BIOESTADÍSTICA BÁSICA

MEDIDAS DE:
POSICIÓN /TENDENCIA CENTRAL/DISPERSIÓN

CLASE 2

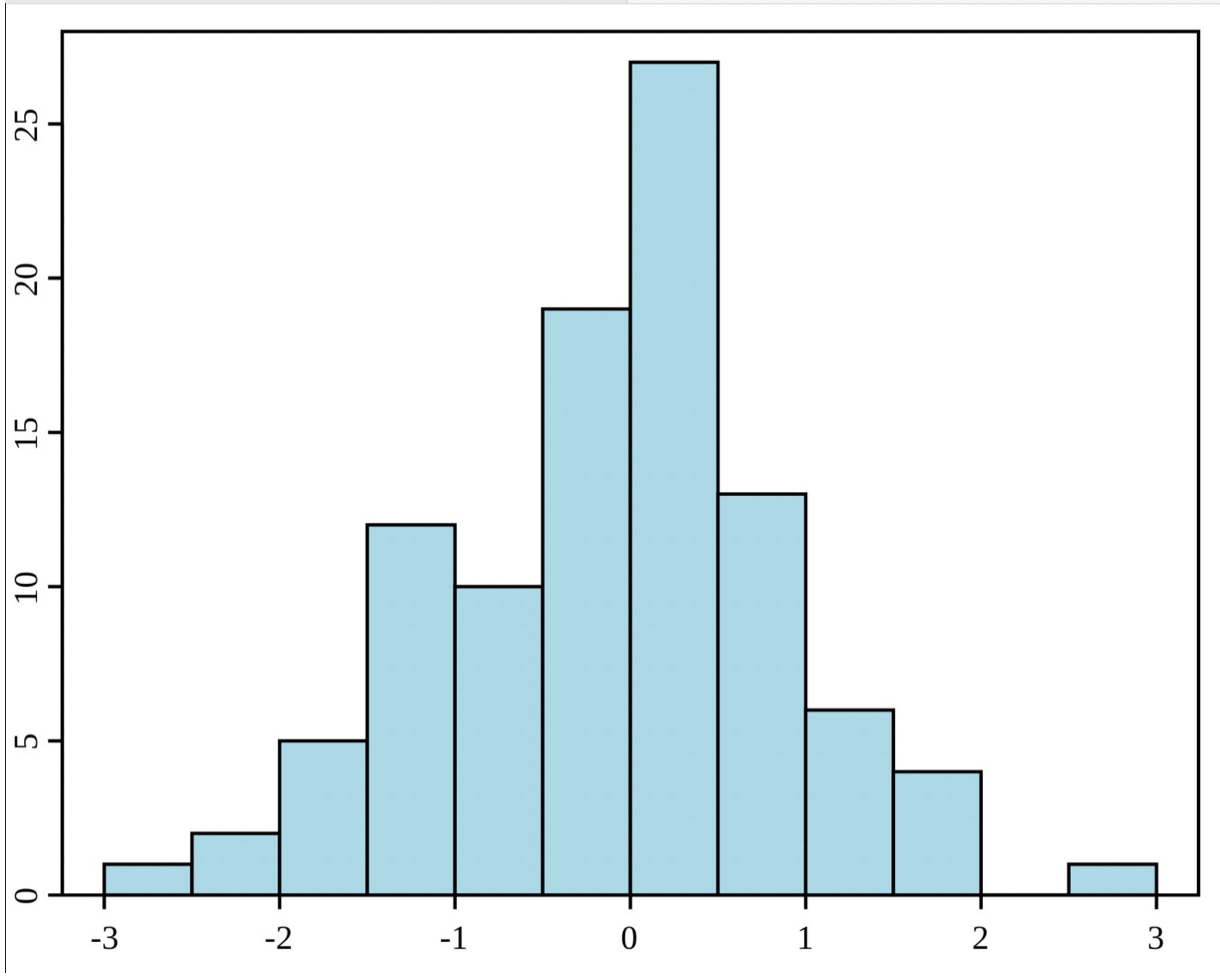
ESTADÍSTICA DESCRIPTIVA

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.

