular a partir de cualquier tipo de datos.
- La podemos encontrar incluso cuando nuestros datos son descripciones cualitativas, en lugar de números.
- Ciertos procedimientos estadísticos que utilizan la mediana son más complejos que aquellos que utilizan la media.
- Debemos ordenar los datos antes de llevar a cabo cualquier cálculo.



La Moda

La moda es el valor que más se repite en el conjunto de datos. Como en todos los demás aspectos de la vida, el azar puede desempeñar un papel importante en la organización de datos. En ocasiones, el azar hace que un solo elemento no representativo se repita lo suficiente para ser el valor más frecuente del conjunto de datos. Es por esto que rara vez utilizamos la moda de un conjunto de datos no agrupados como una medida de tendencia central.



Ventajas y Desventajas

- Se puede utilizar como una posición central para datos tanto cualitativos como cuantitativos.
- También, al igual que la mediana, los valores extremos no afectan indebidamente a la moda.
- La podemos utilizar aun cuando una o más clases sean de extremo abierto.
- No se utiliza tan a menudo como medida de tendencia central.
- A veces, no existe un valor modal debido a que el conjunto de datos no contiene valores que se presenten más de una vez.
- Cuando los conjuntos de datos contienen dos, tres o más modas, es difícil interpretarlos y compararlos.

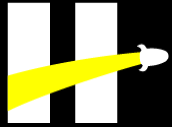


Resumen

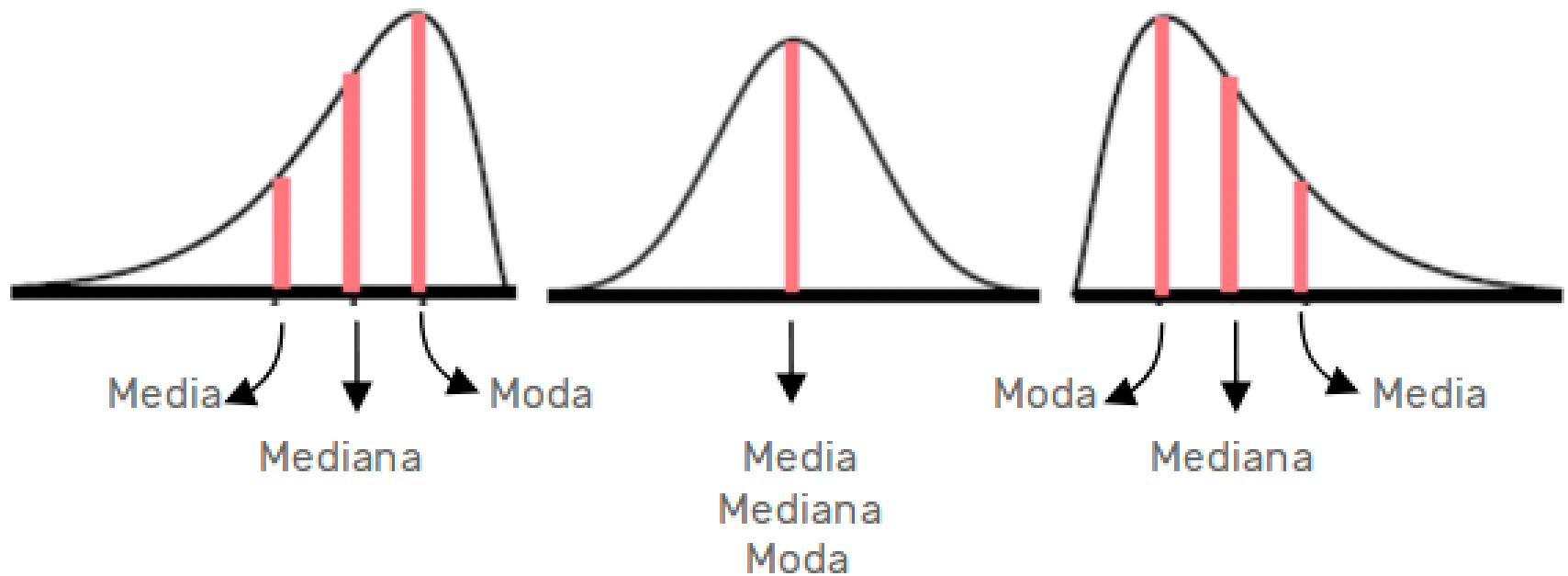
Las distribuciones simétricas que sólo contienen una moda siempre tienen el mismo valor para la media, la mediana y la moda.

Sesgo positivo (es decir, sesgada a la derecha): la moda se encuentra en el punto más alto de la distribución, la mediana está a la derecha de la moda y la media se encuentra todavía más a la derecha de la moda y la mediana.

Sesgo negativo (es decir, sesgada a la izquierda), la moda se encuentra en el punto más alto de la distribución, la mediana está a la izquierda y la media se encuentra todavía más a la izquierda de la moda y la mediana.



