

NOMBRE DE LA ASIGNATURA:	Estadística Descriptiva	TUTOR:	Deivis Galván Cabrera
---------------------------------	-------------------------	---------------	-----------------------

Nombre del estudiante: _____ Fecha: _____

- Al comenzar el curso se pasó una encuesta a los alumnos del primer curso de un colegio, preguntándoles, entre otras cuestiones, por el número de hermanos que tenían, obteniéndose los siguientes resultados.

3, 3, 2, 2, 8, 5, 2, 4, 3, 1, 4, 5, 3, 3, 3, 3, 3, 2, 5

1, 3, 3, 2, 2, 4, 3, 3, 2, 2, 4, 4, 3, 6, 3, 3, 2, 2, 4

3, 4, 3, 2, 2, 4, 4, 3, 3, 4, 2, 5, 4, 1, 2, 8, 2, 3, 3, 4

- Representa este conjunto de datos con un diagrama de barras
 - Calcula las medidas de tendencia central para los datos (Media, Mediana y Moda)
 - Calcula las medidas de dispersión central para los datos (Rango, Varianza, Desviación típica y Coeficiente de variación de Pearson)
- Para realizar un determinado experimento se ha medido la anchura interorbital, en mm. De una muestra de 40 palomas, obteniéndose los siguientes datos:

12'2, 12'9, 11'8, 11'9, 11'6, 11'1, 12'3, 12'2, 11'8, 11'8

10'7, 11'5, 11'3, 11'2, 11'6, 11'9, 13'3, 11'2, 10'5, 11'1

12'1, 11'9, 10'4, 10'7, 10'8, 11'0, 11'9, 10'2, 10'9, 11'6

10'8, 11'6, 10'4, 10'7, 12'0, 12'4, 11'7, 11'8, 11'3, 11'1

Se pide:

- Construya una distribución de frecuencias y calcule la media, desviación típica y coeficiente de variación.
- Agrupe los datos en intervalos con la amplitud más adecuada, calculando de nuevo los parámetros anteriores y comparándolos con los resultados obtenidos a partir de los datos no agrupados. Dibuje el histograma.

En lo que sigue trabaje con la distribución por intervalos.

- ¿En qué intervalo de centro de la media se encuentra, al menos, el 75 % de la distribución?
- Calcule la mediana y la moda.
- Obtenga el intervalo donde se encuentra el 40 % central de la distribución.

3. Suponga que los salarios pagados a los empleados de una compañía, son como se representa a continuación:

CARGOS	NÚMERO	SALARIO MENSUAL
Directores	2	1.930.000
Supervisores	4	1.510.000
Economistas	6	1.370.000
Contadores	4	1.350.000
Auxiliares	26	646.000
Obreros	110	590.000

Se pide:

- a) Calcular la media, mediana y moda.
b) En su opinión. ¿Cuál de estos promedios considera más representativo?
4. Durante diez días se observó en un almacén de autos, el número de estos vendidos por su empleado estrella.

DÍA:	1	2	3	4	5	6	7	8	9	10
AUTOS:	3	4	2	1	3	2	4	6	5	4

Cada auto vale \$18.500.000 y al empleado se le paga un sueldo de \$270.000 básico mensual más el 0,5% del valor de cada auto vendido. ¿Cuál será el sueldo promedio para el vendedor estrella en este lapso de 10 días?

5. En una distribución unidimensional se sabe, que la media aritmética es igual a 3,04 y que la media cuadrática es igual a 4,26. Se pide calcular la varianza y la desviación típica de dicha distribución.
6. Con los siguientes datos:

28	42	64	32	68	25	34	40	51	62
36	56	76	40	70	28	38	46	58	64

Se pide:

- a) Trabajando con los datos originales calcular la desviación típica, desviación mediana y el coeficiente de variación.
b) Agrupar los datos en una tabla de frecuencias, teniendo en cuenta que la amplitud es constante (la variable es continua) y $m = 6$, calcular el coeficiente de variación.

Las medidas de dispersión nos informan sobre cuánto se alejan del centro los valores de la distribución.

Las medidas de dispersión son:

Rango o recorrido: El rango es la diferencia entre el mayor y el menor de los datos de una distribución estadística.

Desviación media: La desviación respecto a la media es la diferencia entre cada valor de la variable estadística y la media aritmética.

$$D_i = x - \bar{x}$$

La desviación media es la media aritmética de los valores absolutos de las desviaciones respecto a la media.

La desviación media se representa por D_x $D_{\bar{x}} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{N}$

$$D_x = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{N}$$

Ejemplo: Calcular la desviación media de la distribución:

9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

$$D_{\bar{x}} = \frac{|9-9| + |3-9| + |8-9| + |8-9| + |9-9| + |8-9| + |9-9| + |18-9|}{8} = 2.25$$

Desviación media para datos agrupados: Si los datos vienen agrupados en una tabla de frecuencias, la expresión de la desviación media es:

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}|f_1 + |x_2 - \bar{x}|f_2 + \dots + |x_n - \bar{x}|f_n}{N}$$

$$D_x = \frac{\sum_{i=1}^n |x_i - \bar{x}|f_i}{N}$$

Ejemplo: Calcular la **desviación media** de la distribución:

	x_i	f_i	$x_i \cdot f_i$	$ x - x $	$ x - x \cdot f_i$
[10, 15]	12.5	3	37.5	9.286	27.858
[15, 20]	17.5	5	87.5	4.286	21.43
[20, 25]	22.5	7	157.5	0.714	4.998
[25, 30]	27.5	4	110	5.714	22.856
[30, 35]	32.5	2	65	10.174	21.428
		21	457.5		98.57

$$\bar{x} = \frac{457.5}{21} = 21.786$$

$$D_{\bar{x}} = \frac{98.57}{21} = 4.69$$

Varianza: La varianza es la media aritmética del cuadrado de las desviaciones respecto a la media de una distribución estadística. La varianza se representa por σ^2 .

