

Análisis estadístico
Unidad 1

MSc. Francisco Paniagua



Contenidos

Unidad 1 Introducción a la Estadística



Concepto de Estadística

- La **Estadística es la ciencia** cuyo objetivo es reunir una información cuantitativa concerniente a individuos, grupos, series de hechos, etc. y deducir de ello gracias al análisis de estos datos unos significados precisos o unas previsiones para el futuro.
- *La estadística, en general, es la ciencia que trata de la recopilación, organización presentación, análisis e interpretación de datos numéricos con e fin de realizar una toma de decisión más efectiva.*



Hablar con datos

- “Hablar” con datos constituye la única posibilidad de seguir un proceso racional de solución de problemas.
- Identificar el problema de raíz, evitando que se desperdicie esfuerzo, tiempo y dinero en los síntomas.
- Hablar en términos de procesos y responsabilidades, no de culpas.



Tipos de datos

Variables cuantitativas

Continuas

- Toma cualquier valor.
- Puede ser decimal.
- Ejemplo: temperatura, presión, altura, edad.

Discretas

- Toma valores enteros.
- Ejemplo: número de alumnos, cantidad de estudiantes.

Variables cualitativas

Discreta Nominal

- Categorías no ordenadas.
- Ejemplo: género, grupo sanguíneo, lugar de nacimiento.

Discreta Ordinal

- Categoría ordenada.
- Ejemplo: escolaridad, grado de satisfacción.



Análisis Descriptivo de los datos

Gráficas:

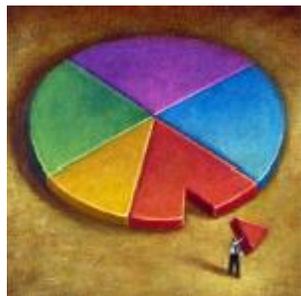
Permiten visualizar los resultados obtenidos

Medidas de Variabilidad:

Determinan la cantidad de variación de la variable; si los datos son o no dispersos

Medidas de Tendencia Central:

Describen alrededor de que valores fluctúan los datos de la variable



ESTADÍSTICA

Concepto:

Fórmulas de estadística descriptiva

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

Descriptiva

RECOLECCIÓN

TABULACIÓN

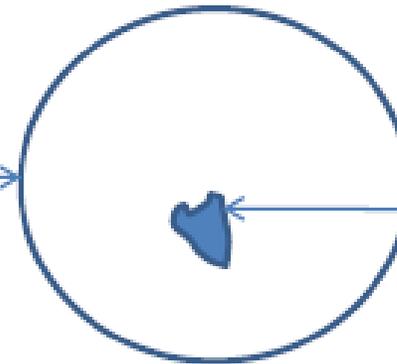
ANÁLISIS

REPRESENTACIÓN
GRÁFICA

CONCLUSIONES

Inferencial

POBLACIÓN (S)



MUESTRA

Las 3 características fundamentales de una muestra:

1. Representativa
2. Aleatoria
3. Estratificada

“La Estadística es un instrumento de análisis”