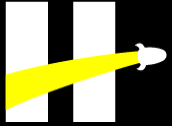
Resumen





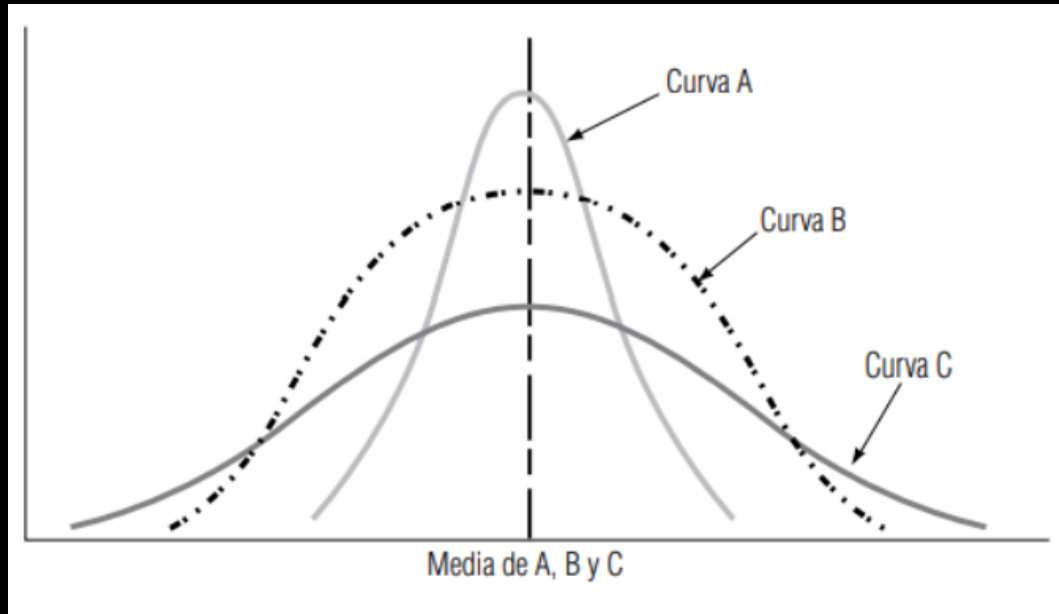
Resumen

Cuando la población está sesgada negativa o positivamente, la mediana suele ser la mejor medida de posición, debido a que siempre está entre la moda y la media. La frecuencia de ocurrencia de un solo valor no influye mucho en la mediana como es el caso de la moda, ni la distorsionan los valores extremos como la media. En cualquier otro caso, no existen guías universales para la aplicación de la media, la mediana o la moda como medidas de tendencia central. Toca a los analistas decidir cual describe mejor el conjunto de datos.



La dispersión

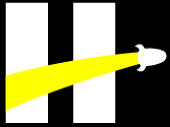
Al igual que sucede con cualquier conjunto de datos, la media, la mediana y la moda sólo nos revelan una parte de la información que debemos conocer acerca de las características de los datos. Para aumentar nuestro entendimiento del patrón de los datos, debemos medir también su dispersión, separación o variabilidad.





El Rango

El rango es la diferencia entre el más alto y el más pequeño de los valores observados. Sólo toma en cuenta los valores más alto y más bajo de una distribución y ninguna otra observación del conjunto de datos. Como resultado, ignora la naturaleza de la variación entre todas las demás observaciones, y tiene una gran influencia de los valores extremos.



La Varianza

Es la suma de los cuadrados de las distancias entre la media y cada elemento de la población, dividido entre el número total de observaciones.

Al elevar al cuadrado cada distancia, logramos que todos los números sean positivos y, al mismo tiempo, asignamos más peso a las desviaciones más grandes (desviación es la distancia entre la media y un valor).

La fórmula para calcular la varianza es:

Varianza de población

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N} = \frac{\sum x^2}{N} - \mu^2$$

donde:

- σ^2 = varianza de la población
- x = elemento u observación
- μ = media de la población
- N = número total de elementos de la población
- Σ = suma de todos los valores $(x - \mu)^2$, o todos los valores x^2



La Desviación Estándar

Para poder realizar una interpretación intuitiva de la varianza, debemos hacer un cambio significativo. Esta medida se conoce como la desviación estándar y es la raíz cuadrada de la varianza. La desviación estándar, entonces, queda en las mismas unidades que los datos originales.

Desviación estándar de la población

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x - \mu)^2}{N}} = \sqrt{\frac{\sum x^2}{N} - \mu^2}$$

donde,

- x = observación
- μ = media de la población
- N = número total de elementos de la población
- Σ = suma de todos los valores $(x - \mu)^2$, o todos los valores x^2
- σ = desviación estándar de la población
- σ^2 = varianza de la población



Coeficientes de variación

El coeficiente de variación es una medida relativa de dispersión. Relaciona la desviación estándar y la media, expresando la desviación estándar como porcentaje de la media. La unidad de medida, entonces, es “porcentaje”, en lugar de las unidades de los datos originales.

Coeficiente de variación

$$\text{Coeficiente de variación de la población} = \frac{\sigma}{\mu} (100)$$

Desviación estándar de la población $\rightarrow \sigma$

Media de la población $\rightarrow \mu$