

Ensayo

Nombre: Lisseth Esmeralda Mendoza Moreno

Docente: Juan José Ojeda

Grado: 1 cuatrimestre

Grupo: LTS21EMCO120-A

Tema: medidas de posición y variación para datos agrupados y no agrupados.



Datos agrupados y no agrupados.

Desarrollo

La presente investigación sobre los datos agrupados y no agrupados, en los datos agrupados son aquellos datos que tienen como característica principal la "FRECUENCIA" con que se presentan es decir que serán aquellos que se encuentran contados y clasificados.

Los datos no agrupados veremos que son aquellos que carecerán de frecuencia ya sea porque no se encuentran contabilizados o clasificados.

También veremos la frecuencia de clase, la frecuencia relativa, el punto medio entre otros temas, el punto de esta introducción es llegar a que el docente tenga el conocimiento de lo que contiene este ensayo para que comprenda de una manera breve los temas que anteriormente se dio a conocer.

DESARROLLO

Como bien mencionamos que los datos agrupados son aquellos que se necesitan agrupar para saber la muestra de la población pero solo cuando son de 30 o más de donde se fue tomada y se pueden agrupar por medio de la media, mediana moda entre otras más y también cuando los datos se agrupan por muestra de clase es necesario separar los datos.

En el caso de los datos no agrupados son aquellos que se debe tener menos de 20 elementos en la muestra ya en este caso se analizan y ya no hay necesidad de formar clases y por esto se llaman de esa forma "no agrupados".

Cuando hablamos **frecuencia de clase** se habla de dos tipos los cuales son;

Marca de clase "punto medio": punto que divide a la clase en dos partes iguales. Es el promedio entre los límites superior e inferior de la clase.

Intervalo de clase: para una distribución de frecuencias que tiene clases del mismo tamaño, el intervalo de clase se obtiene restando el límite inferior de una clase del límite inferior de la siguiente.

Frecuencia relativa

Hace referencia a una media estadística mediante la cual se puede obtener el número de veces que un mismo suceso es repetido dentro de un momento determinado. Estos datos de estadística que arroja la frecuencia relativa, nos ayudan a descubrir cómo se repite o cada cuánto se repite un mismo suceso o evento. En pocas palabras como dice la antología, Es la relación o cociente entre la frecuencia absoluta y el número total de observaciones. Es la proporción entre la frecuencia de un intervalo y el número total de datos.

PUNTO MEDIO

es el valor medio de un conjunto de datos. Para un conjunto de datos con un número impar de miembros, es el miembro del conjunto de datos con apenas tantos valores menos que su valor como mayor que él. Punto medio es el punto que divide a un segmento en dos partes iguales. El punto medio de un segmento, es único y equidista de los extremos del segmento. Cumpliendo esta última condición, pertenece a la mediatriz del segmento.

La fórmula para determinar el punto medio de un segmento en el plano, con coordenadas:

(x_1, y_1) y (x_2, y_2) es: $[(x_1 + x_2) / 2] + [(y_1 + y_2) / 2]$.

LIMITES

Son los valores extremos que tiene el intervalo de clase, inferior y superior, entre los cuáles van a estar los valores de los datos agrupados en ese intervalo de clase.

TRATAMIENTO PARA DATOS NO AGRUPADOS .

Medidas de tendencia central: las medidas estadísticas que pretenden resumir en un solo valor a un conjunto de valores. Representan un centro en torno al cual se encuentra ubicado el conjunto de los datos. Los cuales son media aritmética, la mediana, la media geométrica, la moda, etc. estas tienden a estar localizadas generalmente en su parte central.

Algunas medidas de tendencia central son: MEDIA ARITMÉTICA, MEDIANA Y MODA

1.-Media aritmética (\bar{x})

También se le conoce como promedio ya que es el promedio de las lecturas o mediciones individuales que se tienen en la muestra, se determina con la fórmula siguiente:

$$\text{Media aritmética} = \frac{\sum_{i=1}^n x_i}{N} = \frac{x_1 + x_2 + x_3 \dots + x_n}{N}$$

Formula media aritmetica

Dónde

-
 \bar{x} = media aritmética
 x_i = dato i

n = número de datos en la muestra.

EJEMPLOS media aritmética,

1.- Se han tomado como muestra las medidas de seis cables usados en un arnés para lavadora, las cuales son; 15.2 cm, 15.0, 15.1, 15.2, 15.1 y 15.0, determine su mediar aritmética.