Estadística Descriptiva

Medidas de tendencia central

Media

Mediana

Moda

Medidas de variabilidad

Rango

Varianza

Desviación estándar

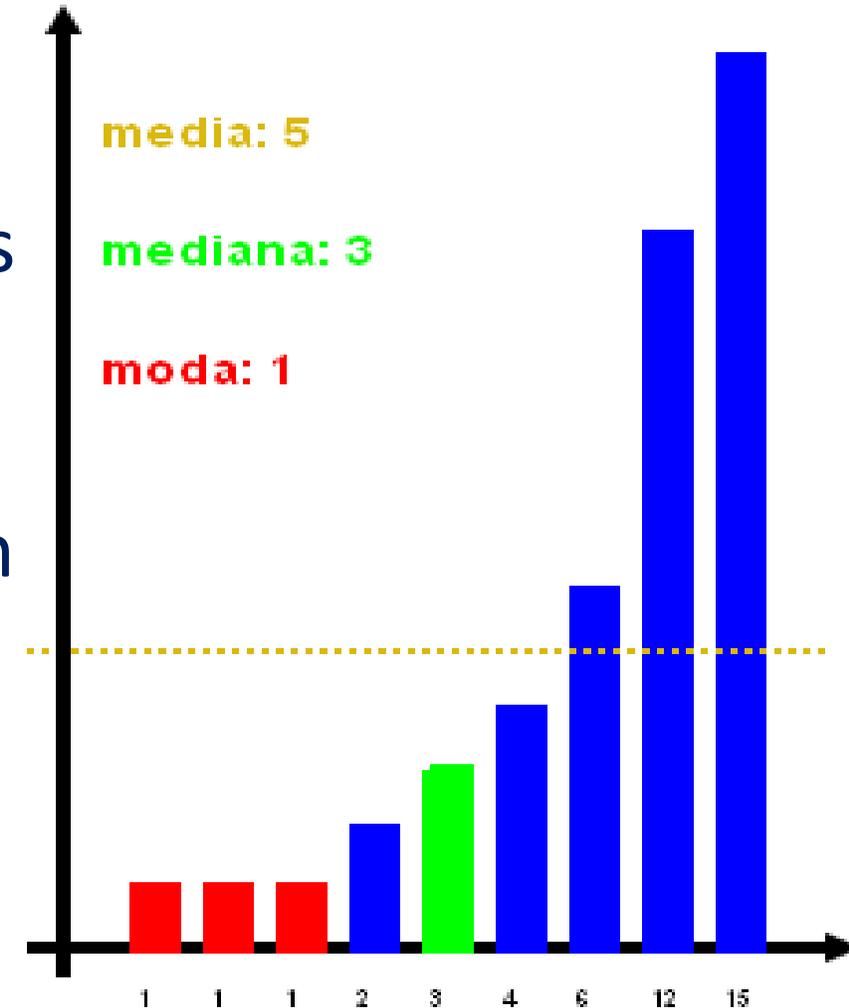


Estadística Descriptiva

Medidas de tendencia central

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

- Media: se obtiene de dividir la sumatoria de los valores individuales entre el total de datos.
- Mediana: es el centro de los datos en un conjunto ordenando.
- Moda: el valor que más se repite.

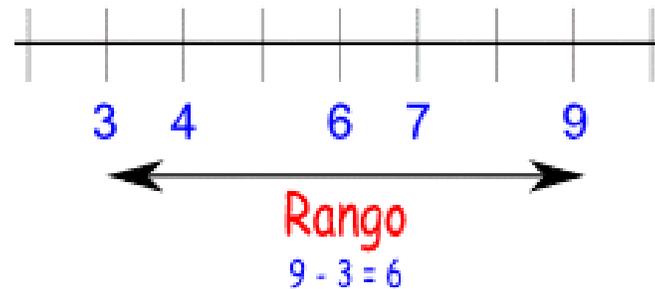


Estadística Descriptiva

Medidas de variabilidad

- Rango: da la información de la variabilidad total o distancia total del conjunto de datos.

$$R = \text{Max } x_i - \text{Min } x_i$$

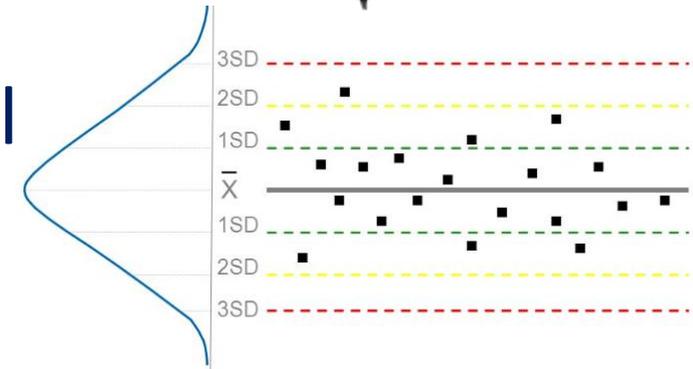


Estadística Descriptiva

Medidas de variabilidad

- Desviación estándar: es la variación promedio del conjunto de datos con respecto a la media.
- Varianza: proporciona la dispersión total de los datos. Se expresa en unidades cuadradas.

