Estadística – Tecnicatura Universitaria en Programación

ESTADÍSTICA

UNIDAD 2

Ing. Sofía Pezzi

MEDIDAS NUMÉRICAS

¿Por qué son necesarias las medidas numéricas?

- Las gráficas son sumamente útiles para la descripción visual de un conjunto de datos, pero no siempre son la mejor herramienta cuando se desea hacer inferencias acerca de una población a partir de la información contenida en una muestra.

- Medidas numéricas

Para una muestra. → Estadísticas.

Para una población. → Parámetros.

MEDIDAS DE CENTRALIZACIÓN

La **media aritmética** o **promedio** de un conjunto de n mediciones es igual a la suma de las mediciones dividida entre n.

- Medida de centralización muy utilizada

  $\bar{X}$ indica la muestra de una muestra.

  μ indica la muestra de una población.

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

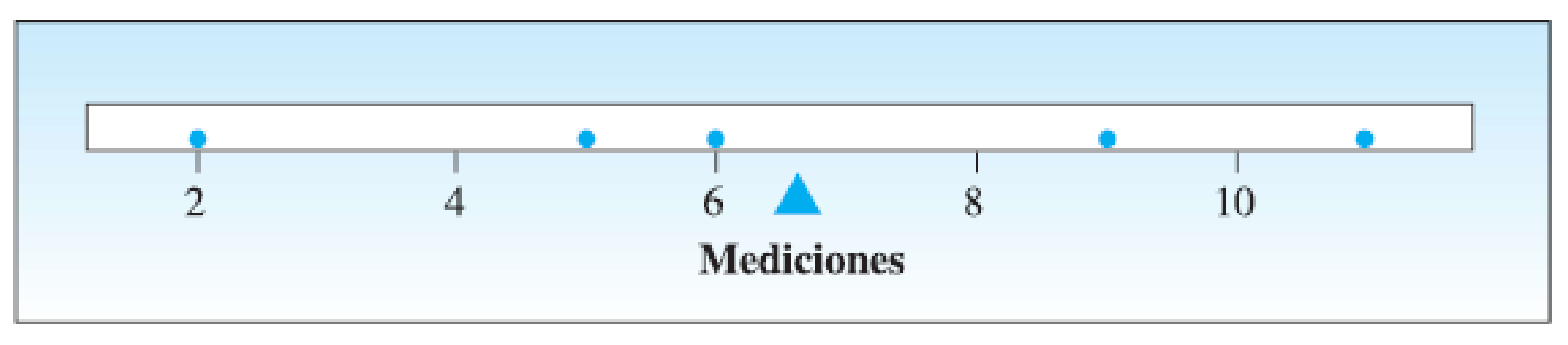
$\sum x_i$ → Sumatoria de las mediciones

n → Cantidad de mediciones

MEDIDAS DE CENTRALIZACIÓN

EJEMPLO:

Dadas las siguientes 5 mediciones: 2, 9, 11, 5, 6.



$$\bar{x} = \frac{2 + 9 + 11 + 5 + 6}{5} = \frac{33}{5}$$

$$\boxed{\bar{x} = 6{,}6}$$

El ...... promedio es de 6,6.

MEDIDAS DE CENTRALIZACIÓN

La **mediana, me** de un conjunto de n mediciones es el valor de x que cae en la posición media cuando las mediciones son ordenadas de mayor a menor.

- Posición: → $Pme = \frac{n+1}{2}$

EJEMPLO:

Dadas las siguientes 5 mediciones: 2, 9, 11, 5, 6.

2 – 5 – $\boxed{6}$ – 9 – 11

$me = 6$

El 50 % de ...... es mayor 6.

El 50 % de ...... es menor 6.

MEDIDAS DE CENTRALIZACIÓN

La **mediana, me** de un conjunto de n mediciones es el valor de x que cae en la posición media cuando las mediciones son ordenadas de mayor a menor.

EJEMPLO:

Dadas las siguientes 6 mediciones: 2, 9, 11, 5, 6, 27.

2 – 5 – 6 – 9 – 11 – 27

$$me = \frac{6+9}{2} = 7{,}5$$

El 50 % de …… 7,5.

MEDIDAS DE CENTRALIZACIÓN

- Mediana

  Buena medida de centralización.

  No es tan sensible como la media.

  No se ve afectada por valores atípicos.

EJEMPLO:

Dadas las siguientes 5 mediciones: 2, 9, 11, 5, 6.