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.

VARIABLES CUANTITATIVAS



DATOS AGRUPADOS

MEDIDAS DE POSICIÓN

PERCENTILES, CUARTILES QUINTILES, DECILES

Medidas de posición NO central

Percentiles son las medidas más utilizadas para propósitos de ubicación o clasificación de las personas en características tales como peso, estatura, etc.

A TENER EN CUENTA:

La preferencia para elegir Cuartiles, Deciles o Percentiles está condicionada al grado de detalle con que se contextualiza la posición de una observación en el conjunto de datos referenciados

Pero, también la elección entre las distintas medidas de posición no central está asociada a la heterogeneidad de las distribuciones.

Mientras más dispersión presenten los datos de una variable, se aconseja incrementar las unidades divisionales

- Recién nacidos - peso demasiado bajo. ¿Qué peso se considera “demasiado bajo”?
- Consultas de emergencia en hospitales según causa.
- Fecundidad adolescente según países.

PERCENTILES:

Cada uno de los 99 valores que dividen a la distribución de datos en 100 partes iguales.

P_k

$$\frac{nk}{100}$$

DONDE:

$$P_k = L_{k-1} + \frac{\frac{nk}{100} - N_{k-1}}{n_k} A$$

L_{k-1} es el limite inferior del intervalo h (cuya frecuencia acumulada es la primera mayor o igual a $\frac{nk}{100}$)

nk es la frecuencia absoluta del intervalo h

N_{k-1} es la frecuencia acumulada hasta L_{k-1}

A es la longitud del intervalo h

EJEMPLO

Calcular el percentil de orden 86 de los ingresos mensuales de algunos médicos en EEUU

INGRESO MENSUAL	NÚMERO DE PERSONAS (ni)	FECUENCIA ACUMULADA (NI)
2400-3000	3	3
3000-4200	20	23
4200-5400	35	58
5400-7250	25	83
7250-9000	15	98
9000-12000	2	100

$$\frac{nk}{100}$$

$$\frac{nk}{100} = \frac{100 * 86}{100} = 86$$

INGRESO MENSUAL	NÚMERO DE PERSONAS (ni)	FECUENCIA ACUMULADA (NI)
2400-3000	3	3
3000-4200	20	23
4200-5400	35	58
5400-7250	25	83
7250-9000	15	98
9000-12000	2	100

L_{k-1} es el limite inferior del intervalo h (cuya frecuencia acumulada es la primera mayor o igual a $\frac{nk}{100}$)

P_{86} es (7250, 9000)
L_{k-1} = **7250**

INGRESO MENSUAL	NÚMERO DE PERSONAS (ni)	FECUENCIA ACUMULADA (NI)
2400-3000	3	3
3000-4200	20	23
4200-5400	35	58
5400-7250	25	83
7250-9000	15	98
9000-12000	2	100

N_{k-1} es la frecuencia acumulada **hasta** L_{k-1}

$$N_{k-1} = 83$$

nk es la frecuencia absoluta del intervalo h

$$nk = 15$$

A es la longitud del intervalo h

$$A=9000-7250 = \mathbf{1750}$$

Con esos datos, reemplazo:

$$P_k = L_{k-1} + \frac{\frac{nk}{100} - N_{k-1}}{n_k} A$$

$$P_{86} = 7250 + \frac{86 - 83}{15} 1750$$

$P86 =$ 7600 – El 86% de los sueldos anuales de los médicos están ubicados en 7.600$ o menos

CUARTILES:

$$\boxed{\frac{nk}{4}}$$

Son 3 valores o elementos que dividen a la distribución o conjunto de datos en 4 partes iguales, es decir, cada una engloba el 25% de los mismos.