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



$$\sigma_X^2 = \frac{\sum_{i=1}^N (X_i - \mu_X)^2}{N}$$

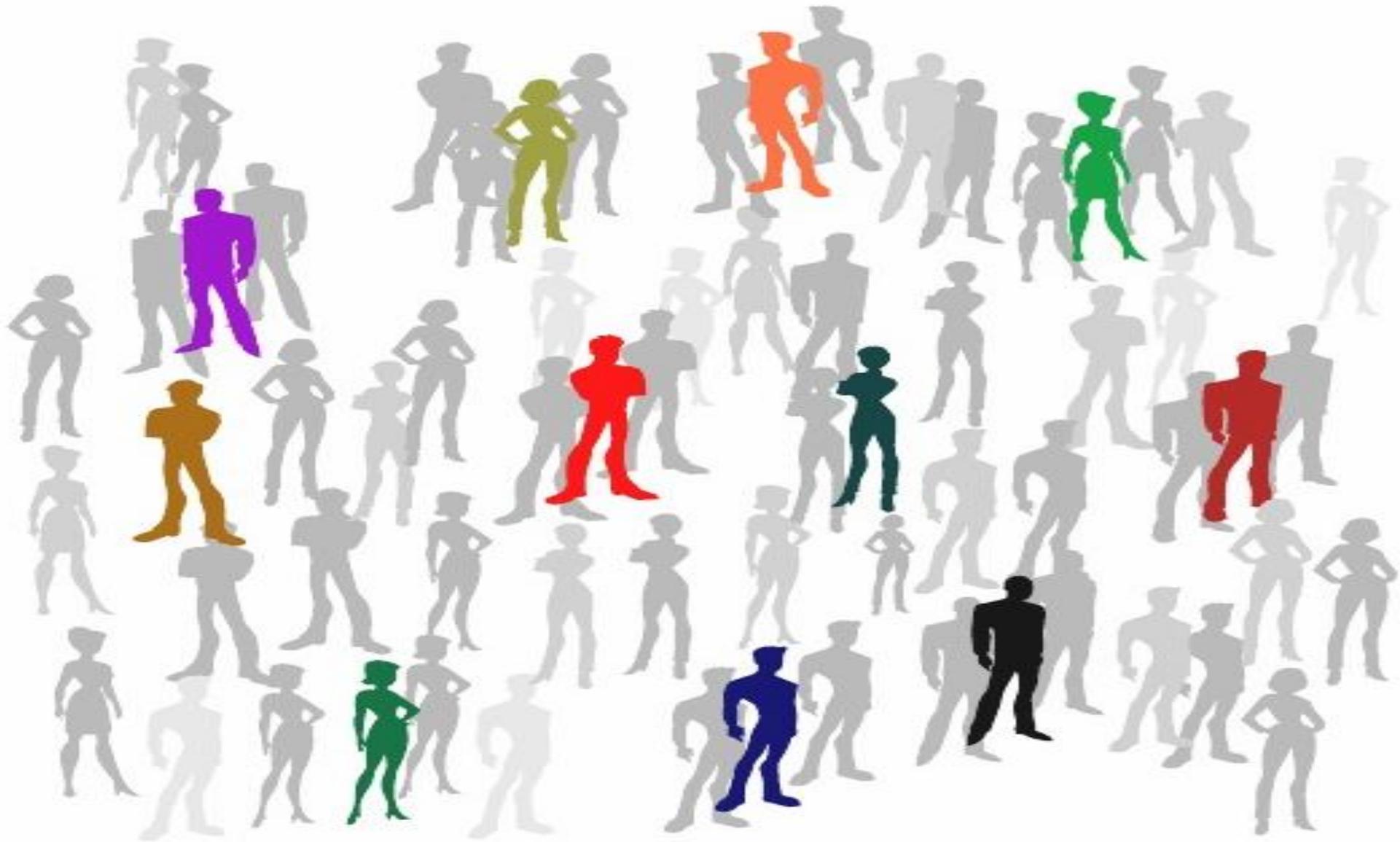


Aproximación conceptual

Características colectivas

- Supongamos un grupo de personas que comparten características comunes, obviamente observaremos también características diferenciales.
- Supongamos, además, que estamos realizando la tarea de cuantificar esas respectivas características.
- Se representa la situación así:



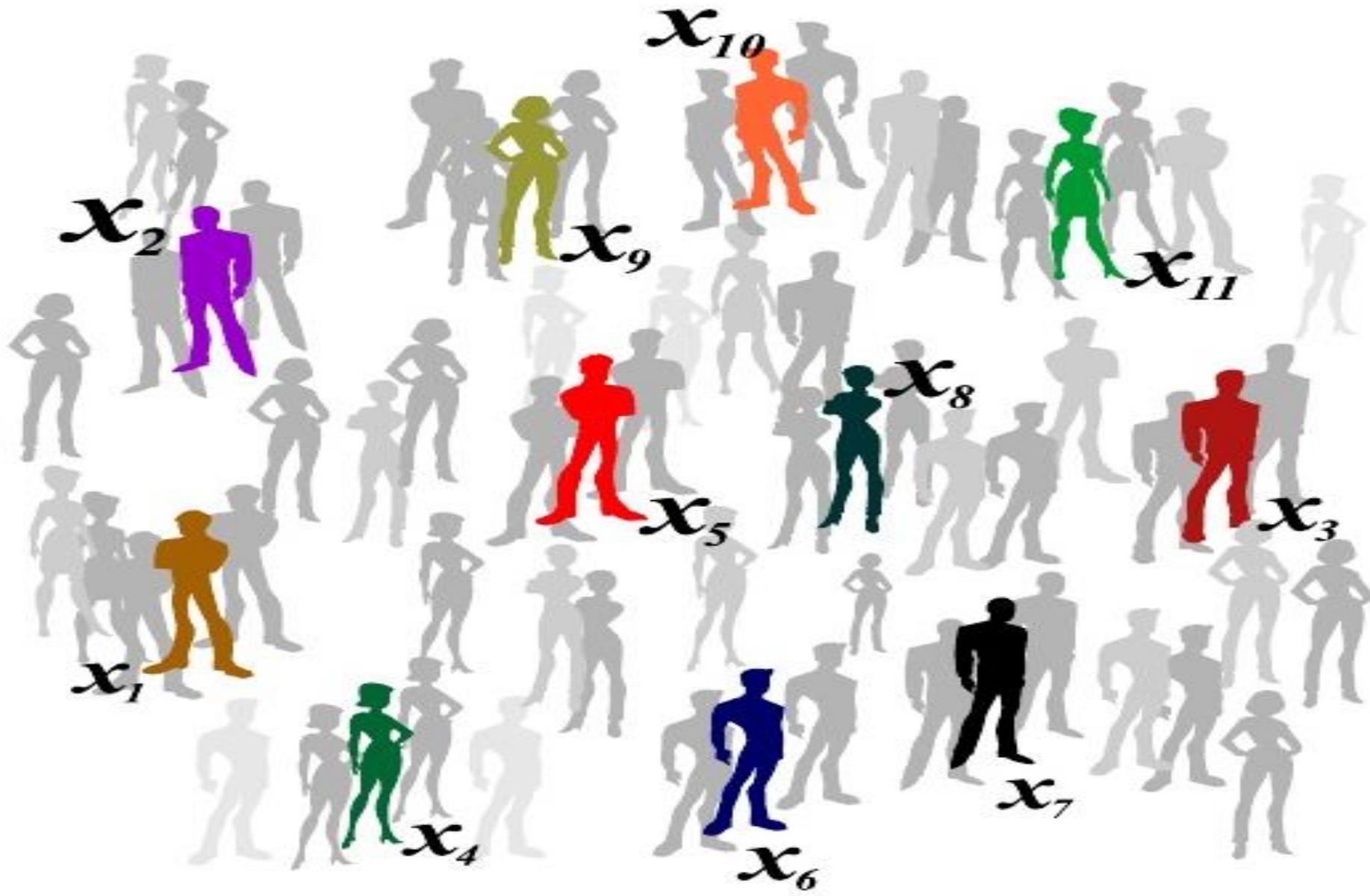


Aproximación conceptual

Cuantificación de variables

- Denominemos estas variables evaluadas o cuantificadas “X”, asignándole valores según un sistema específico.
- En la gráfica anterior el grupo de personas se conformaría con asignaciones diferentes de “X”, lo cual podríamos representar de la siguiente manera:





Aproximación conceptual

Medidas “promedio”

- El conjunto de medidas podrían ser resumidas mediante promedios o medidas de tendencia central.
- En este caso utilizamos la media aritmética (pudo haber sido la mediana o la moda). Ese valor está representado por un valor central y un “caso modelo” que representaría las características promedio del grupo:





Aproximación conceptual

Dispersión de la media

- Una vez calculada la medida promedio resulta fácil notar que existe una diferencia (distancia) entre las medidas individuales y el promedio.
- Cuanto mayor dispersión se observe (distancia respecto a la media) menos homogéneas son las observaciones.
- La dispersión puede cuantificarse calculando la diferencia entre las medidas individuales y el promedio.





Aproximación conceptual

Muestreo

- Supongamos que el grupo de personas pertenecen a un grupo de dentro del cual comparten características comunes.
- Entendemos que estos aspectos comunes hacen que sean más similares entre sí. Teóricamente, si dependieran de esos factores comunes las características deberían ser iguales a la media.



Aproximación conceptual

Muestreo

- Pero se conocen otros factores que influyen en las diferencias individuales como el peso, altura, edad, etc. Dicho de otro modo, las puntuaciones varían conforme fueran más altos o pequeños, más jóvenes o mayores que la edad promedio.
- Estas variables explican parte de las variaciones o desviaciones de la media.