$\bar{x} = 6,6$

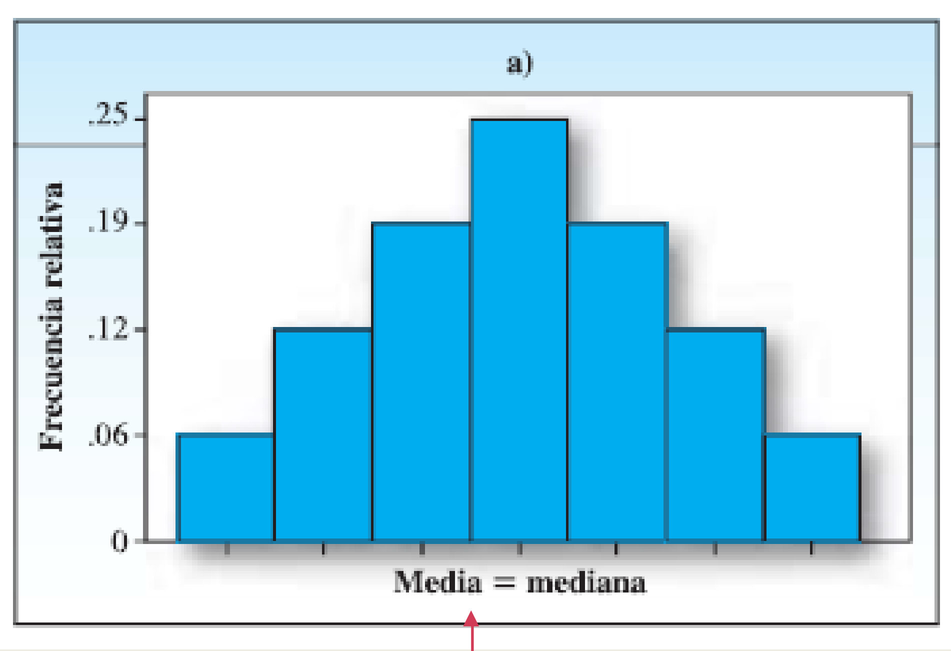
$me = 6$

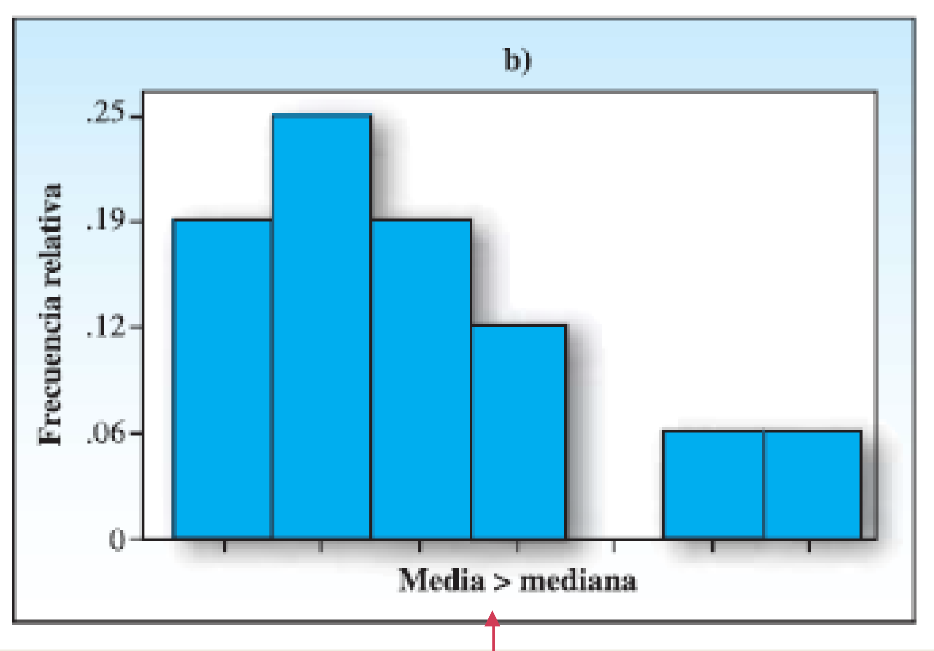
Dadas las siguientes 5 mediciones: 2, 9, 27, 5, 6.

