UAP
UNIVERSIDAD ALAS PERUANAS

www.uap.edu.pe

ESTADISTICA Y PROBABILIDADES

Tema N°

MEDIDA DE DISPERSION

PROFESOR: ALEX IPARRAGUIRRE ZAVALETA

Nombres y Apellidos: _____

Eslcuela Académica Ingenieria Ambiental **III - Ciclo:**

4.1. MEDIDAS DE DISPERSIÓN O VARIABILIDAD

Las medidas de dispersión son aquellas que cuantifican el grado de concentración o de dispersión de los valores de la variable en torno de un valor central, generalmente la media aritmética. Las medidas de dispersión se utilizan para dos propósitos básicos:

a) Para verificar la confiabilidad de los promedios y

b) Para que sirva como base para el control de la variación misma.

También podemos decir que los términos concentración y dispersión pueden ser utilizados indistintamente, pues se da la relación.

Alta dispersión ⟶ Baja concentración

Baja dispersión ⟶ Alta concentración

Las medidas de dispersión que se utilizan con mayor frecuencia son:

Rango o recorrido de la variable.

- Rango Intercuartil
- Rango semicuartil
- Varianza
- Desviación estándar
- Coeficiente de variación
- Diagrama de caja

4.1.1. VARIANZA

Es una medida que cuantifica el grado de dispersión o de variación de los valores de una variable cuantitativa con respecto a su media aritmética. Si los valores tienden a concentrarse alrededor de su media, la varianza será pequeña. Si los valores tienen a distribuirse lejos de la media, la varianza será grande.

La varianza calculada a partir de una muestra se denota por s^2 y referida a la población se denota por σ^2 o V[x].

La varianza se define como la media aritmética de los cuadrados de las desviaciones de los datos respecto a su media aritmética. La varianza es una medida de dispersión con unidades de medición al cuadrado, $S/.^2$, $\$^2$, km^2, etc.

A. Formas de Cálculo

A.1. Para datos no agrupados

Se utiliza la siguiente fórmula:

Para n > 30

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n} = \frac{\sum_{i=1}^{n} y_i^{\,2}}{n} - \bar{y}^2$$

Para n ≤ 30 [varianza de Cochran]

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^{\,2} - n\bar{y}^2}{n-1}$$

Ejemplo:

Calcular e interpretar la varianza de los puntajes de 10 estudiantes en un examen de estadística.

X_i = 13, 14, 16, 15, 14, 16, 17, 16, 12, 11

Solución:

De los datos dados obtenemos:

$$\sum_{i=1}^{10} y_i^{\,2} = 2108$$

$$\bar{x} = 14.4$$

$$n = 10$$

Reemplazando en la fórmula (Varianza de Cochran n < 30)

$$s^2 = \frac{\sum_{i=1}^{10} y_i^{\,2} - n\bar{y}^2}{n-1} = \frac{2108 - 10(14.4)^2}{9} = 3.82\,\text{puntos}^2$$

Interpretación: La variación de los puntajes es de 3.82 puntos^2

A.2. <u>Para datos agrupados</u>

Se utiliza la siguiente fórmula:

Para n > 30

$$s^2 = \frac{\sum_{i=1}^{m}(y_i - \bar{y})^2}{n} = \frac{\sum_{i=1}^{m} y_i^{\,2} \times f_i}{n} - \bar{y}^2$$

Para n ≤ 30 [varianza de Cochran]

$$s^2 = \frac{\sum_{i=1}^{m}(y_i - \bar{y})^2 \times f_i}{n-1} = \frac{\sum_{i=1}^{m} y_i^{\,2} \times f_i - n\bar{y}^2}{n-1}$$

Ejemplo:

Calcular e interpretar la varianza de los datos dados: Distribución de Turistas según sus gastos semanales en dólares

.

Solución:

TABLA N°

Gastos semanales en soles $LI - LS$	y_i	f_i	y_i^2	$y_i^2 \times f_i$
[400 – 520)	460	2	211600	423200
[520 – 640)	580	3	336400	1009200
[640 – 760)	700	8	490000	3920000
[760 – 880)	820	4	672400	2689600
[880 – 1000)	940	3	883600	2650800
Total	-	20		9809200

De los datos obtenemos:

$$\sum_{i=1}^{5} y_i^2 \times f_i = 10692800$$

$$\overline{y} = 718$$

$$n = 20$$

Reemplazando estos valores en la fórmula obtenemos (Varianza de Cochran, n<30)

$$s^2 = \frac{\sum_{i=1}^{5} y_i^2 \times f_i - n.\overline{x}^2}{n-1}$$

$$s^2 = \frac{10692800 - 20x(718)^2}{19}$$

$$s^2 = 20122.11 \text{ dólares}^2$$

Interpretación: La variación de los gastos semanales es de 20122.11 dólares^2.