Distribución normal

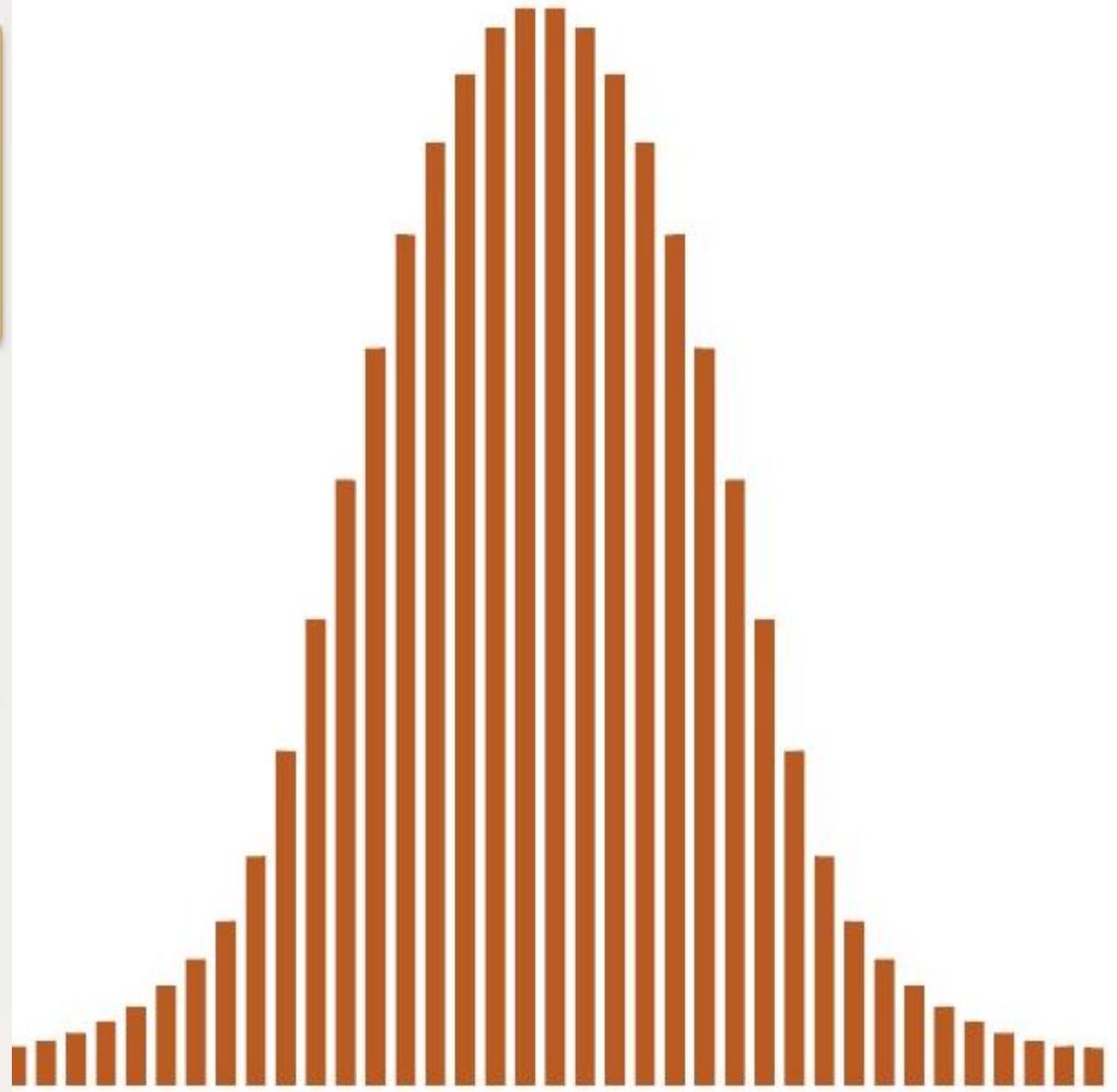
Distribución de probabilidades

- La forma en cómo se distribuyen las probabilidades de ocurrencia de un determinado fenómeno ha generado modelos que facilitan mecanismos de estimación.
- Uno de los modelos cuyo uso se ha extendido es el de la distribución normal, cuya forma intuitiva se representa a continuación:



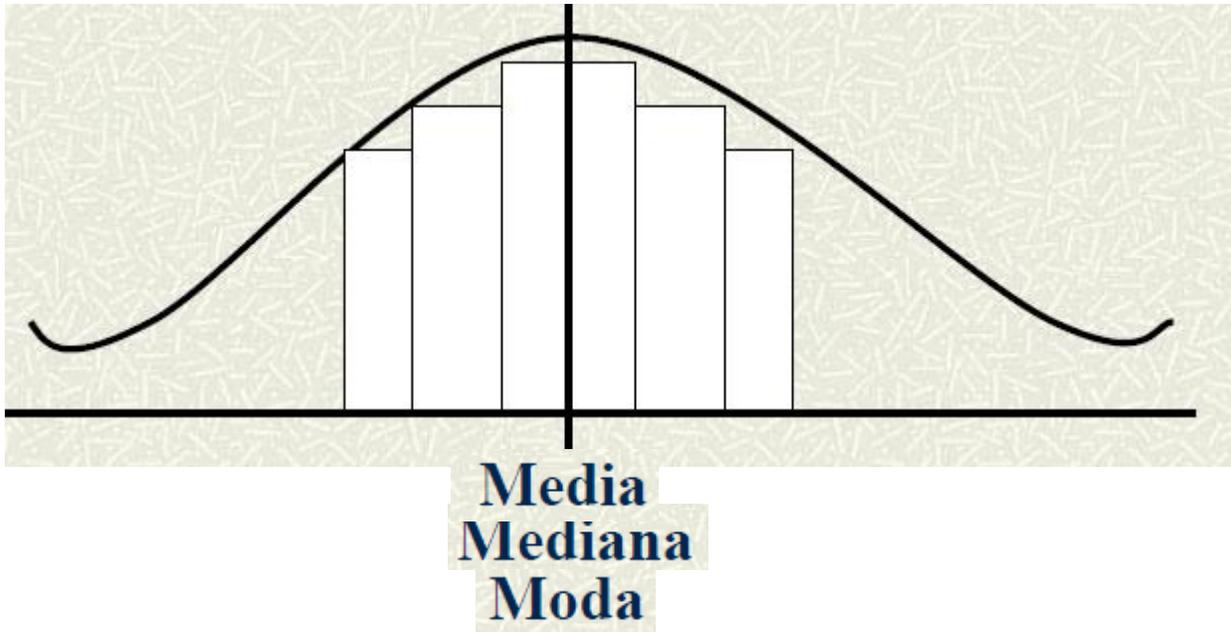


Aquí se puede observar que la mayoría de los casos observados presentan valores cercanos a la media.