Solución: -

$$X = \frac{15.2+15.0+15.1+15.2+15.1+15.0}{6} = 15.1\text{cm}$$

2.- Se toman varias muestras de cierto tipo de queso y se determina la cantidad de proteína por cada 100 gramos de queso, encontrándose lo siguiente: 26.5 gramos, 24.8, 25.3, 30.5, 21.4, determine la cantidad promedio de proteína encontrada en la muestra por cada 100 gramos de queso que se elabora.

Solución: :

$$X = \frac{26.5+24.8+25.3+30.5+21.4}{5} = 25.7 \text{ grs}$$

3.-Se hacen varias lecturas de una muestra que contiene cobre, las lecturas se hacen en un espectrofotómetro de absorción atómica y son la siguientes: 12.3%, 12.28, 12.27, 12.3, 12.24, 15.01, determine la concentración promedio de Cu en la muestra.

Solución: -

$$X = \frac{12.3+12.28+12.27+12.3+12.24+15.01}{6} = \frac{76.4}{6} = 12.73 \% \text{Cu}$$

Si observamos las lecturas del espectrofotómetro nos damos cuenta que el valor de 15.01% es un valor diferente al de las lecturas anteriores, por lo que se descarta el valor ya que se considera un valor atípico, es decir un valor que es debido a circunstancias especiales, en este caso puede ser que se deba al hecho de que se está descalibrando el aparato de absorción atómica o simplemente que se ha

equivocado el operador del aparato al tomar la lectura, por lo que la media se debe calcular con las primeras cinco lecturas; como se muestra a continuación:

Solución _

$$X = \frac{12.3+12.28+12.27+12.3+12.24}{5} = \frac{61.39}{5} = 12.278 \%Cu$$

4.- Si deseamos determinar la edad promedio de los estudiantes de una escuela de nivel superior al iniciar sus estudios, suponga que se toman las edades de algunos de los alumnos de cierta clase y estas son las que siguen: 20, 18, 18, 19, 18, 19, 35, 20, 18, 18, 19.

Solución: Luego, la media se determinará con solo 10 de las edades ya que es necesario descartar la edad de 35 años, que es un dato atípico o un caso especial, por lo que;

—

$$X = \frac{20+18+18+19+18+19+20+18+18+19}{10} = \frac{187}{10} = 18.7 \text{ años}$$

2.-Media geométrica $\approx G$:

Es la raíz en enésima del producto de los valores de los elementos de la muestra, es usada cuando los valores de los datos de la muestra no son lineales, es decir que su valor depende de varios factores a la vez, se determina de la siguiente forma:

$$G = \sqrt[n]{x_1 * x_2 * \dots * x_n}$$

Donde:

G = media geométrica

xi = dato i

n = número de datos en la muestra

EJEMPLOS de media geométrica (G)

1.- Las siguientes temperaturas han sido tomadas de un proceso químico, 13.4°C, 12.8, 11.9, 13.6, determine la temperatura promedio de este proceso.

Solución

$$G = \sqrt[4]{13.4 \times 12.8 \times 11.9 \times 13.6} = \sqrt[4]{27758.7968} = 12.9077 \text{ } ^\circ\text{C}$$

2.- Las siguientes temperaturas han sido tomadas de un proceso para fabricar queso chihuahua, 21.4°C, 23.1, 20.2, 19.7, 21.0, determine la temperatura promedio de este proceso.

Solución

$$G = \sqrt[5]{21.4 \times 23.1 \times 20.2 \times 19.7 \times 21.0} = \sqrt[5]{4131070.852} = 21.048 \text{ } ^\circ\text{C}$$

3.- Media aritmética ponderada (\bar{x}^w).

Esta media se usa cuando el peso que tiene cada uno de los datos de la muestra es diferente, se calcula de la siguiente manera:

$$\bar{x}_w = \frac{\sum_{i=1}^p w_i X_i}{\sum_{i=1}^p w_i}$$

Donde

\bar{x}^w = media aritmética ponderada

x^i = dato i

w^i = peso del dato i

EJEMPLO Media aritmética ponderada (X^w)

A continuación se mencionan las materias que Luis Pérez llevó en el primer semestre de Ingeniería Química, el número de créditos y la calificación obtenida.

