MEDIDAS DE TENDENCIA CENTRAL, DE POSICIÓN Y DISPERSIÓN:

❖ MEDIDAS DE TENDENCIA CENTRAL:

Las medidas de tendencia central nos permiten comparar datos o en cierto sentido resumir la información.
Existen tres medidas de tendencia central: media, modo o moda, mediana.

CÁLCULO PARA DATOS SIN AGRUPAR.

Moda: Mo
La moda es el valor de la variable que se repite más veces, es decir, aquel que tiene mayor frecuencia absoluta. Dicha medida se puede calcular en cualquier topo de variable.

Media o promedio: $\bar{x}$
Sea X una variable que puede estar formada por los siguientes valores: $x_1, x_2, x_3, ..., x_n$.

La media es: $\bar{x} = \frac{n_1 x_1 + n_2 x_2 + ... + n_m x_m}{N}$. Siendo n_i con $1 \le i \le m$ las frecuencias absolutas.
Esta medida se puede calcular para variables cuantitativas.

Mediana: M_e
Si los elementos de la variable están colocados en orden creciente, el que ocupa el lugar central se llama mediana.
La mediana divide al conjunto de observaciones en dos conjuntos iguales.

- Si el número de observaciones es par, la mediana es la semi-suma de los dos valores centrales.
- Si el número de observaciones es impar, la mediana es el elemento que ocupa el lugar central.

Ejercicio 1):
El Ministerio de Viviendas realizó una encuesta para saber si el número medio de hijos por familia ha descendido respecto a la década anterior. Para ello ha encuestado a 50 familias respecto al número de hijos y ha obtenido los siguientes datos:

2	4	2	3	1	2	4	2	3	0	2	2	2	3	2	6	2	3	2	2	3	2	3	3	4
3	3	4	5	2	0	3	2	1	2	3	2	2	3	1	4	2	3	2	4	3	3	2	2	1

a) Construye la tabla de frecuencias a partir de estos datos.
b) ¿Cuántas familias tienen exactamente tres hijos?
c) ¿Qué porcentaje de familias tienen exactamente 3 hijos?
d) ¿Qué porcentaje de las familias de la muestra tienen más de dos hijos? ¿Y menos de 3?
e) Calcula las medidas de tendencia central.

CÁLCULO PARA DATOS AGRUPADOS.

Moda (Mo) es el valor más frecuente en el conjunto de datos, se encuentra en el intervalo de máxima frecuencia, que se llama clase modal. También se puede indicar como **intervalo modal**; siendo el intervalo con mayor frecuencia.

Media o promedio: $\bar{x}$

Sea X una variable que puede estar formada por los siguientes marcas de clase: $c_1, c_2, c_3, ..., c_n$.

La media es: $\bar{x} = \dfrac{n_1 c_1 + n_2 c_2 + ... + n_m c_m}{N}$. Siendo n_i con $1 \le i \le m$ las frecuencias absolutas.

Esta medida se puede calcular para variables cuantitativas.

Mediana: M_e

La mediana para datos agrupados se encuentra en el intervalo donde la frecuencia acumulada relativa alcanza al menos 50%; este intervalo se llama clase mediana.

La mediana se obtiene mediante la fórmula: $M_e = L_{i-1} + a_i \dfrac{\frac{N}{2} - N_{i-1}}{n_i}$ siendo:

L_{i-1} el límite inferior del intervalo en el que se alcanza al menos 50% de los datos;

a_i la longitud de dicho intervalo;

N_{i-1} la frecuencia absoluta acumulada del intervalo inmediatamente anterior

n_i la frecuencia absoluta del intervalo que se está considerando.

- **MEDIDAS DE POSICIÓN:**

Las medidas de posición son los cuartiles, deciles percentiles, sirven para encontrar puntos específicos en los datos no agrupados o agrupados de un estudio estadístico. También siendo una medida de centralización como es la mediana que representa el 50% de los datos porque divide a la serie o a la distribución en dos partes iguales se considera una medida de posición.

CÁLCULO PARA DATOS SIN AGRUPAR.

Cuartiles: Se dividen los datos en cuatro partes iguales. El cuartil **Q1** es el valor de la variable que ocupa la posición Q_1 correspondiente al primer 25% de los datos, el **Q2** es el valor de la variable que ocupa la posición Q_2 corresponde al 50% por lo tanto también debe ser igual a la mediana, y **Q3** es el valor de la variable que ocupa la posición Q_3 al 75%.