$\bar{x} = 9,8$ → Se corre hacia los valores extremos

$me = 6$

MEDIDAS DE CENTRALIZACIÓN





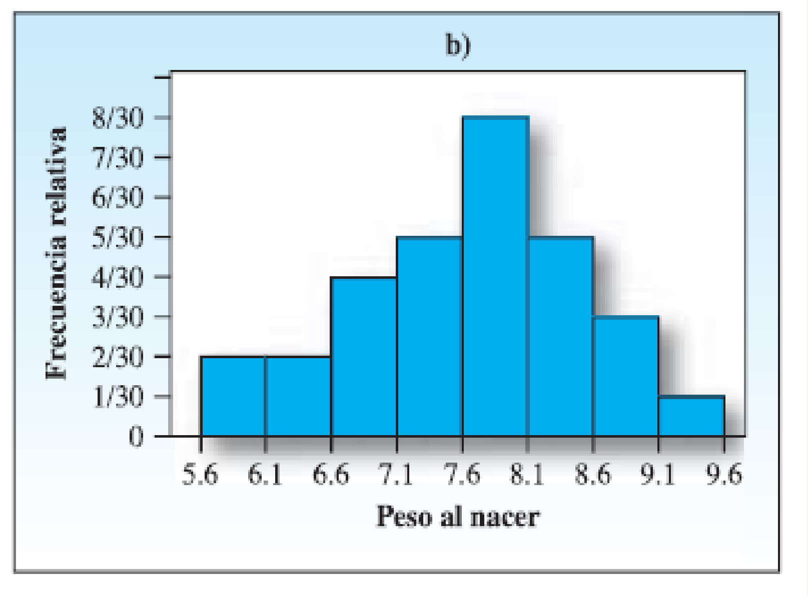
Distribución simétrica

Distribución asimétrica

MEDIDAS DE CENTRALIZACIÓN

La **moda, mo** es la categoría o medición que se presenta con mayor frecuencia.

- Datos agrupados: → Intervalo de mayor frecuencia. → CLASE MODAL
- Se utiliza para describir conjuntos grandes de datos.
- En un histograma se observa donde está el pico.



EJERCITACIÓN

1. Se toman 5 mediciones: 0, 5, 1, 1, 3.

a. Trazar una gráfica de puntos para los datos. Calcular el "centro" aproximado.

b. Encontrar la media, mediana y moda.

c. Localizar las tres mediciones de centro en la gráfica de puntos en el inciso a). Con base en las posiciones relativas de la media y mediana, ¿las mediciones son simétricas o son sesgadas?

EJERCITACIÓN

3. Se toman 10 mediciones: 3, 5, 4, 6, 10, 5, 6, 9, 2, 8.
    a. Calcular la media.
    b. Encontrar la mediana.
    c. Encontrar la moda.

EJERCITACIÓN

4. Un reproductor de discos de video es un aparato común en casi todas las casas en Estados Unidos. De hecho, casi todas las familias los tienen y muchas tienen más de uno. Una muestra de 25 familias produjo las siguientes mediciones en x, el número de los DVD en la casa

a. La distribución de x, el número de los DVD en una familia, ¿es simétrica o sesgada? Explicar

b. Calcular la media, la mediana y la moda para estas mediciones.

c. Trazar un histograma de frecuencia relativa para el conjunto de datos. Localice la media, mediana y moda a lo largo del eje horizontal. ¿Las respuestas a los incisos a) y b) son correctas?

1	0	2	1	1
1	0	2	1	0
0	1	2	3	2
1	1	1	0	1
3	1	0	1	1

EJERCITACIÓN

6. En un experimento psicológico, fue registrado el tiempo en un trabajo para 10 personas bajo una limitación de 5 minutos. Estas mediciones son en segundos:

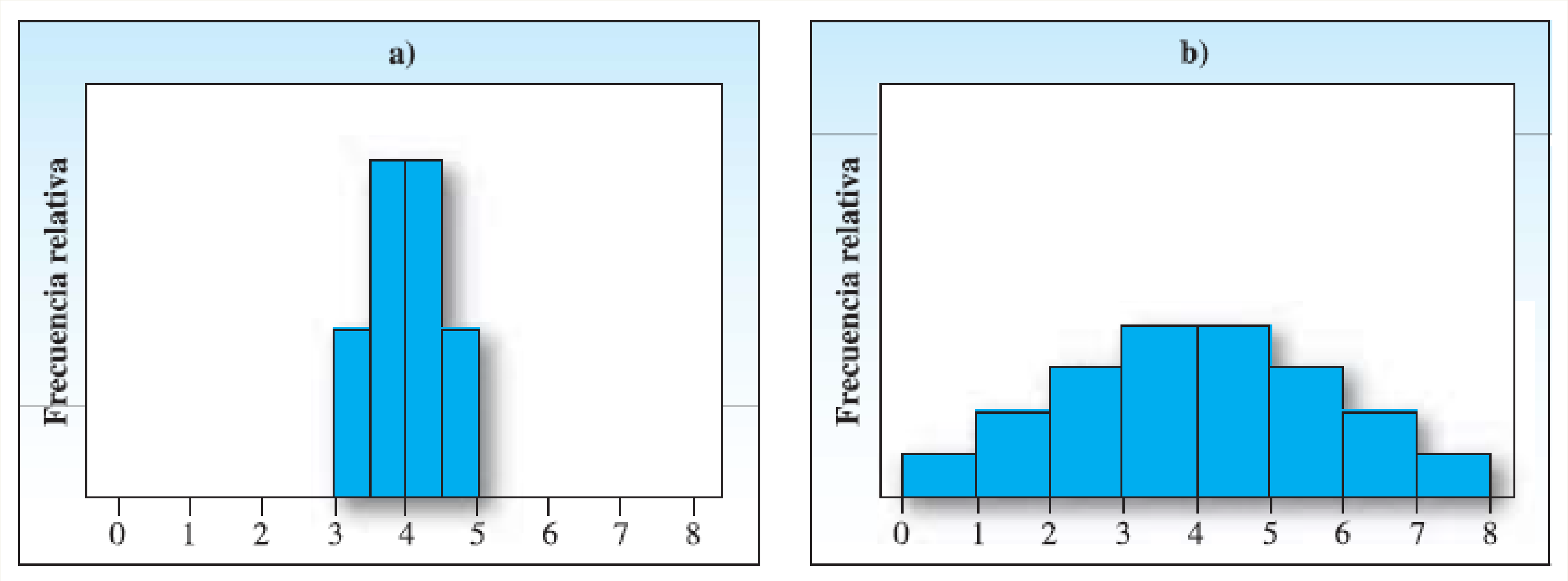
175 – 190 – 250 – 230 – 240 – 200 – 185 – 190 – 225 – 265

a. Encontrar el tiempo promedio en el trabajo.
b. Encontrar la mediana del tiempo en el trabajo.
c. Si usted está escribiendo un informe para describir estos datos, ¿qué medida de tendencia central usaría?

MEDIDAS DE DISPERSIÓN

- El valor de la media sola puede no expresar realmente como es la distribución de los datos.