MATERIA	NUMERO CREDITOS	CALIFICACION
Metodología de la investigación	8	90.5
Matemáticas I	10	100.0
Programación	8	81.0
Química	10	78.0
Dibujo	4	100.0
Economía	8	84.0

Determine la calificación promedio que obtuvo Luis Pérez en su primer semestre.

Solución

$$X^w = \frac{(8 \times 90.5) + (10 \times 100.0) + (8 \times 81.0) + (10 \times 78.0) + (4 \times 100.0) + (8 \times 84.0)}{8 + 10 + 8 + 10 + 4 + 8} = \frac{724 + 1000 + 648 + 780 + 400 + 672}{48} = \frac{4224}{48} = 88.0$$

4.-Media armónica (H)

La media armónica se define como el recíproco del promedio de los recíprocos de cada uno de los datos que se tienen en la muestra, y se determina de la siguiente manera:

$$H = \frac{1}{\frac{1}{n} \sum_{i=1}^n 1/x_i} = \frac{n}{\sum_{i=1}^n 1/x_i}$$

EJEMPLOS de media armónica (H).

Solución :

Determine la media armónica de los siguientes datos, 3.1, 2.8, 2.84, 3.05, 3.09

$$H = \frac{5}{\frac{1}{3.1} + \frac{1}{2.8} + \frac{1}{2.84} + \frac{1}{3.05} + \frac{1}{3.09}}$$
$$= \frac{5}{0.3226 + 0.3571 + 0.3521 + 0.3279 + 0.32361} = \frac{5}{1.6833} = 2.9703$$

5.-Mediana (X_{med})

La mediana es aquel valor que se encuentra en la parte central de los datos que se tienen en la muestra una vez que estos han sido ordenados según su valor o magnitud. Para calcular la mediana se presentan dos casos:

a. Cuando el número de datos en la muestra es impar.- En este caso después de ordenar los datos de la muestra en cuanto a su magnitud, es decir de mayor a menor valor o de menor a mayor valor, se procede a localizar aquel dato que se encuentra justo en el centro de los datos o en la parte central de los mismos, el valor de este dato será el que dé valor a la mediana.

EJEMPLO de la mediana (X_{med})

1.- Los siguientes datos son las mediciones obtenidas de un circuito utilizado en un arnés de lavadora; se toman como muestra siete circuitos y sus mediciones son: 11.3, 11.2, 11.5, 11.2, 11.2, 11.4, 11.5 cm.

Solución: Ordenando los datos de menor a mayor valor; 11.2, 11.2, 11.2, 11.3, 11.4, 11.5, 11.5.

Se observa que el dato 11.3 es el que queda en la parte central, por lo que este es el que dará valor a la mediana; entonces, $x_{med} = 11.3$ cm.

b. Cuando el número de datos en la muestra es par.- En este caso después de ordenar los datos en cuanto a su magnitud, observamos que en la parte central de los datos no se encuentra dato alguno, en este caso, la mediana tomará el valor del promedio de dos datos; el que se encuentra antes de la parte central y el que se encuentra después de la parte central.

2.- Los siguientes datos son las mediciones obtenidas de un circuito utilizado en un arnés de lavadora; se toman como muestra ocho circuitos y sus mediciones son: 11.3, 11.2, 11.5, 11.2, 11.2, 11.4, 11.5, 11.4 cm.

Solución

Ordenando los datos de mayor a menor valor:

11.5, 11.4, 11.4, 11.3, 11.2, 11.2, 11.2, 11.1 cm.

Se observa que en la parte central de los datos no hay dato alguno por lo que la mediana se determina con el promedio de los datos subrayados, entonces,

$$X_{med} = \frac{11.3 + 11.2}{2} = 11.25\text{cm}$$

6.-Moda (X_{mod}).

La moda se define como aquel valor o valores que más se repiten o que tienen mayor frecuencia entre los datos que se han obtenido en una muestra, la muestra de una población nos genera la distribución de los datos una vez que estos se han graficado y en esta gráfica es posible observar la moda o modas de la misma, es por esto que una distribución de datos puede ser amodal (carece de moda), unimodal (tiene una sola moda), bimodal (tiene dos modas) o polimodal (tiene más de dos modas).

EJEMPLOS de Moda (x_{mod}).

1.-Determine la moda de los datos que se muestran a continuación, se refieren a la estatura de un grupo de jóvenes; 1.60m, 1.65, 1.70, 1.71, 1.70, 1.70, 1.70, 1.71, 1.70, 1.93, 1.87, 1.85.