El cuartil inferior (Q1)

Deja a su izquierda los datos (es el menor valor pero, a su vez, el mayor valor de una cuarta parte de los datos) y se cumple que Q1 = P_{25}

El cuartil medio (Q2)

Deja a su izquierda el 50% de los datos y coincide exactamente con la mediana y se cumple Q2= P50

El cuartil superior (Q3)

Deja a su izquierda el 75% de los datos (mayor que tres cuartas partes de los datos) y se cumple Q3 = P75

INTERPRETAR:

Q1 = El 25% de los datos es <= al valor

Q2 = El 50% de los datos es <= al valor

Q3 = El 75% de los datos es <= al valor

Datos agrupados

$Q_1 = \frac{1}{4}N$

$Q_2 = \frac{2}{4}N$

$Q_3 = \frac{3}{4}N$

N de obs de las posiciones de los cuartiles

$$Q_m = L_i + \frac{\frac{mN}{4} - N_{ai-1}}{n_m} \times a$$

Donde:

m = cuartil que se desea conocer.
L_{inf} = límite inferior del intervalo donde se encuentra el cuartil.
N = Número total de observaciones [frecuencia absoluta total].
N_{ai-1} = Frecuencia absoluta acumulada del intervalo anterior a aquel en que se encuentra el cuartil.
n_m= frecuencia absoluta total del intervalo cuartil.
a = amplitud del intervalo.

EJEMPLO

La distribución de edades de personas con Covid19 en un hospital público de la cuidad de Quito.

Edad	Frecuencia absoluta	Frecuencia absoluta acumulada N_{ai}
30-35	3	3
36-40	3	6
41-45	6	12
46-50	14	26
51-55	9	35
56-60	8	43
61-65	7	50
66-70	6	56
71-75	4	60

$$Q_1 = \frac{1}{4}N$$

$$Q_2 = \frac{2}{4}N$$

$$Q_3 = \frac{3}{4}N$$

Edad	Frecuencia absoluta	Frecuencia absoluta acumulada N_{ai}
30-35	3	3
36-40	3	6
41-45	6	12
46-50	14	26
51-55	9	35
56-60	8	43
61-65	7	50
66-70	6	56
71-75	4	60

Q1 = 15

Q2 = 30

Q3 = 45

$$Q_m = L_i + \frac{\frac{mN}{4} - N_{ai-1}}{n_m} \times a$$

Donde:

m = cuartil que se desea conocer.

L_{inf} = límite inferior del intervalo donde se encuentra el cuartil.

N = Número total de observaciones [frecuencia absoluta total].

N_{ai-1} = Frecuencia absoluta acumulada del intervalo anterior a aquel en que se encuentra el cuartil.

n_m= frecuencia absoluta total del intervalo cuartil.

a = amplitud del intervalo.

Q1 =

Q2=

Q3 =

$$Q_1 = 46 + \frac{15 - 12}{14} \times 4 = 46{,}86 \text{ años}$$

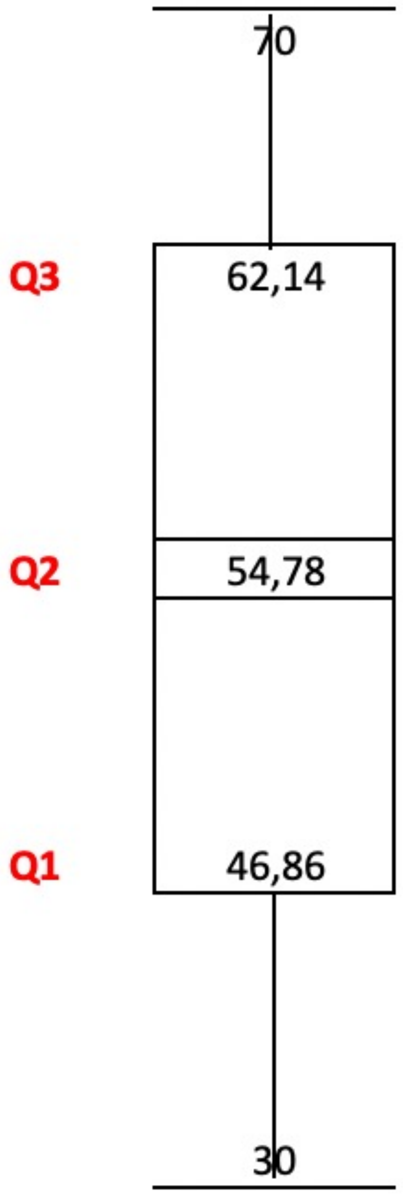
$$Q_2 = 51 + \frac{30 - 26}{9} \times 4 = 52{,}78 \text{ años}$$

$$Q_3 = 61 + \frac{45 - 43}{7} \times 4 = 62{,}14 \text{ años}$$

25% de los pacientes con Covid19 en un hospital público de la cuidad de Quito, tiene menos o igual que 46,86 años.