No tiene en cuenta la dispersión de ellos

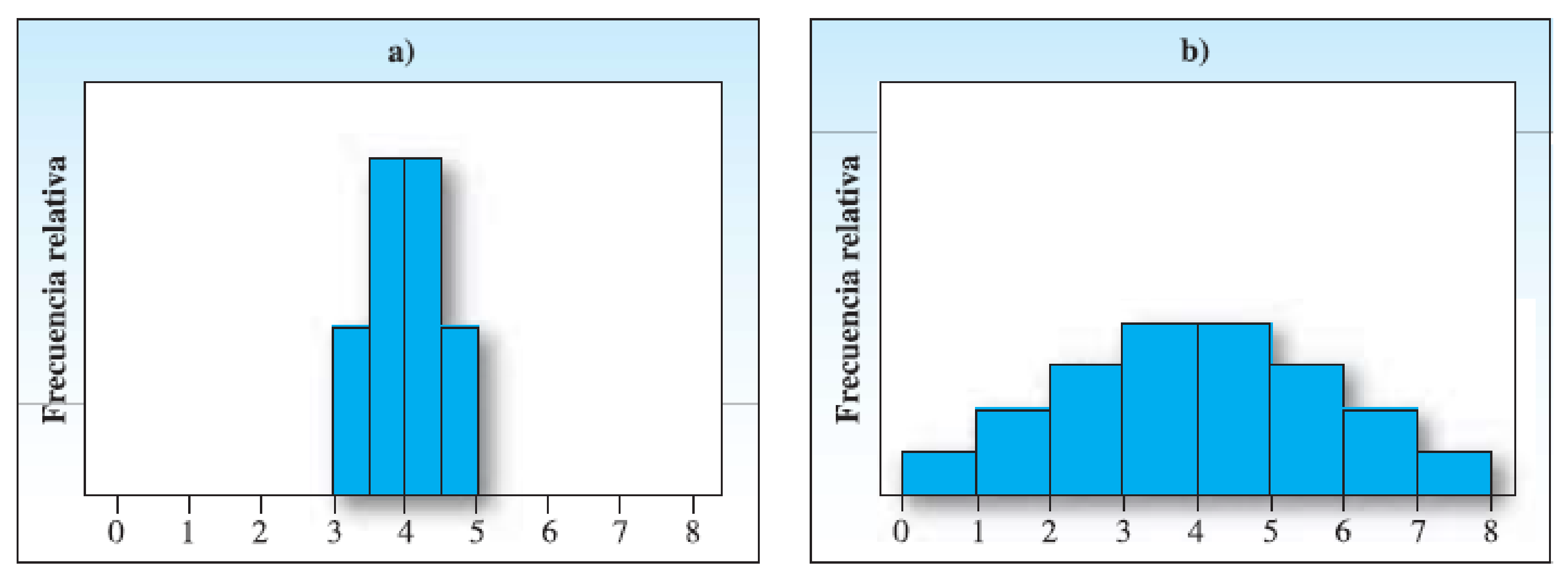


Datos menos dispersos

Datos más dispersos

MEDIDAS DE DISPERSIÓN

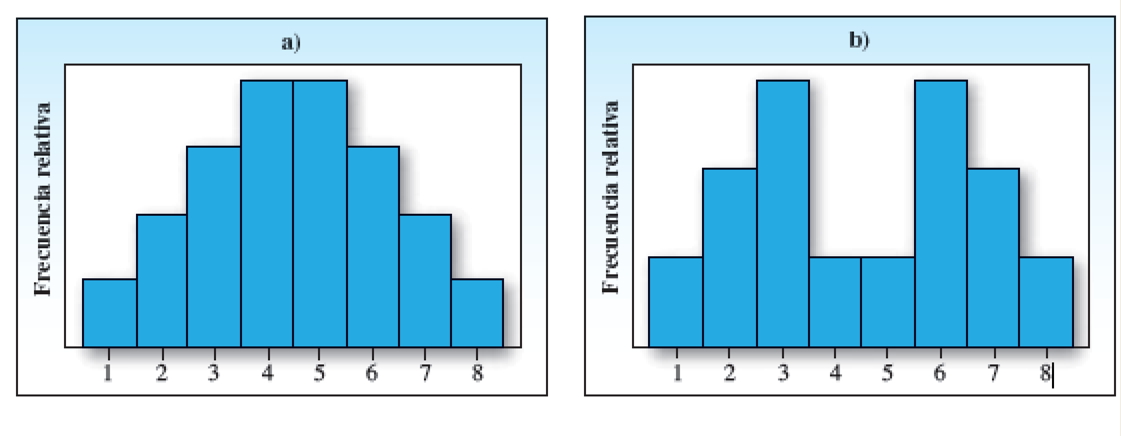
- La **variabilidad** o **dispersión** es una muy importante característica de datos.
- Las **medidas de variabilidad** pueden ayudarle a crear una imagen mental de la dispersión de los datos.



MEDIDAS DE DISPERSIÓN

El **rango, R**, de un conjunto de n mediciones se define como la diferencia entre la medición más grande y la más pequeña.

- Muy fácil de calcular. $\longrightarrow$ $R = x_{MAX} - x_{MIN}$
- Muy fácil de interpretar.
- Es una buena medida de dispersión para muestras pequeñas