Solución :

Estatura	Frecuencia
1.60	1
1.65	1
1.70	5*
1.71	2
1.85	1
1.87	1
1.93	1

La tabla muestra la distribución de frecuencias de los datos o el número de veces que estos se repiten, la mayor frecuencia que es 5 corresponde a una estatura de 1.70m, por lo que esta sería la moda.

Luego, $x_{mod} = 1.70m$

2.- Determine la moda de los siguientes datos que se refieren a la edad de alumnos de primer semestre del tecnológico de Chihuahua, 18 años, 17, 19, 21, 19, 18, 22, 22, 18, 18, 17, 19, 19, 19, 18, 20, 21, 20, 18, 19, 18, 19, 18,19, 22, 35.

Solución

Edad	Frecuencia
17	2
18	7*

19	8*
20	2
21	2
22	3
35	1

En este caso se observa que las edades que más frecuencia tienen son las de 18 y 19 años, por lo que se concluye que existen dos modas,

$X_{\text{mod1}} = 18$ años , $X_{\text{mod2}} = 19$ años

Hay que hacer notar que la frecuencia para ambas modas puede ser de igual magnitud o diferente, como en el caso que se ilustra.

b2. Medidas de Dispersión.

Cuando se tiene una muestra de datos obtenida de una población cualquiera, es importante determinar sus medidas de tendencia central así como también es básico el determinar que tan dispersos están los datos en la muestra, por lo que se hace necesario determinar su rango, la varianza, la desviación estándar, etc., ya que una excesiva variabilidad o dispersión en los datos indica la inestabilidad del proceso en análisis en la mayoría de los casos.

1.- *Rango o recorrido*. El rango es la diferencia entre el valor mayor y el valor menor encontrados en la muestra, también se le denomina recorrido ya que nos dice entre que valores hace su recorrido la variable de interés; y se determina de la siguiente manera:

$$R = V_M - V_m$$

Dónde

R = rango o recorrido

V_M = valor mayor en la muestra

V_m = valor menor en la muestra

EJEMPLOS de RANGO o RECORRIDO

1.-Se han tomado como muestras las mediciones de la resistencia a la tensión de la Soldadura usada para unir dos cables, estas son: 78.5kg, 82.4, 87.3, 78.0, 90.0, 86.5, 77.9, 92.4, 75.9, determine su rango o recorrido.

Solución

$$V_M = 92.4 \text{ kg}$$

$$V_m = 75.9 \text{ kg}$$

$$R = V_M - V_m = 92.4 - 75.9 = 16.5 \text{ kg}$$

2.-Se toman las mediciones de la cantidad de grasa de la leche en gramos por cada 100 ml de leche que entra a un proceso de pasteurización, a continuación se enumeran; 14.85, 15.32, 12.76, 16.29, 15.84, 17.3, 17.61, 16.33, determine el rango o recorrido de la cantidad de grasa de la leche.

Solución

$$V_M = 17.61$$

$$V_m = 12.76$$

$$R = 17.61 - 12.76 = 4.85 \text{ gramos}$$

2. Desviación absoluta media (\bar{d})

Esta medida de dispersión nos representa la diferencia absoluta promedio que existe entre cada dato que se encuentra en la muestra y la media de los datos y se determina de la siguiente manera:

donde :

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

x_i = dato i

\bar{x} = media aritmética de la muestra.

n = número de datos en la muestra.

EJEMPLO de DESVIACIÓN absoluta media (d)

1.- Determine la desviación absoluta media de los siguientes datos que son las concentraciones de plomo de algunas muestras, las que a continuación se enumeran: 18gr, 12, 21, 19, 16, 20, 22.

Solución

Para determinar la desviación absoluta media o promedio, lo primero que hay que hacer es calcular la media aritmética de los datos de la muestra, la que es $128/7 = 18.286$, luego se procede a calcular el promedio de las diferencias absolutas entre cada dato y la media calculada.

$$\bar{d} = \frac{|18 - 18.286| + |12 - 18.286| + \dots + |20 - 18.286| + |22 - 18.286|}{7} =$$

$$\bar{d} = \frac{0.286 + 6.286 + 2.714 + 0.714 + 2.286 + 1.714 + 3.714}{7} = \frac{17.714}{7} = 2.5305 \text{ gr}$$

3: Varianza o variancia (s^2).