Para encontrar la posición: $Q_1 = \dfrac{N+1}{4}$, $Q_2 = \dfrac{2(N+1)}{4}$, $Q_3 = \dfrac{3(N+1)}{4}$

Dónde: Q_1, Q_2, Q_3 son la posición de los cuartiles, y N es el total de los datos.

Deciles: Se dividen los datos en 10 partes iguales.

Para encontrar la posición:

$D_1 = \dfrac{N+1}{10}$, $D_2 = \dfrac{2(N+1)}{10}$, $D_3 = \dfrac{3(N+1)}{10}$, $D_5 = \dfrac{5(N+1)}{10}$, … , $D_9 = \dfrac{9(N+1)}{10}$

Dónde: D_1, D_2, $D_3, \ldots, D_9$ son la posición de los deciles, y N es el total de los datos.

Percentiles: Se dividen los datos en 100 partes iguales.
Para encontrar la posición:

$$P_1 = \frac{N+1}{100}\ , \quad P_2 = \frac{2(N+1)}{100}, \quad P_3 = \frac{3(N+1)}{10}, \ldots, \quad P_{50} = \frac{50(N+1)}{100}, \ldots, \ P_{99} = \frac{99(N+1)}{100}$$

Dónde: P_1, P_2, $P_3, \ldots, P_{50}, \ldots P_{99}$, son la posición de los percentiles, y n es el total de los datos.

Ejemplo 1) :
De 20 estudiantes tenemos sus evaluaciones de un examen. Calcularemos el Q_1, D_5, P_{75}.

5,5,8,7,9,10,7,6,8,7,8,9,10,10,8,7,6,5,9,6.

- Primero deben ordenarse los números de forma ascendente:

5, 5, 5, 6, **6, 6,** 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 10, 10, 10.

$Q_1 = \frac{20+1}{4} = 5{,}25$ esta es la posición y la vamos a buscar en la serie de datos ya ordenada.
Por lo tanto; $\boldsymbol{Q1} = \mathbf{6}$.

- Ahora vamos a calcular el D_5

5, 5, 5, 6, 6, 6, 7, 7, 7, **7, 8**, 8, 8, 8, 9, 9, 9, 10, 10, 10.

$D_5 = \frac{5(20+1)}{10} = 10{,}5$ esta es la posición y la vamos a buscar en la serie de datos ya ordenada.

Por lo tanto, se calcula después de haber encontrado la posición que es 10.5, lo siguiente:

$\frac{7+8}{2} = 7{,}5$ por lo que el decile 5 (que coincide con la mediana) es: $\boldsymbol{D5} = \mathbf{7{,}5}$

- Por último calculamos P_{75}.

5, 5, 5, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, **9, 9**, 9, 10, 10, 10.

$P_{75} = \frac{75(20+1)}{100} = 15{,}75$ esta es la posición y la vamos a buscar en la serie de datos ya ordenada.
Por lo tanto; $\boldsymbol{P75} = \mathbf{9}$

CÁLCULO PARA DATOS AGRUPADOS.

Antes de utilizar la fórmula general, debemos de encontrar la posición en una distribución de frecuencias y esta se calcula de la siguiente forma:

Posiciones de los cuartiles: $Q_1 = \dfrac{N}{4}$, $Q_2 = \dfrac{2N}{4}$, $Q_3 = \dfrac{3N}{4}$

Posiciones de los deciles: $D_1 = \dfrac{N}{10}$,………., $D_5 = \dfrac{5N}{10}$,…, $D_9 = \dfrac{9N}{10}$

Posiciones de los percentiles: $P_1 = \dfrac{N}{100}$, …..., $P_{50} = \dfrac{50N}{100}$,…, $P_{99} = \dfrac{99N}{100}$

Fórmulas:

Cuartiles: $\mathbf{Qi} = L_{i-1} + \left(\dfrac{Q_i - N_{i-1}}{n_i}\right).a$ **Qi:** cuartil i-ésimo que se desea calcular.

Q_i: posición del cuartil i-ésimo que se desea calcular.