50% de los pacientes con Covid19 en un hospital público de la cuidad de Quito, tiene 52,78 años.

75% de los pacientes con Covid19 en un hospital público de la cuidad de Quito, tiene más o igual a 62,14 años.


70
Q3
62,14
Q2
54,78
Q1
46,86
30

DECILES:

Son valores que dividen los datos ordenados en diez partes iguales (9 divisiones). Datos clasificados en orden ascendente

Li	D1	D2	D3	D4	D5	D6	D7	D8	D9	Ls

Los deciles dan los valores correspondientes al 10 %, al 20 %... y al 90 % de los datos

El decil 5 es la mediana o el Q2

$$\frac{nk}{10}$$

$$D_k = L_i + \left(\frac{\frac{kN}{10} - F_{i-1}}{f_i}\right) \cdot A$$

N es la cantidad de datos de la muestra.

K es el número del decil.

F_{i-1} es la ***frecuencia absoluta acumulada ANTERIOR*** al intervalo de trabajo.

f_i es la ***frecuencia absoluta*** del intervalo de trabajo.

L_i es el ***límite inferior*** del intervalo de trabajo.

A es la **amplitud** del intervalo de trabajo.

EJEMPLO

La distribución de edades de personas con Zika, en un hospital público de Esmeraldas. Calcular el decil 4

Edad	NÚMERO DE PERSONAS (ni)	FECUENCIA ACUMULADA (NI)
10-19	5	5
20-29	11	16
30-39	8	24
40-49	5	29
50-59	8	37
60-69	6	43
70-79	7	50

$$\boxed{\frac{50*4}{10} = 20}$$

Medidas de posición	Cálculo	
	Distribución de datos no agrupados	Distribución de datos agrupados
Cuartiles [Q_m]	Localización	$Q_m = L_i + \frac{mN/4 - N_{ai-1}}{n_m} \times a$
Deciles [D_m]		$D_m = L_i + \frac{mN/10 - N_{ai-1}}{n_m} \times a$
Centiles o Percentiles [C_m] o [P_m]		$P_m = L_i + \frac{mN/100 - N_{ai-1}}{n_m} \times a$

MEDIDAS DE TENDENCIA CENTRAL

NORMALMENTE SE UBICAN EN EL CENTRO DE LA DISTRIBUCIÓN

MEDIA
MEDIANA
MODA

μ
$\bar{X}$

Datos agrupados por intervalos

LIC	LSC	Punto medio	Frec. Absoluta
l1	s1	x1	n1
l2	s2	x2	n2
.	.	.	.
.	.	.	.
lk	sk	xk	nk

Punto medio $xi = \frac{li+si}{2}$

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

EJEMPLO

El MSP realizó una tabla de frecuencias del presupuesto destinado a 10 pequeños proyectos de salud. Calcular el promedio

Min	Max	Punto medio	Frec. Absoluta
0	100		12
100	200		28
200	300		46
300	400		71
400	500		186
500	600		224
600	700		209
700	800		122
800	900		53
900	1000		19

RESULTADO

$$\bar{x} = \quad 555.15$$

El promedio del presupuesto destinado a proyectos pequeños de salud es de 555,15 USD

MEDIANA: La mediana es el valor medio cuando

$$M_e = L_i + \frac{\frac{n}{2} - F_{i-1}}{f_i} . A_i$$

Donde:

- L_i: límite inferior del intervalo en el cual se encuentra la mediana.
- n: número de datos del estudio. Es la sumatoria de las frecuencias absolutas.
- F_{i-1}: frecuencia acumulada del intervalo anterior al que se encuentra la mediana.
- A_i: amplitud del intervalo en el que se encuentra la mediana.
- f_i: frecuencia absoluta del intervalo en el que se encuentra la mediana.