MEDIDAS DE DISPERSIÓN

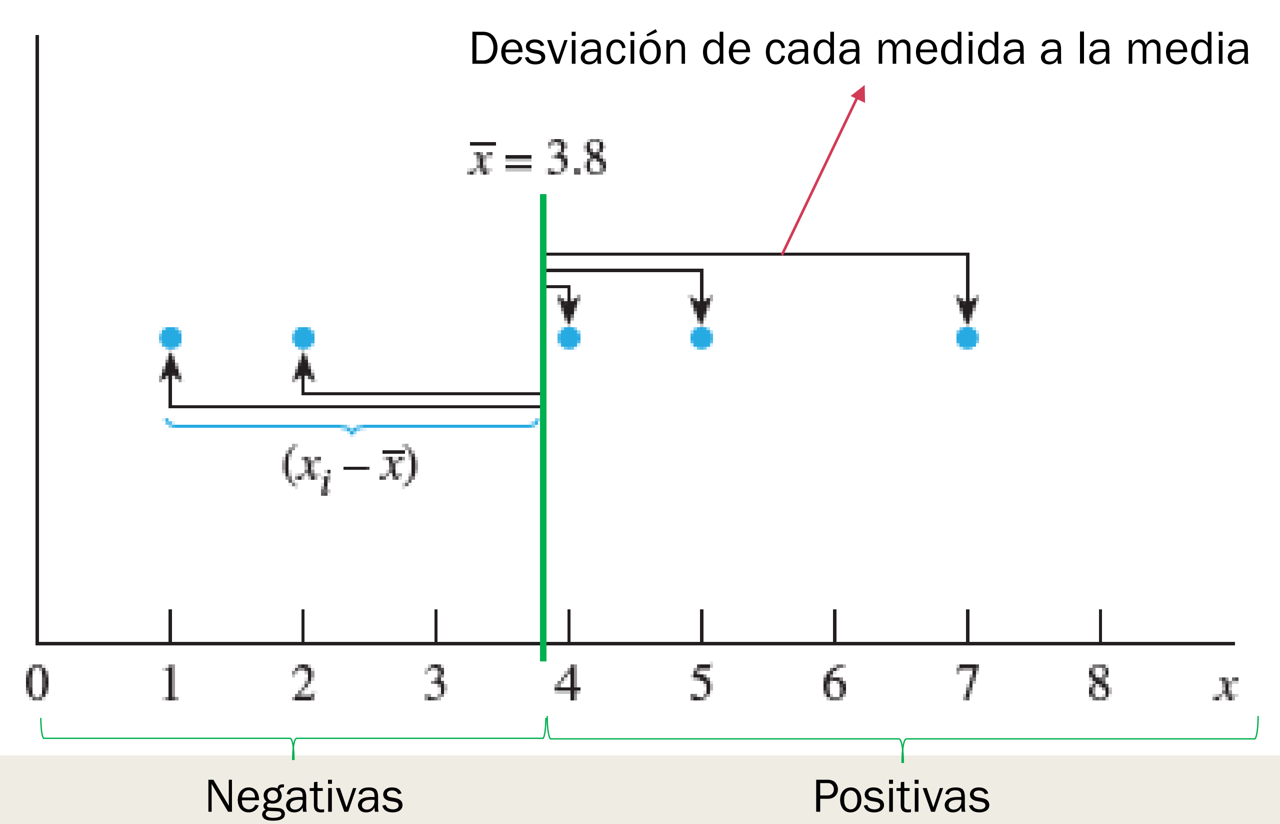
EJEMPLO:

Dadas las siguientes mediciones muestrales 5, 7, 1, 2, 4:

$$\bar{x} = \frac{5+7+1+2+4}{5}$$

$$\bar{x} = \frac{19}{5} = 3{,}8$$

$$R = 7 - 1 = 6$$



MEDIDAS DE DISPERSIÓN

La **varianza** de una población de n mediciones es el promedio de los cuadrados de las desviaciones de las mediciones alrededor de su media x̄.

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N} \qquad s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

- Grande $\longrightarrow$ Muestras muy dispersas.
- Chica $\longrightarrow$ Muestras poco dispersas.
- Se hace al cuadrado, dado que al haber desviaciones positivas y negativas, el numerador se anularía.

MEDIDAS DE DISPERSIÓN

EJEMPLO: Dadas las siguientes mediciones muestrales $\boxed{5, 7, 1, 2, 4:}$ → N = 5

X	$(x_i - \bar{x})$	$(xi - \bar{x})^2$
5	1,2	1,44
7	3,2	10,24
1	-2,8	7,84
2	-1,8	3,24
4	-0,2	0,04
19	**0**	**22,80**

$$\bar{x} = \frac{19}{5} = 3{,}8$$

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{N}$$

$$s^2 = \frac{22{,}80}{4}$$

$$s^2 = 5{,}7$$

MEDIDAS DE DISPERSIÓN

La **desvío estándar** raíz cuadrada positiva de la varianza

$$s = \sqrt{s} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N}}$$

EJEMPLO: Dadas las siguientes mediciones muestrales 5, 7, 1, 2, 4:

$$\bar{x} = 3{,}8 \qquad s^2 = 5{,}7 \qquad s = \sqrt{5{,}7} = 2{,}39$$

- En promedio las mediciones dieron 3,8 ± 2,39.