Es el promedio de las diferencias elevadas al cuadrado entre cada valor que se tiene en la muestra (x_i) y la media aritmética (\bar{x}) de los datos y se determina de la siguiente manera:

donde : n es el número de datos en la muestra.

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

EJEMPLO Varianza o variancia (s²).

1.-Los siguientes datos es la cantidad de glucosa en miligramos encontrada en muestras de sangre de algunos pacientes, 14.2, 12.1, 15.6, 18.1, 14.3, determine su varianza.

Solución

Lo primero que hay que calcular es la media aritmética de la muestra como ya se ha hecho anteriormente.

$$\bar{x} = \frac{14.2+12.1+15.6+18.1+14.3}{5} = \frac{74.3}{5} = 14.86mg$$

$$s^2 = \frac{(14.2-14.86)^2 + (12.1-14.86)^2 + \dots + (14.3-14.86)^2}{5-1} =$$

$$s^2 = \frac{0.4356 + 7.6176 + 0.5476 + 10.4976 + 0.3136}{4} = \frac{19.412}{4} = 4.853mg^2$$

4: Desviación estándar (s).

Es la desviación o diferencia promedio que existe entre cada dato de la muestra y la media aritmética de la muestra. Y se obtiene a partir de la varianza, sacándole raíz cuadrada.

donde : s

s² = varianza o variancia

$$s = \sqrt{s^2}$$

Por tanto la desviación estándar de la muestra anterior sería;

$$s = \sqrt{4.853mg^2} = 2.2029mg$$

La interpretación de este resultado sería, que la cantidad de glucosa encontrada en la muestra es en promedio de 14.86 miligramos y que la cantidad de glucosa en la muestra se aleja o dispersa en promedio 1.9704 mg alrededor de la media.

1. Media aritmética para datos agrupados

Se calcula sumando todos los productos de marca clase con la frecuencia absoluta respectiva y su resultado dividirlo por el número total de datos:

$$\bar{X} = \frac{\text{Suma (marca clase} \times \text{frecuencia absoluta)}}{\text{Total de datos}}$$

La marca clase de una tabla para datos agrupados en intervalos corresponde al promedio de los extremos de cada intervalo.

El intervalo

26 – 30

26: corresponde al extremo inferior del intervalo.
30: corresponde al extremo superior del intervalo

En el intervalo anterior la **marca de clase** es 28 es decir $\bar{x} = \frac{26 + 30}{2} = 28$

2- Moda

1.-Es el valor que representa la mayor frecuencia absoluta. En tablas de frecuencias con datos agrupados, hablaremos de intervalo modal.

La moda se representa por Mo.

2.1- Todos los intervalos tienen la misma amplitud.

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot t_i$$

L_i Extremo inferior del intervalo modal (intervalo que tiene mayor frecuencia absoluta).

f

i Frecuencia absoluta del intervalo modal.

f

$i-1$ Frecuencia absoluta del intervalo anterior al modal.

f

$i+1$ Frecuencia absoluta del intervalo posterior al modal.

t_i Amplitud de los intervalos.

Si los intervalos tienen amplitudes distintas.

En primer lugar tenemos que hallar las alturas.

$$H_i = f_i / t_i$$

Donde

h_i : altura correspondiente a cada intervalo.

f_i : Frecuencia absoluta del intervalo (también se puede utilizar la frecuencia acumulada o relativa)

t_i : Amplitud de los intervalos

Luego la clase modal es la que tiene mayor altura.

3- Mediana

Es el valor que ocupa el lugar central de todos los datos cuando éstos están ordenados de menor a mayor. La mediana se representa por Me. La mediana se puede hallar sólo para variables cuantitativas.

Cálculo de la mediana para datos agrupados

La mediana se encuentra en el intervalo donde la frecuencia acumulada llega hasta la mitad de la suma de las frecuencias absolutas. Es decir tenemos que buscar el intervalo en el que se encuentre. $N / 2$.

Fórmula:

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot t_i$$

L_{i-1} es el límite inferior de la clase donde se encuentra la mediana.

$N / 2$ es la semisuma de las frecuencias absolutas.

F_{i-1} es la frecuencia acumulada anterior a la clase mediana.

f_i es la frecuencia absoluta del intervalo mediano.

T_i es la amplitud de los intervalos.

EJEMPLO

En la siguiente tabla se muestran las edades de un grupo de personas.