Distribución normal

Simetría

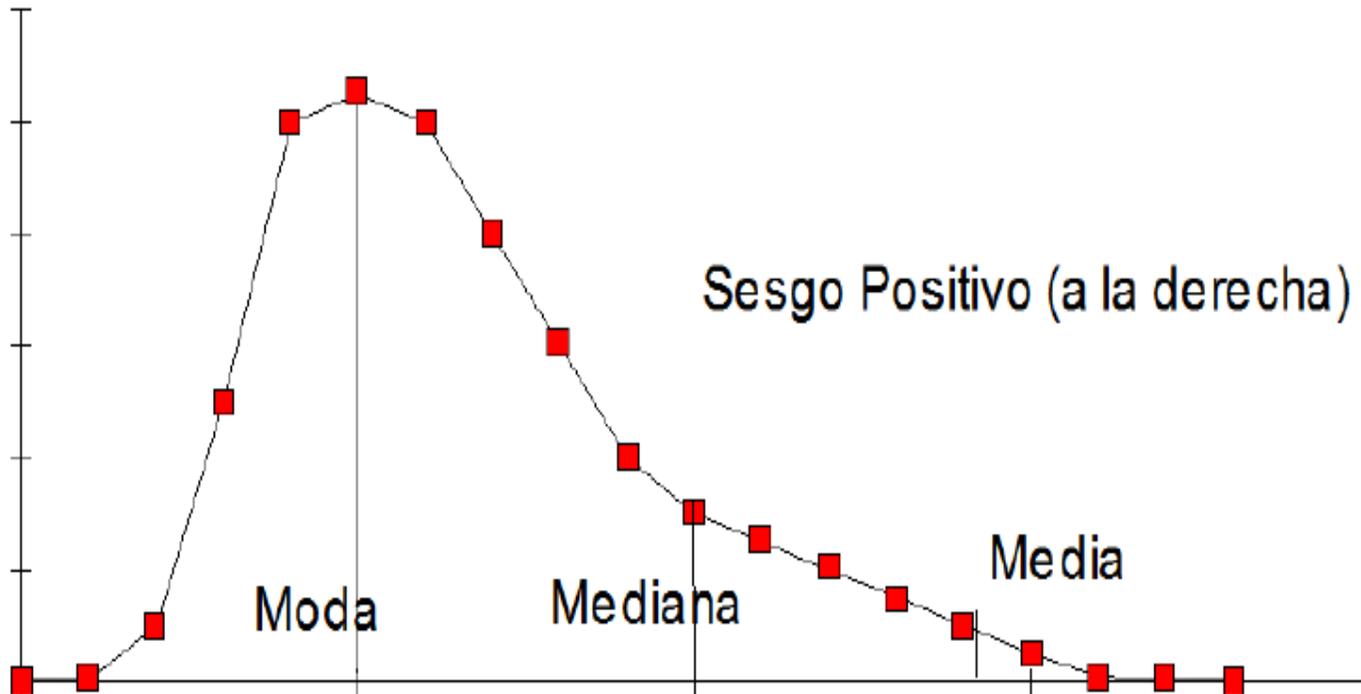


- Curva simétrica o normal.
- $\text{Media} = \text{Mediana} = \text{Moda}$



Distribución normal

Sesgo positivo



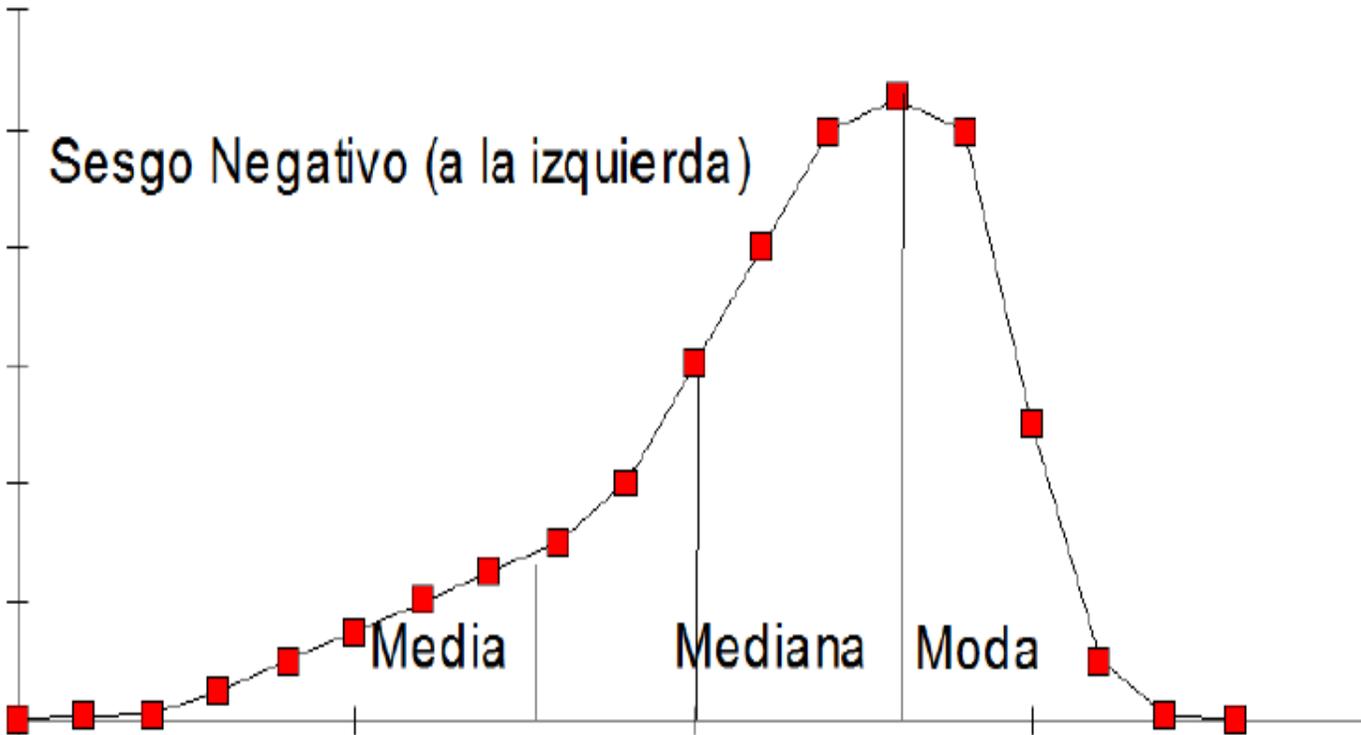
- La mayoría de los datos se encuentran por debajo de la media.



Distribución normal

Sesgo negativo

Sesgo Negativo (a la izquierda)



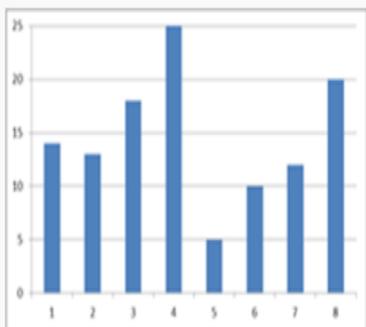
- La mayoría de los datos se encuentran por arriba de la media.



Distribución normal

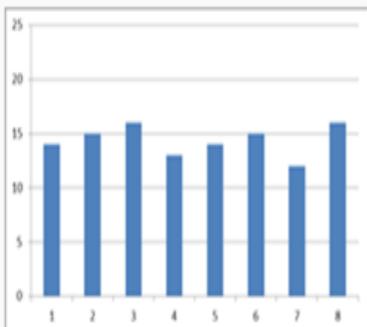