4.1.2. LA DESVIACIÓN ESTÁNDAR O TÍPICA

Se define como la raíz cuadrada positiva de la varianza:

$$s = \sqrt{\text{var } ianza}$$

Es uno de los estadígrafos de dispersión de mayor uso, la cual se expresa en unidades reales de la variable, es decir ya no están elevadas al cuadrado. La desviación estándar, al igual que la varianza, es no negativa (S ≥ 0), puesto que es la raíz positiva de la varianza. A mayor dispersión le corresponderá una mayor desviación estándar.

La desviación estándar o típica.

Ejemplo 51:

Calcular e interpretar la desviación estandar de los datos del ejemplo N° 49. Solución:

$$s^2 = \sqrt{8.82} = 1.95 \text{ puntos}$$

Interpretación: Los puntajes de los alumnos se alejan en promedio de su valor central en 1.95 puntos.

Ejemplo 52:
Calcular e interpretar la desviación estándar de los datos referidos a los gastos semanales en dólares de 20 turistas.
Solución:

$$s^2 = \sqrt{20122.11} = S/.141.86$$

Interpretación: Los gastos semanales en dólares de los 20 turistas se dispersan de su valor central en S/.141.86.

4.1.3. Coeficiente de variación

Es una medida de dispersión relativa exenta de unidades y expresada en porcentaje, se utilizan para comparar la variación de dos distribuciones siempre que las variables se expresen en las mismas unidades de medida y sean aproximadamente del mismo tamaño promedio. Sin embargo, a veces es necesario comparar dos conjuntos de datos expresados en unidades diferentes (tales como soles y kilogramos). En estos casos las medidas de dispersión absoluta no son comparables y deben utilizarse medidas de dispersión relativa.

El coeficiente de variación de un conjunto de datos se denota por C.V. y se expresa como:

$$C.V. = \frac{s}{\bar{y}} \times 100$$

- Si C.V. $\leq$ 15%, los datos son homogéneos, es decir tienen una baja variabilidad.
- Si C.V. > 15%, los datos son heterogéneos, es decir tienen una alta variabilidad.

Ejemplo 53:
Calcular e interpretar el coeficiente de variación de los datos referidos a los gastos semanales en dólares de 20 turistas.
.
Solución:

$$CV = \frac{141.86}{718} \times 100 = 19.80\% > 15\%$$

Interpretación: Los gastos semanales en soles de los turistas son heterogéneos, es decir presentan un alto grado de variabilidad.

4.2. MEDIDAS DE ASIMETRIA

Son medidas que miden el grado de formación horizontal de una distribución de frecuencias.

Se dice que una distribución de frecuencias es simétrica, si los intervalos equidistantes del intervalo central tienen iguales frecuencias. También se dice que una distribución es simétrica si su curva de frecuencias es simétrica con respecto al centro de los datos.

Dos distribuciones pueden tener la misma media y la misma desviación estándar, pero pueden diferir en el grado de asimetría.

Si la distribución es simétrica, entonces la media, la mediana y la moda coinciden. En contraposición, si estos 3 promedios no coinciden la distribución es asimétrica.

Entre las medidas de asimetría más usuales tenemos:

- **<u>El coeficiente de asimetría de Pearson</u>**

Se expresa como:

1) $$As = \frac{3(\bar{x} - Me)}{s}$$

Otra forma de expresar el coeficiente de asimetría de Pearson es:

2) $$As = \frac{(\bar{x} - Md)}{s}$$

- **<u>La Media Asimétrica</u>**

3) $$Ax = \frac{Q_3 - 2Me + Q_1}{Q_3 - Q_1}$$

4) $$A_K = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$

El Primer Coeficiente de Pearson (1º) es el más usual.