1° Calculemos la media aritmética:

Edad	Marca clase (X_j)	Frecuencia absoluta (f_j)	Frecuencia acumulada (F_j)
[0 - 10)	5	3	3
[10 - 20)	15	6	9
[20 - 30)	25	7	16
[30 - 40)	35	12	28
[40 - 50)	45	3	31

$$N = 31$$

$$\bar{X} = \frac{5 \times 3 + 15 \times 6 + 25 \times 7 + 35 \times 12 + 45 \times 3}{31} =$$

$$\bar{X} = \frac{15 + 90 + 175 + 420 + 135}{31} = \frac{835}{31} = 26,94$$

$$\bar{X} = 26,94$$

2° Ahora calculemos la mediana (Me) según las fórmulas explicadas más arriba:

Lo primero que debemos hacer para poder calcular la mediana es identificar la clase mediana. Para esto tenemos que buscar el intervalo en el que se encuentre. $N / 2$ en este caso $N / 2 = 31 / 2 \Rightarrow 15,5$

Ahora debemos buscar el intervalo donde la frecuencia acumulada (F_i) contenga el valor obtenido (15,5)

Edad	Marca clase (X _j)	Frecuencia absoluta (f _j)	Frecuencia acumulada (F _j)
[0-10)	5	3	3
[10-20)	15	6	9
[20-30)	25	7	16
[30-40)	35	12	28
[40-50)	45	3	31

N = 31

$$\frac{N}{2} = \frac{31}{2} = 15,5$$

Ahora reemplazamos los datos en la fórmula:

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot t_i$$

$$Me = 20 + \frac{15,5 - 9}{7} \cdot 10$$

$$Me = 20 + 9,29$$

$$Me = 29,285$$

3° Calculemos la moda Mo:

Edad	Marca clase (X _j)	Frecuencia absoluta (f _j)	Frecuencia acumulada (F _j)
[0-10)	5	3	3
[10-20)	15	6	9
[20-30)	25	7	16
[30-40)	35	12	28
[40-50)	45	3	31

$$N = 31$$

Intervalo modal:
mayor frecuencia absoluta

Ahora podemos reemplazar los datos en la fórmula:

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot t_i$$

$$Mo = 30 + \frac{12 - 7}{(12 - 7) + (12 - 3)} \cdot 10$$

$$Mo = 30 + 3,57$$

$$Mo = 33,6$$

- Si la moda está en el primer intervalo, entonces $f_{i-1} = 0$. Si la moda está en el último intervalo, entonces $f_{i+1} = 0$.
- Puede haber más de una moda en el caso en que dos o más valores de la variable presenten la misma frecuencia.

4.-Cuantiles

Los cuantiles son medidas de posición que se determinan mediante un método que determina la ubicación de los valores que dividen un conjunto de observaciones en partes iguales. Son los valores de la distribución que la dividen en partes iguales, es decir, en intervalos que comprenden el mismo número de valores. Cuando la distribución contiene un número alto de intervalos o de marcas y se requiere obtener un promedio de una parte de ella, se puede dividir la distribución en cuatro, en diez o en cien partes.

Algunos cuantiles más usados son los siguientes los cuales son cuatro:

Los deciles, cuando dividen la distribución en diez partes y los centiles o percentiles, cuando dividen la distribución en cien partes. Los cuantiles, como los deciles y los percentiles, son en cierta forma una extensión de la mediana.

A continuación hay una tabla de algunos valores:

Para algunos valores u , se dan nombres particulares a los cuantiles, $Q(u)$:

u	$Q(u)$
0.5	Mediana
0.25, 0.75	Cuartiles
0.1, ... , 0.99	Deciles
0.01, ..., 0.99	Centiles

5.- cuartiles

En los cuartiles son los tres valores que dividen al conjunto de datos ordenados en cuatro partes porcentualmente iguales.

Hay tres cuartiles denotados usualmente Q_1 , Q_2 , Q_3 . El segundo cuartil es precisamente la mediana. El primer cuartil, es el valor en el cual o por debajo del cual queda un cuarto (25%) de todos los valores de la sucesión (ordenada); el tercer cuartil, es el valor en el cual o por debajo del cual quedan las tres cuartas partes (75%) de los datos.

Los datos agrupados son los siguientes:

$$Q_k = L_k + \frac{k\left(\frac{n}{4}\right) - F_k}{f_k} * c$$

La fórmula para el cálculo de los cuartiles cuando se trata de datos agrupados es la siguiente:

$k= 1,2,3$

Donde :

Lk = Límite real inferior de la clase del cuartil k

n = Número de datos

Fk = Frecuencia acumulada de la clase que antecede a la clase del cuartil k.

fk = Frecuencia de la clase del cuartil k

c = Longitud del intervalo de la clase del cuartil k.