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{N}$$

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

Varianza para datos agrupados

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \dots + (x_n - \bar{x})^2 f_n}{N}$$

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

Para simplificar el cálculo de la varianza vamos a utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{N} - \bar{x}^2$$

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2}{N} - \bar{x}^2$$

Varianza para datos agrupados

$$\sigma^2 = \frac{x_1^2 f_1 + x_2^2 f_2 + \dots + x_n^2 f_n}{N} - \bar{x}^2$$

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2 f_i}{N} - \bar{x}^2$$

Ejercicios de varianza: Calcular la varianza de la distribución: 9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9 + 3 + 8 + 8 + 9 + 8 + 9 + 18}{8} = 9$$

$$\sigma^2 = \frac{(9 - 9)^2 + (3 - 9)^2 + (8 - 9)^2 + (8 - 9)^2 + (9 - 9)^2 + (8 - 9)^2 + (9 - 9)^2 + (18 - 9)^2}{8} = 15$$

Calcular la varianza de la distribución de la tabla:

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[10, 20]	15	1	15	225
[20, 30]	25	8	200	5000

$$\bar{x} = \frac{1820}{42} = 43.33$$

[30,40)	35	10	350	12 250
[40, 50)	45	9	405	18 225
[50, 60)	55	8	440	24 200
[60,70)	65	4	260	16 900
[70, 80)	75	2	150	11 250
		42	1 820	88 050

$$\sigma^2 = \frac{88050}{42} - 43.33^2 = 218.94$$

Propiedades de la varianza

1. 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.
4. Si tenemos varias distribuciones con la misma media y conocemos sus respectivas varianzas se puede calcular la varianza total.

Si todas las muestras tienen el mismo tamaño:

$$\sigma^2 = \frac{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}{n}$$

Si las muestras tienen distinto tamaño:

$$\sigma^2 = \frac{k_1 \cdot \sigma_1^2 + k_2 \cdot \sigma_2^2 + \dots + k_n \cdot \sigma_n^2}{k_1 + k_2 + \dots + k_n}$$

Observaciones sobre la varianza

1. La varianza, al igual que la media, es un índice muy sensible a las puntuaciones extremas.
2. En los casos que no se pueda hallar la media tampoco será posible hallar la varianza.
3. La varianza no viene expresada en las mismas unidades que los datos, ya que las desviaciones están elevadas al cuadrado.

Desviación típica: La desviación típica es la raíz cuadrada de la varianza. Es decir, la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación.

La desviación típica se representa por σ .

$$\sigma = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{N}}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{N}}$$

Desviación típica para datos agrupados

$$\sigma = \sqrt{\frac{(X_1 - \bar{x})^2 f_1 + (X_2 - \bar{x})^2 f_2 + \dots + (X_n - \bar{x})^2 f_n}{N}}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{x})^2 f_i}{N}}$$

Para simplificar el cálculo vamos a utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma = \sqrt{\frac{X_1^2 + X_2^2 + \dots + X_n^2}{N} - \bar{x}^2}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n X_i^2}{N} - \bar{x}^2}$$

Desviación típica para datos agrupados

$$\sigma = \sqrt{\frac{X_1^2 f_1 + X_2^2 f_2 + \dots + X_n^2 f_n}{N} - \bar{x}^2}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n X_i^2 f_i}{N} - \bar{x}^2}$$

Ejercicios de desviación típica: Calcular la desviación típica de la distribución:

9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9 + 3 + 8 + 8 + 9 + 8 + 9 + 18}{8} = 9$$

$$\sigma = \sqrt{\frac{(9-9)^2 + (3-9)^2 + (8-9)^2 + (8-9)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2 + (18-9)^2}{8}} = 3.87$$

Calcular la desviación típica de la distribución de la tabla:

	xi	fi	xi · fi	xi ² · fi
[10, 20)	15	1	15	225
[20, 30)	25	8	200	5000
[30,40)	35	10	350	12 250
[40, 50)	45	9	405	18 225
[50, 60)	55	8	440	24 200
[60,70)	65	4	260	16 900
[70, 80)	75	2	150	11 250
		42	1 820	88 050

$$\bar{x} = \frac{1820}{42} = 43.33$$

$$\sigma = \sqrt{\frac{88050}{42} - 43.33^2} = 14.797$$

Propiedades de la desviación típica

1. La desviación típica 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 desviación típica no varía.
3. Si todos los valores de la variable se multiplican por un número la desviación típica queda multiplicada por dicho número.

CAT – Caucaasia

4. Si tenemos varias distribuciones con la misma media y conocemos sus respectivas desviaciones típicas se puede calcular la desviación típica total.

Si todas las muestras tienen el mismo tamaño:

$$\sigma = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}{n}}$$

Si las muestras tienen distinto tamaño:

$$\sigma = \sqrt{\frac{k_1 \cdot \sigma_1^2 + k_2 \cdot \sigma_2^2 + \dots + k_n \cdot \sigma_n^2}{k_1 + k_2 + \dots + k_n}}$$

Observaciones sobre la desviación típica

1. La desviación típica, al igual que la media y la varianza, es un índice muy sensible a las puntuaciones extremas.
2. En los casos que no se pueda hallar la media tampoco será posible hallar la desviación típica.
3. Cuanta más pequeña sea la desviación típica mayor será la concentración de datos alrededor de la media.