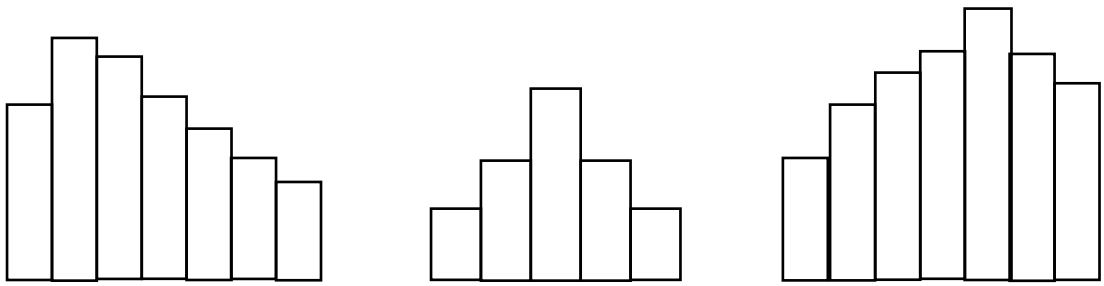
El segundo coeficiente de Pearson (2º) se usa cuando la distribución es unimodal.

El tercer y cuarto coeficientes de Asimetría (3º y 4º) llamados también Media Asimétrica se usa cuando existen intervalos con extremos abiertos ilimitados y no sea posible calcular la Media y consecuentemente la varianza.

Si:

$A_K = 0$ ⟹ La distribución es simétrica. Ver fig.

$A_K > 0$ ⟹ La distribución es asimétrica positiva (sesgada a la Derecha). Ver fig.



Asimetría Positiva Distribución Simétrica Asimétrica Negativa

Ejemplo 54:

Calcular e interpretar el coeficiente de asimetría de los datos en la Tabla N° 6.

Solución:

a) Utilizando el coeficiente de Pearson (1º)

$$As = \frac{3(\bar{x} - Me)}{s}$$

De los datos dados se ha obtenido:

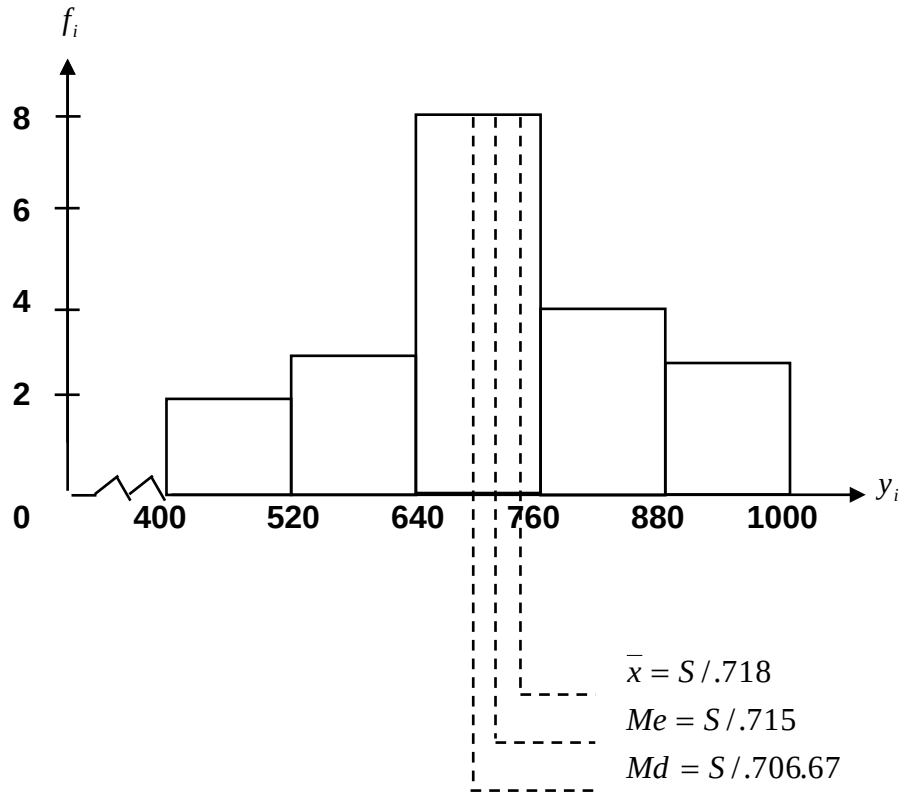
$\bar{x} = S/.718,$ $Me = S/.715,$ $s = S/.141.86$

Reemplazando en la fórmula obtenemos:

$$As = \frac{3(718-715)}{141.86} = 0.06$$

Interpretación: Este valor indica que la distribución es asimétrica positiva.

b) Construyendo el histograma de frecuencias.



Como podemos observar:

$$Md < Me < \bar{x}$$

Lo cual indica que la distribución es asimétrica.