RECORDAR

Si los datos están agrupados en una tabla de distribución de frecuencias, se debe establecer el intervalo donde se encuentra la mediana.

Para ello se determina el intervalo cuya frencuencia acumulada sea mayor o igual a n/2.

Presupuesto anual	Número de programas de salud en un Municipio	Frecuencia acumulada
2400-3000	3	3
3000-4200	20	23
4200-5400	35	58
5400-7250	25	83
7250-9000	15	98
9000-12000	2	100

MODA: Es aquel valor que tiene la mayor frecuencia absoluta

$$M_o = L_i + \frac{f_i - f_{i-1}}{f_i - f_{i-1} + f_i - f_{i+1}} . A_i$$

Donde:

- L_i: límite inferior del intervalo en el cual se encuentra la moda.
- f_{i-1}: frecuencia absoluta del intervalo anterior en el que se encuentra la moda.
- f_i: frecuencia absoluta del intervalo en el que se encuentra la moda.
- f_{i+1}: frecuencia absoluta del intervalo siguiente en el que se encuentra la moda.
- A_i: amplitud del intervalo en el que se encuentra la moda.

	Intervalos	Marca de clase x	Frecuencia absoluta f	Frecuencia acumulada F
1	[0 - 4)	2	3	3
2	[4 - 8)	6	5	8
3	[8 - 12)	10	6	14
4	[12 - 16)	14	4	18
5	[16 - 20)	18	3	21
	Total		21	