Deciles: : $\mathbf{Di} = L_{i-1} + \left(\dfrac{D_i - N_{i-1}}{n_i}\right).a$ **Di:** decil i-ésimo que se desea calcular.

D_i: posición del decil i-ésimo que se desea calcular.

Percentiles: : $\mathbf{Pi} = L_{i-1} + \left(\dfrac{P_i - N_{i-1}}{n_i}\right).a$ **Pi:** percentil i-ésimo que se desea calcular.

P_i: posición del percentil i-ésimo que se desea calcular.

Tener en cuenta que:

L_{i-1}: límite inferior donde se encuentra el cuartil a calcular.
N_{i-1}: frecuencia acumulada anterior al cuartil considerado.
n_i: frecuencia de la clase del cuartil donde se localiza.
a: amplitud del intervalo.

❖ MEDIDAS DE DISPERSIÓN:

Ejemplo 2:

El entrenador de un equipo de básquetbol duda entre seleccionar a Luciano o a Mario, para participar en un torneo. Los puntos conseguidos por cada uno en una semana de entrenamiento fueron:

Luciano	18	23	22	24	19	25	16
Mario	18	26	18	28	22	17	18

¿A cuál de los dos debe seleccionar justifica?

$\bar{x}(L) = \bar{x}(M) = 21$

Aquí podemos observar que las medidas de tendencia central no nos sirven, en estos casos se utilizan las **medidas de dispersión.**

La media es un dato importante, pero no dice nada acerca de la dispersión de los valores de la variable alrededor de ésta.

Y lo que se llama **varianza** es la que nos permite medir esto.

CÁLCULO PARA DATOS SIN AGRUPAR.

Varianza:

Es la suma de las distancias de los valores a la media al cuadrado, por las frecuencias absolutas dividido la suma de las frecuencias.

Es decir: $$S^2 = \frac{n_1\left(x_1 - \overline{x}\right)^2 + n_2\left(x_2 - \overline{x}\right)^2 + ... + n_n\left(x_n - \overline{x}\right)^2}{N}$$

Existe otra medida de dispersión llamada desvío que es la raíz cuadrada de la varianza, que es un parámetro más razonable, pues se expresa en la misma unidad que los datos y que la media.

Desvío:

Es la raíza cuadrada de la varianza.

$S = \sqrt{S^2}$

CÁLCULO PARA DATOS AGRUPADOS.

La varianza se calcula ponderando los cuadrados de las desviaciones de cada modalidad respecto de la media, por las frecuencias de clase y la desviación estándar es la raíz cuadrada de ese resultado.

VARIANZA : $S^2 = \dfrac{\sum_{i=1}^{k}\left(c_i - \overline{x}\right)^2 n_i}{N}$ **DESVÍO ESTÁNDAR** $S = \sqrt{S^2}$.

COEFICIENTE DE VARIACIÓN:

La dispersión de los datos estudiados no puede determinarse exclusivamente a partir de la desviación típica, pues no siempre una mayor desviación típica indica una mayor dispersión. Por ejemplo, la desviación típica del peso de un grupo de cinco caballos suele ser mayor que la desviación típica del peso de un grupo de 5 conejos. Es evidente que hay que considerar otro factor importante como la media de los pesos de ambos grupos de animales, para establecer una comparación relativa proporcional de la dispersión de los datos.

Una medida de la dispersión relativa de dos conjuntos de datos es el coeficiente de variación, que se define cómo:

$CV = \dfrac{S}{\overline{x}}$.

Para interpretar el coeficiente de variación hay que saber:

Dados dos conjuntos, aquel que tenga un coeficiente de variación mayor es el más disperso, el más heterogéneo. Además, su valor no depende de la unidad de medida utilizada, pues la media y la desviación típica se ven afectadas por la misma unidad de medida.

Ejercicio 2):

La estatura (en cm.), el número de calzado y el peso de 6 alumnas de tercero de bachillerato se dan en la siguiente tabla:

Estatura	164	158	162	168	172	166
Nro. de calzado	37	37	36	38	39	41
Peso	51	53	57	49	56	55

Halla la media y la desviación típica de cada conjunto de datos.

¿Qué conjunto es más disperso?

Obtiene los datos de 6 compañeras de tu clase y realiza la misma actividad.

Compara los resultados con las otras distribuciones.