Para calcular los cuartiles individuales las formulas son las siguientes que te muestro :

Fórmula de Q1, para series de Datos agrupados:

$$Q_1 = l_i + \frac{P - f_{a-1}}{f_1} * l_c, P = \frac{n}{4}$$

Donde

L1 = límite inferior de la clase que lo contiene

P = valor que representa la posición de la medida

f1 = la frecuencia de la clase que contiene la medida solicitada.

Fa-1 = frecuencia acumulada anterior a la que contiene la medida solicitada.

lc = intervalo de clase.

Q2 (coincide, es idéntico o similar a la mediana, Q2 = Md), es el menor valor que es mayor que la mitad de los datos, es decir el 50% de las observaciones son mayores que la mediana y el 50% son menores.

En el segundo cuartil la fórmula es la siguiente:

$$Q_2 = l_i + \frac{P - f_{a-1}}{f_1} * l_c, P = \frac{2n}{4}$$

Donde

L1 = límite inferior de la clase que lo contiene

P = valor que representa la posición de la medida

f1 = la frecuencia de la clase que contiene la medida solicitada.

Fa-1 = frecuencia acumulada anterior a la que contiene la medida solicitada.

lc = intervalo de clase

El tercer cuartil Q3, es el menor valor que es mayor que tres cuartas partes de los datos, es decir aquel valor de la variable que supera al 75% y es superado por el 25% de las observaciones.

La fórmula del Q3 es la siguiente:

$$Q_3 = l_i + \frac{P - f_{a-1}}{f_1} * l_c, P = \frac{3n}{4}$$

Donde

L1 = límite inferior de la clase que lo contiene

P = valor que representa la posición de la medida

f1 = la frecuencia de la clase que contiene la medida solicitada.

Fa-1 = frecuencia acumulada anterior a la que contiene la medida solicitada.

lc = intervalo de clase

Ahora para los **datos NO agrupados son los siguientes:**

Si se tienen una serie de valores X1, X2, X3 ... Xn, se localiza mediante las siguientes

Fórmulas:

- El primer cuartil:

Cuando n es par:

$$\frac{1 * n}{4}$$

Cuando n es impar:

$$\frac{1 (n+1)}{4}$$

Para el tercer cuartil

Cuando n es par:

$$\frac{3 * 4}{4}$$

Cuando n es impar:

$$\frac{3 (n+1)}{4}$$

6.-DECILES

Los deciles son ciertos números que dividen la sucesión de datos ordenados en diez partes porcentualmente iguales. Son los nueve valores que dividen al conjunto de datos ordenados en diez partes iguales, son también un caso particular de los percentiles. Los deciles se denotan D1, D2, ..., D9, que se leen primer decil, segundo decil, etc. Los deciles, al igual que los cuartiles, son ampliamente utilizados para fijar el aprovechamiento académico.

Sus **datos agrupados** son los siguientes:

Para datos agrupados los deciles se calculan mediante la fórmula.

$$D_k = L_k + \frac{k\left(\frac{n}{10}\right) - F_k}{f_k} * c$$

k= 1,2,3,... 9

Donde

L_k = Límite real inferior de la clase del decil k

n = Número de datos

F_k = Frecuencia acumulada de la clase que antecede a la clase del decil k.

f_k = Frecuencia de la clase del decil k

c = Longitud del intervalo de la clase del decil k

Otra fórmula para calcular los deciles:

El cuarto decil es aquel valor de la variable que supera al 40%, de las observaciones y es superado por el 60% de las observaciones.

$$D_4 = l_1 + \frac{P - f_{a-1} * l_c}{f_1} \quad P = \frac{4n}{10}$$

El quinto decil corresponde a la mediana.

$$D_5 = l_i + \frac{P - f_{a-1} * I_c}{f_1} \quad P = \frac{5n}{10}$$

El noveno decil supera al 90% y es superado por el 10% restante.

$$P = \frac{9n}{10}$$

$$D_9 = l_i + \frac{P - f_{a-1} * I_c}{f_1}$$

Donde (para todos):

L1 = límite inferior de la clase que lo contiene

P = valor que representa la posición de la medida

f1 = la frecuencia de la clase que contiene la medida solicitada.