MEDIDAS DE TENDENCIA DISPERSIÓN

MIDE EL GRADO DE DISPERSIÓN DE LOS DATOS COMPARADOS CON EL CENTRO

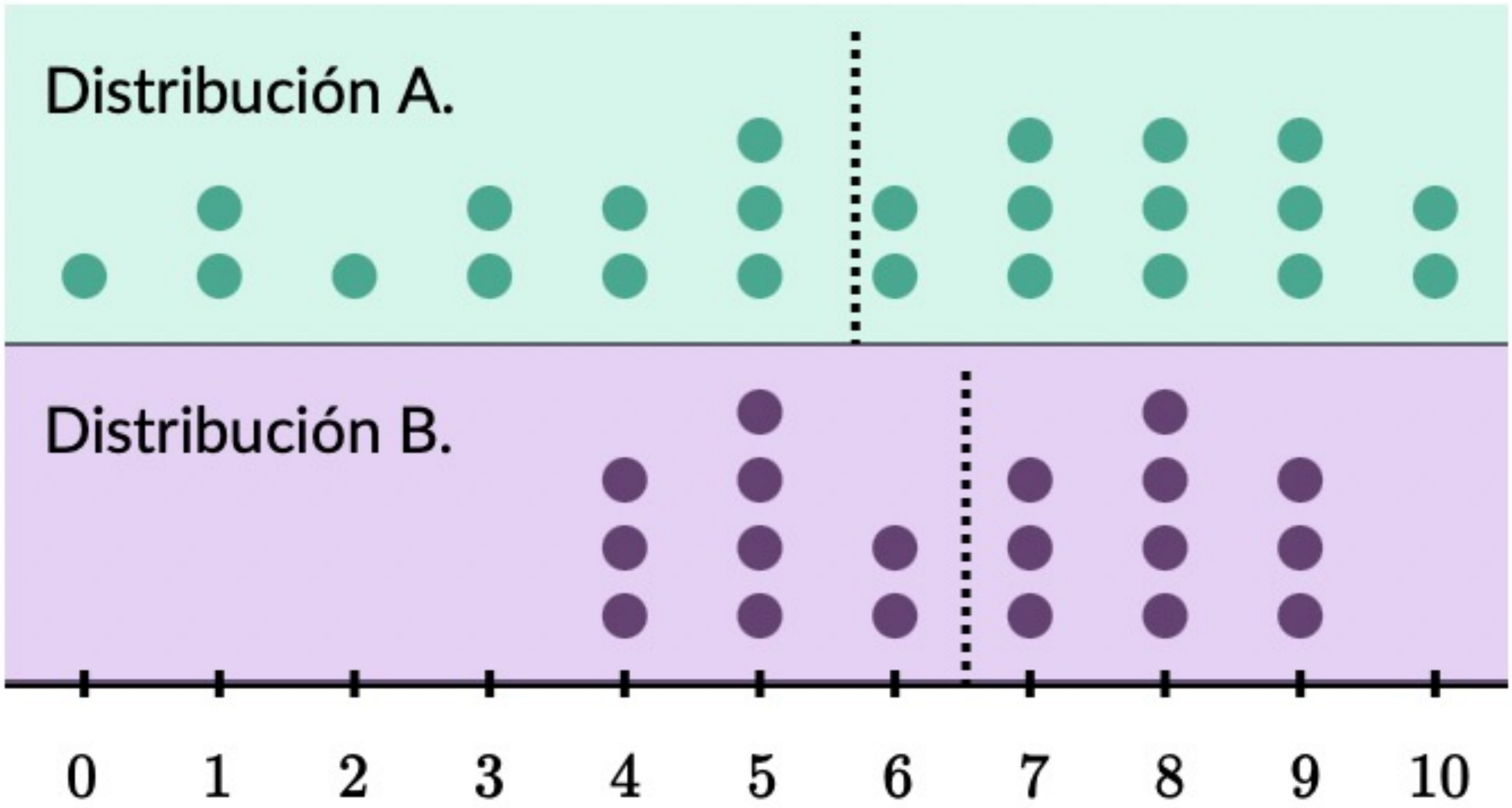
UNA MEDIDA DE VARIABILIDAD

MIENTRAS QUE UNA MEDIDA DE TENDENCIA CENTRAL DESCRIBE EL VALOR TÍPICO, LAS MEDIDAS DE VARIABILIDAD DEFINEN CUÁN LEJOS TIENDEN A CAER LOS PUNTOS DE DATOS DEL CENTRO.

MEASURES OF DISPERSION

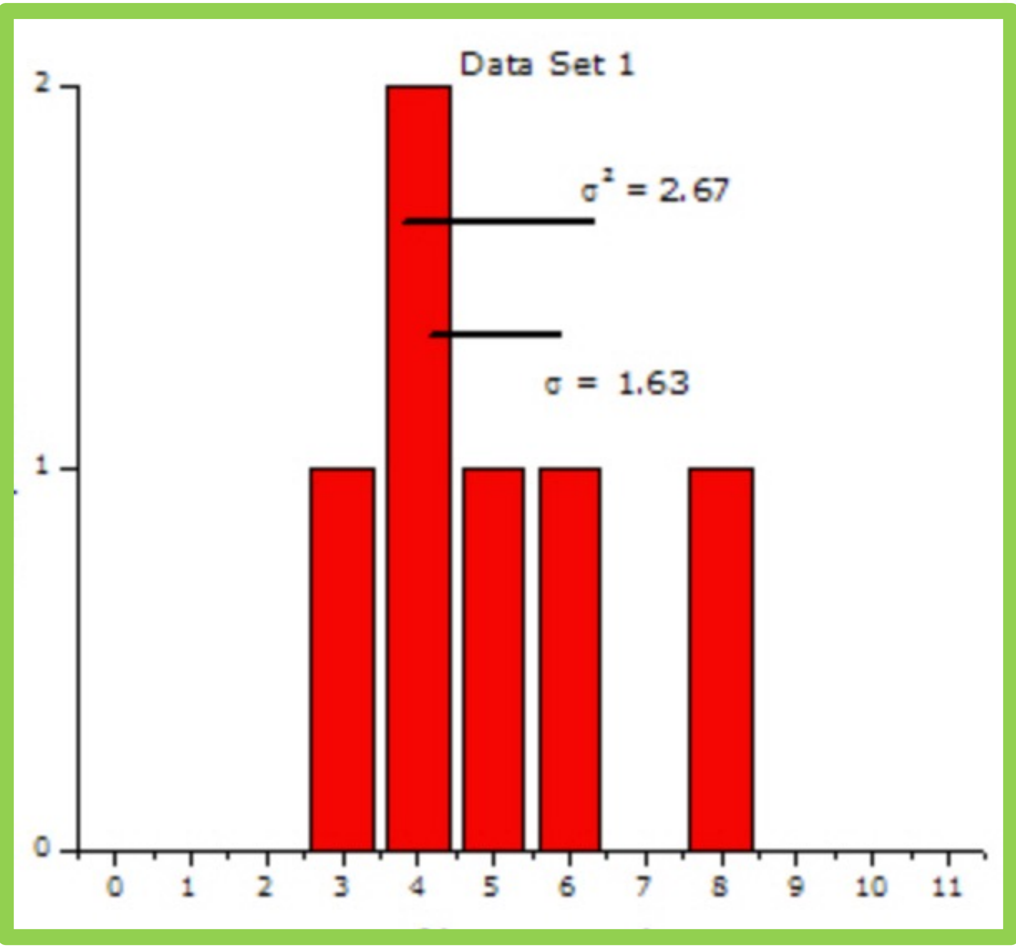
- **Absolute Measures**
  - **Distance Deviation Measures**
    - Range
    - Percentile
    - Quartile Deviation
    - Semi-interquartile Deviation
  - **Average Deviation Measures**
    - Variance
    - Mean Deviation
    - Standard Deviation
- **Relative Measures**
  - Coefficient of Quartile Deviation
  - Coefficient of Quartile Variation
  - Coefficient of Mean Deviation