4.3. MEDIDAS DE KURTOSIS

La Kurtosis es el grado de apuntamiento de una distribución. La Kurtosis sea naliza amparando la distribución con la forma de una curva normal o simétrica, con igual media aritmética y desviación estándar de la distribución que se estudia.

Si una distribución tiene relativamente un elevado pico o apuntamiento, se llama leptokúrtica, mientras si es achatada se denomina platikúrtica. La distribución normal constituye una distribución mesokúrtica, tal como se puede ver en las siguientes figuras:

Leptokúrtica

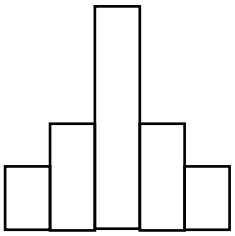


Fig. 7

Mesokúrtica

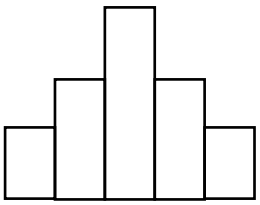


Fig. 8

Platikúrtica

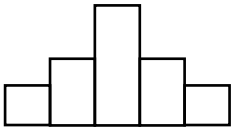


Fig. 9

El estadígrafo para analizar el apuntamiento es el coeficiente de kúrtosis y se expresa como:

1) $K = \dfrac{m^4}{s^4}$

Donde:

$s^4 = (s^2)^2$

$s^2 = Varianza$

$$m^4 = \frac{\sum_{i=1}^{m} (y_i - \overline{y})^4 \times f_i}{n}$$

m^4 se llama: “cuarto momento respecto a la media”

Si:

$a_4 = 3$, la distribución es normal o mesokúrtica.

$a_4 < 3$, la distribución es platikúrtica.

$a_4 > 3$, la distribución es leptokúrtica.

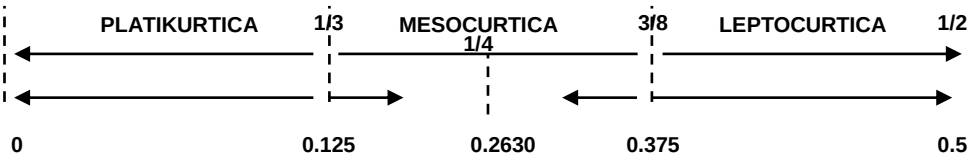
Otro estadígrafo de kúrtosis es:

2) $K = \dfrac{Q_3 - Q_1}{2(P_{90} - P_{10})}$

Si:

Para la “curva normal” K = 0.2630

Una distribución será Mesokúrticaa si K tiende α 0.2630 por ambos lados. Si el valor de K se aleja hacia la derecha tendiendo a y_2, la distribución será Leptokúrtica. Si se aleja hacia la izquierda tendiendo a cero, la distribución será Platikúrtica. Puede ser útil observar el siguiente esquema para determinar la kúrtosis.



NOTA: Para las distribuciones notablemente discrepantes de la normal estas medidas pueden resultar contradictorias.

Ejemplo:

Calcular e interpretar el coeficiente de kurtósis para los datos dados referidos a los gastos semanales en dólares de 20 turistas.

Solución:
Utilizando el coeficiente de Kúrtosis (1) obtenemos:

$$a_4 = \frac{m_4}{(s^2)^2} = \frac{883505232}{(2012211)^2} = 2.18 < 3$$

Interpretación: Este valor indica que la distribución es platikúrtica.

Ejercicios

1. Se presenta un cuadro donde se evidencia el total de visitantes por edades al museo. Hallar las medidas estudiadas e interpretar los resultados.

Intervalo	0 - 5	5 - 10	10 - 15	15 - 20	20 - 25	25 - 30
Frecuencia	3	9	12	9	15	2

2. En un autobús escolar se les pregunta a los alumnos por el tiempo que tardan en llegar de su casa al autobús. Los resultados se recogen en la siguiente tabla:
Hallar las medidas estudiadas e interpretar los resultados**.**

TIEMPO (minutos)	0 - 5	5 - 10	10 - 15	15 - 20	20 - 25
N.º de alumnos	20	13	18	5	4

3. Al preguntar en 50 familias por el número de personas que forman el hogar familiar, hemos obtenido la información que se recoge en la siguiente tabla:
Hallar las medidas estudiadas e interpretar los resultados**.**

N.º de personas	1	2	3	4	5	6
N.º de familias	3	10	23	9	3	2