EJERCITACIÓN

11. La tabla siguiente muestra los nombres de los 42 presidentes de Estados Unidos, junto con el número de sus hijos.

a. Construir un histograma de frecuencia relativa para describir los datos. ¿Cómo describiría usted la forma de esta distribución?

b. Calcular la media y la desviación estándar para el conjunto de datos.

Washington	0	Van Buren	4	Buchanan	0
Adams	5	W.H. Harrison	10	Lincoln	4
Jefferson	6	Tyler*	15	A. Johnson	5
Madison	0	Polk	0	Grant	4
Monroe	2	Taylor	6	Hayes	8
J.Q. Adams	4	Fillmore*	2	Garfield	7
Jackson	0	Pierce	3	Arthur	3
Cleveland	5	Coolidge	2	Nixon	2
B. Harrison*	3	Hoover	2	Ford	4
McKinley	2	F.D. Roosevelt	6	Carter	4
T. Roosevelt*	6	Truman	1	Reagan*	4
Taft	3	Eisenhower	2	G.H.W. Bush	6
Wilson*	3	Kennedy	3	Clinton	1
Harding	0	L.B. Johnson	2	G.W. Bush	2

* Casado dos veces

Fuente: Time Almanac 2007

EJERCITACIÓN

12. Calcular la media y desviación estándar para estas calificaciones.

Calificación de examen	Número de estudiantes
7	1
6	4
5	4
4	4
3	4

MEDIDAS DE POSICIÓN

Los **cuartiles** son medidas numéricas que surgen de separar la cantidad de mediciones en 4 partes iguales.

- Q_1 → Primer cuartil → El 25 % de las mediciones
- Q_2 → Segundo cuartil → El 50 % de las mediciones → = Me
- Q_3 → Tercer cuartil → El 75 % de las mediciones
- Q_4 → Cuarto cuartil → El 100 % de las mediciones

MEDIDAS DE POSICIÓN

- Calculo de posición del cuartil

$$Pos\ Q_k = \frac{k * N}{4}$$

Número de cuartil

1
2
3
4

- Calculo de cuartil
  - Datos si agrupar → Valor de medición cuya frecuencia acumulada toma la posición de ese cuartil
  - Datos agrupados (Con intervalos) → $Q_k = L_i + \frac{PosQ_k - F_{i-1}}{f_i} * a_i$

MEDIDAS DE POSICIÓN

Los **deciles** son medidas numéricas que surgen de separar la cantidad de mediciones en 10 partes iguales.

- D_1 → Primer decil → El 10 % de las mediciones
- D_2 → Segundo decil → El 20 % de las mediciones

⋮

- D_9 → Noveno decil → El 90 % de las mediciones
- D_{10} → Decimo decil → El 100 % de las mediciones

MEDIDAS DE POSICIÓN

- Calculo de posición del decil

$$Pos\ D_k = \frac{k * N}{10}$$

Número de decil

1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 - 10

- Calculo de decil
  - Datos si agrupar → Valor de medición cuya frecuencia acumulada toma la posición de ese decil
  - Datos agrupados (Con intervalos) → $Q_k = L_i + \frac{PosD_k - F_{i-1}}{f_i} * a_i$

MEDIDAS DE POSICIÓN

Los **percentiles** son medidas numéricas que surgen de separar la cantidad de mediciones en 100 partes iguales.

- P_1 → Percentil 1 → El 1 % de las mediciones
- P_2 → Percentil 2 → El 2 % de las mediciones
- P_{49} → Percentil 49 → El 49 % de las mediciones
- P_{84} → Percentil 84 → El 84 % de las mediciones

MEDIDAS DE POSICIÓN

- Calculo de posición del percentil

Número de percentil

1 al 100

$$Pos\ P_k = \frac{k * N}{100}$$

- Calculo de percentil
  - Datos si agrupar → Valor de medición cuya frecuencia acumulada toma la posición de ese percentil
  - Datos agrupados (Con intervalos) → $Q_k = L_i + \frac{PosP_k - F_{i-1}}{f_i} * a_i$