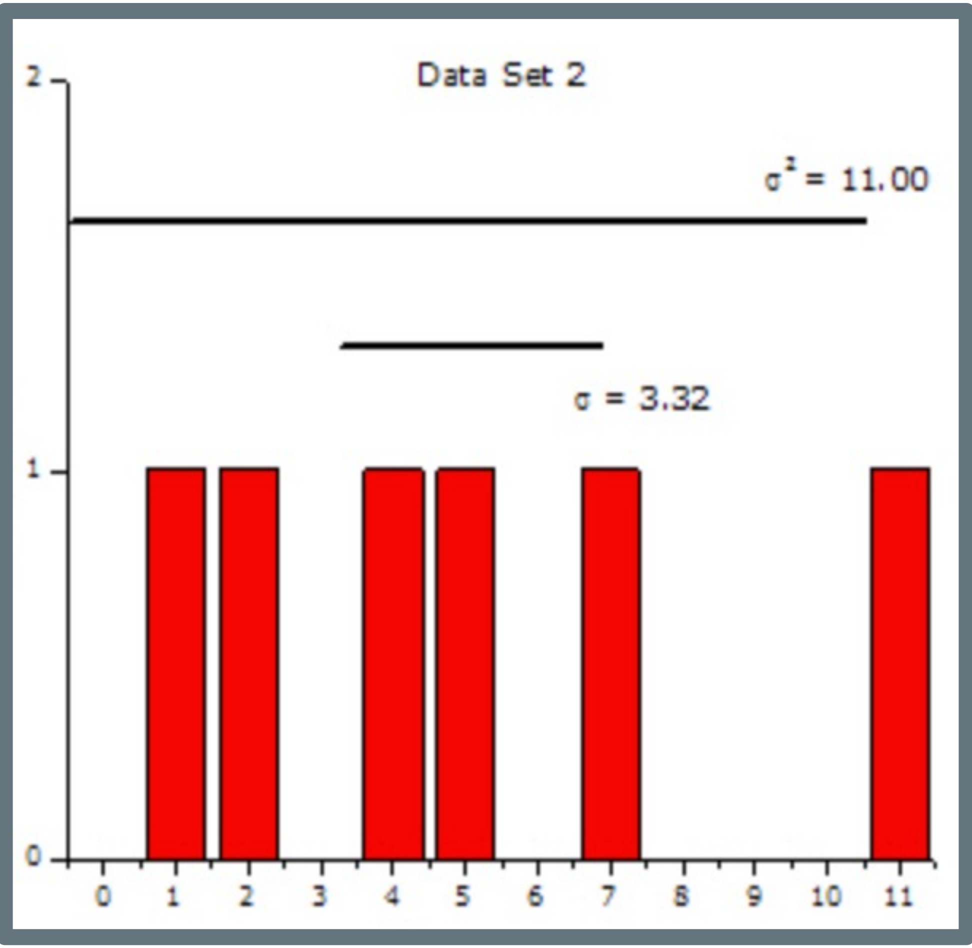
Una baja dispersión indica que los puntos de datos tienden a agruparse estrechamente alrededor del centro. La alta dispersión significa que tienden a caer más lejos.