Dispersión → Desviación estándar

Muestra 1



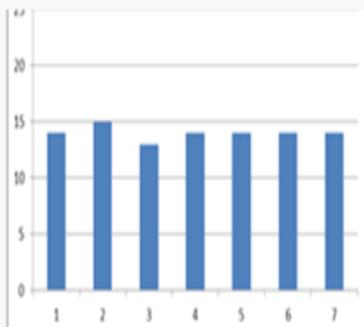
$\sigma = 6,82$

Muestra 2



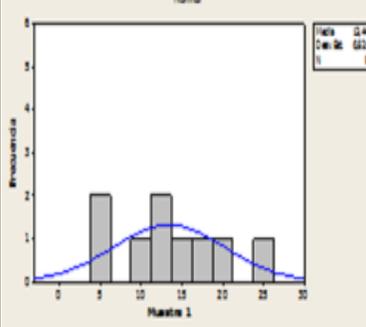
$\sigma = 1,32$

Muestra 3

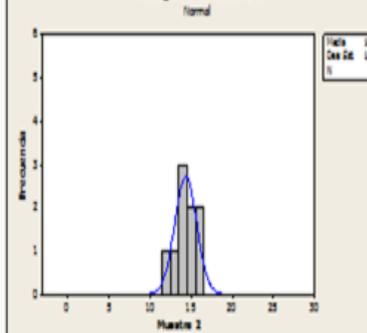


$\sigma = 0,61$

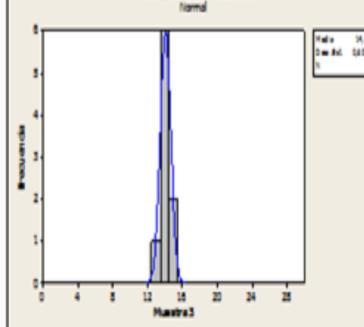
Histograma de Muestra 1



Histograma de Muestra 2



Histograma de Muestra 3



- A mayor dispersión mayor será la desviación estándar, por lo tanto el proceso será mejor a medida que mejore su dispersión o su desviación estándar.