MEDIDAS DE POSICIÓN

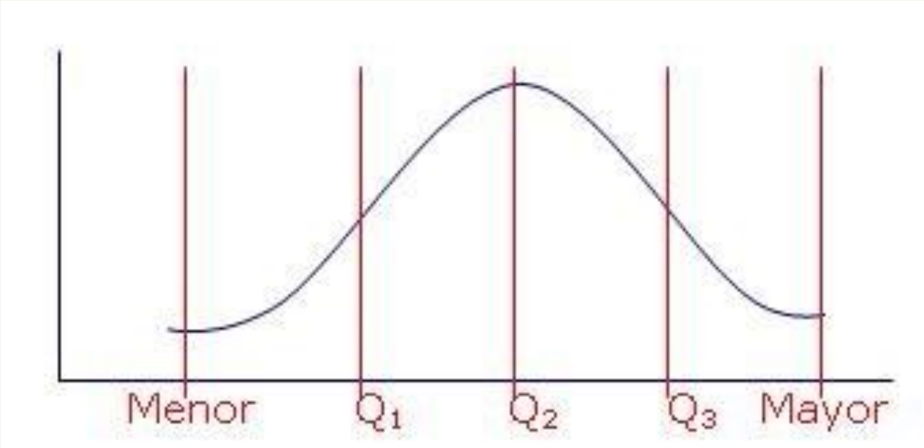
- Resumen

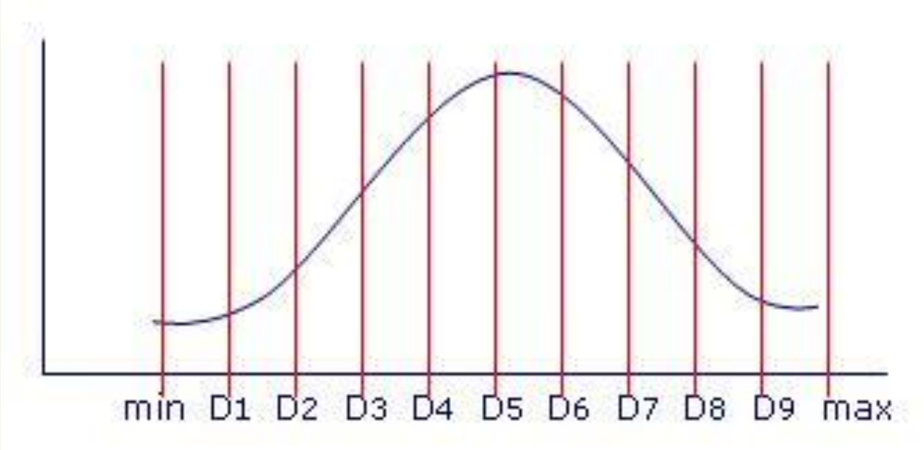
$$Q_1 = P_{25}$$

$$Q_2 = Me = D_5 = P_{50}$$

$$Q_3 = P_{75}$$

$$Q_4 = D_{10} = P_{100}$$





EJERCITACIÓN

14. De la siguiente distribución calcular:

    a. Media, mediana y moda.

    b. Varianza y desvío estándar.

    c. Los 4 cuartiles y el percentil 82.

X	Frecuencia absoluta	Frecuencia acumulada
2	6	6
3	15	21
4	10	31
5	9	40
TOTALES	40	

EJERCITACIÓN

15. De la siguiente distribución, calcular:
   a. Media, mediana y moda.
   b. Varianza y desviación estándar.
   c. Percentil 59, decil 4 y cuartil 3.

X	Fr. abs	Fr acum
10 – 12	5	5
12 – 14	11	16
14 – 16	19	35
16 – 18	21	56
18 - 20	4	60

EJERCITACIÓN

16. La distribución de las estaturas en centímetros de los alumnos de un centro, expresados en porcentajes, es la siguiente:
    a. ¿Entre qué estaturas se encuentra la quinta parte de las estaturas centrales?

X = estatura en cm	Fr relativa
145 – 149	0,3 %
150 – 154	1,6 %
155 – 159	9,4 %
160 – 164	20,5 %
165 – 169	31,5 %
170 – 174	22,5 %
175 – 179	10,7 %
180 - 184	3,5 %

EJERCITACIÓN

19. La altura en cm de los niños de 12 años, examinados durante la última semana en la unidad de crecimiento del centro hospitalario “Crecebien”, viene representada en la siguiente tabla, sabiendo que la altura media de los mismos es 147,75 cm, calcular:

a. La frecuencia A del tercer intervalo.
b. La simetría de la distribución a partir de la comparación de media, mediana y moda.
c. El percentil correspondiente a un niño que mide 1,43 m.

Estatura	Niños
155-159	4
150-154	13
145-149	12
140-144	A
135-139	2
130-134	1