Distribución A.
Distribución B.
0
1
2
3
4
5
6
7
8
9
10

Conjunto de datos	*N*	ΣX	ΣX2	*μ*	*μ2*	*σ2*	*σ*
1	6	30	166	5	25	2.67	1.63
2	6	30	216	5	25	11.00	3.32





DESVIACIÓN ESTÁNDAR: Raíz cuadrada de la varianza, si el valor de la misma es grande, la dispersión de los valores también lo será.

$\boxed{\sigma =}$ → La desviación estándar de **población**

$\boxed{s =}$ → La desviación estándar de **la muestra**

Varianza y desviación estándar para datos agrupados

	Varianza	Desviación estándar	Media	Número de elementos
Población	$\sigma^2 = \dfrac{\sum_{i=1}^{k} f_i(x_i - \mu)^2}{N}$	$\sigma = \sqrt{\sigma^2}$	$\mu = \dfrac{\sum_{i=1}^{k} x_i \cdot f_i}{N}$	$N = \displaystyle\sum_{i=1}^{k} f_i$
Muestra	$s^2 = \dfrac{\sum_{i=1}^{k} f_i(x_i - \bar{x})^2}{n-1}$	$s = \sqrt{s}$	$\bar{x} = \dfrac{\sum_{i=1}^{k} x_i \cdot f_i}{n}$	$n = \displaystyle\sum_{i=1}^{k} f_i$

EJEMPLO

Minutos de traslado en automóvil, al centro de salud más cercano, en mujeres de la Joya de los Sachas.

Minutos	Número de Mujeres (ni)
0-10	2
11-20	3
21-30	3
31-40	7
41-50	5

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

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot f_i}{n}$$

PUNTO MEDIO

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

Minutos	Número de Mujeres (ni)	Punto medio Xi	$\bar{X}$	$Xi - \bar{X}$	$(Xi - \bar{X})^2$	$fi * (Xi - \bar{X})^2$
0-10	2					
11-20	3					
21-30	3					
31-40	7					
41-50	5					

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

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot f_i}{n}$$

Minutos	Número de Mujeres (ni)	Punto medio Xi	$\bar{X}$	Xi - $\bar{X}$	(Xi – $\bar{X}$)^2	fi * (Xi – $\bar{X}$)^2
0-10	2	5	10	-25,45	647,7025	1295,405
11-20	3	15,5	46,5	-14,95	223,5025	670,5075
21-30	3	25,5	76,5	-4,95	24,5025	73,5075
31-40	7	35,5	248,5	5,05	25,5025	178,5175
41-50	5	45,5	227,5	15,05	226,5025	1132,5125
			609			3350,45
			30,45			S^2: 176,34
						S: 13,28

INTERPETAR EN RELACIÓN A LA MEDIA

La sd, toma las unidades de medida del ejercicio; cm, dólares, años, kg.

La desviación estándar de los minutos de traslado, con respecto a su promedio es 13,28 minutos

¿QUÉ TAN DISPERSOS ESTÁN LOS DATOS?

Coeficiente de variación: $\frac{s}{\bar{x}}$

$$\frac{13,28}{30,45} X\ 100 = 43,61\%$$