- La campana de Gauss (línea Azul) debajo de cada uno de los diagramas de barras, se torna más altas y menos ancha a medida que la dispersión disminuye.



Ejercicios Unidad 1

- Estadística descriptiva



■ Ejercicio. Estadísticas descriptivas

Un ingeniero especializado en control de calidad debe garantizar que las tapas de las botellas de champú queden ajustadas correctamente. Si las tapas quedan flojas, podrían caerse durante el envío. Si se aprietan demasiado, será difícil retirarlas. El valor objetivo del par de torsión para ajustar las tapas es 18. El ingeniero recolecta una muestra aleatoria de 68 botellas y prueba la cantidad de par de torsión que se necesita para quitar las tapas.

Como parte de la investigación inicial, el ingeniero analiza los estadísticos descriptivos y las gráficas para evaluar la distribución de los datos.

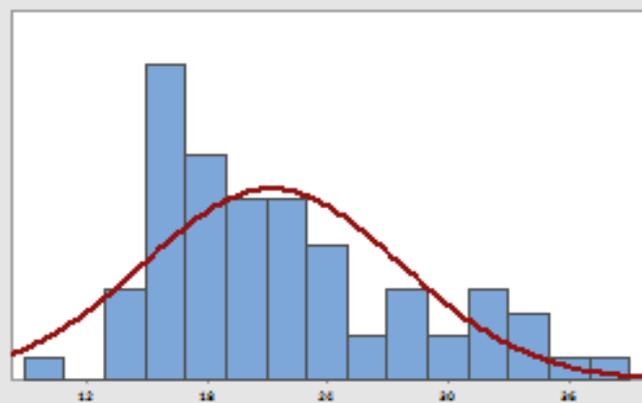
1. Abra los datos de muestra, *TorsiónTapa.MTW*.
2. Elija **Estadísticas > Estadísticas básicas > Resumen gráfico**.
3. En **Variables**, ingrese *Torsión*.
4. Haga clic en **Aceptar**.



Interpretar los resultados

La torsión media que se requiere para retirar las tapas es 21.265, con una desviación estándar de 6.422. El histograma muestra una amplia dispersión en la que muchos valores se ubican lejos del valor objetivo de 18. El intervalo de confianza de 95% muestra que un rango probable para la torsión media de la población está entre 19.710 y 22.819, que es mayor que el objetivo de 18. Para investigar la diferencia entre la torsión media y el valor objetivo, el ingeniero podría utilizar una prueba t de 1 muestra.

Informe de resumen de Torsión



Intervalos de confianza de 95%



Prueba de normalidad de Anderson-Darling

A-cuadrado	2.04
Valor p	<0.005

Media	21.265
Desv.Est.	6.422
Varianza	41.242
Asimetría	0.786618
Curtosis	-0.288434
N	68

Mínimo	10.000
1er cuartil	16.000
Mediana	20.000
3er cuartil	24.750
Máximo	37.000

Intervalo de confianza de 95% para la media

19.710	22.819
--------	--------

Intervalo de confianza de 95% para la mediana

17.000	21.521
--------	--------

Intervalo de confianza de 95% para la desviación estándar

5.495	7.729
-------	-------



Instrucciones Unidad 1

1. Seleccione 2 bases de datos, de la carpeta de Ejercicios del campus y realice una revisión de los datos (sanity check). Para cada base de datos realice una lista de 4 variables y determine el tipo de dato.
2. Seleccione 2 variables cuantitativas de las bases de datos y genere el Resumen de la Estadística Descriptiva para cada una.
3. Realice 3 conclusiones sobre las medidas de tendencia central y las medidas de variabilidad.



Bibliografía

- Besterfield, D.H. (2009) “Control de Calidad”, Prentice Hall. Octava edición.
- Evans, J. & Lindsay, W. (2008) “Administración y control de la calidad”, Internacional Thomson Editores, Séptima edición
- Gómez Barrantes Miguel, Elementos de Estadística Descriptiva, Ed EUNED, 2001
- Manual del Usuario MINITAB 17 www.Minitab.com
- Montgomery, Douglas. “Probabilidad y Estadística aplicada a la Ingeniería”. Mc Graw Hill. México, 2002.
- Moya M, Robles N. “Probabilidad y Estadística”, 2ª. Ed. Cartago, Costa Rica: Editorial Tecnológica de Costa Rica, 2010.
- Walpole et al. “Probabilidad y estadística para ingenieros”. Prentice Hall. México, 2004.