Fa-1 = frecuencia acumulada anterior a la que contiene la medida solicitada.

Ic = intervalo de clase.

Fórmulas Datos No Agrupados

Si se tienen una serie de valores X1, X2, X3 ... Xn, se localiza mediante las siguientes

Fórmulas:

Cuando n es par

$$\frac{A * n}{10}$$

Cuando n es impar

$$\frac{A(n+1)}{10}$$

Siendo A el número del decil.

7.-CENTILES O PERCENTILES

Los percentiles son, tal vez, las medidas más utilizadas para propósitos de ubicación o clasificación de las personas cuando atienden características tales como peso, estatura, etc. Los percentiles son ciertos números que dividen la sucesión de datos ordenados en cien partes porcentualmente iguales. Estos son los 99 valores que dividen en cien partes iguales el conjunto de datos ordenados. Los percentiles (P1, P2,... P99), leídos primer percentil,..., percentil 99.

Datos Agrupados

Cuando los datos están agrupados en una tabla de frecuencias, se calculan mediante la fórmula:

$$P_k = L_k + \frac{k \left(\frac{n}{100} \right) - F_k}{f_k} * c$$

k= 1,2,3,... 99

Donde

Lk = Límite real inferior de la clase del decil k

n = Número de datos

Fk = Frecuencia acumulada de la clase que antecede a la clase del decil k.

fk = Frecuencia de la clase del decil k

c = Longitud del intervalo de la clase del decil k

Otra forma para calcular los percentiles es:

Primer percentil, que supera al uno por ciento de los valores y es superado por el noventa y nueve por ciento restante.

$$P = \frac{1n}{100}$$

$$P_1 = l_i + \frac{P - f_{a-1} * I_c}{f_1}$$

El 60 percentil, es aquel valor de la variable que supera al 60% de las observaciones y es superado por el 40% de las observaciones.

$$P_{60} = l_i + \frac{P - f_{a-1}}{f_1} * I_c \quad P = \frac{60n}{100}$$

$$P_{99} = l_i + \frac{P - f_{a-1}}{f_1} * I_c \quad P = \frac{99n}{100}$$

El percentil 99 supera 99% de los datos y es superado a su vez por el 1% restante.

Fórmulas Datos No Agrupados son las siguientes.

Si se tienen una serie de valores $X_1, X_2, X_3 \dots X_n$, se localiza mediante las siguientes

Fórmulas:

Para los percentiles, cuando n es par:

$$\frac{A * n}{10}$$

10

Cuando n es impar: $\frac{A(n+1)}{100}$

Siendo A, el número del percentil.

100

3. ejemplo

Salarios	No. De	fa
(l. De Clases)	Empleados (fl)	
200-299	85	85
300-299	90	175
400-499	120	295
500-599	70	365
600-699	62	427
700-800	36	463

Determinación del primer cuartil, el séptimo decil y el 30 percentil, de la siguiente tabla:

Como son datos agrupados, se utiliza la fórmula

$$P = l_i + \frac{P - f_{e-1} * I_c}{f_1}$$

Siendo las siguientes:

La posición del primer cuartil.

$$P = \frac{n}{4}$$

4

$$P = \frac{7n}{10}$$

10

La posición del 7 decil.

$$P = \frac{30n}{100}$$

100

La posición del percentil 30.

Son las siguientes:

El primer cuartil:

$$\frac{463}{4} = 115.5$$

4

$$115.5 - 85 = 30.75$$

$$Li = 300, lc = 100, fi = 90$$

$$Q1 = 300 + \frac{30.75}{90} * 100 = 334$$

El 7 decil:

$$\frac{7(463)}{100} = \frac{3241}{100} = 324.1$$

Posición:

$$324.1 - 295 = 29.1$$

$$Li = 500, fi = 70$$

$$D_7 = 500 + \frac{29.1}{70} * 100 = 541.57$$

El percentil 30

$$\frac{30(463)}{100} = \frac{13890}{100} = 138.9$$

$$138.9 - 85 = 53.9$$

$$fi = 90$$

$$P_{30} = 300 + \frac{53.9}{90} * 100 = 359.88$$

Estos resultados nos indican que el 25% de los empleados ganan salarios por debajo de \$ 334; que bajo 541.57 gana el 57% de los empleados y sobre \$359.88, gana el 70% de los empleados.

Coeficiente de Pearson

La covariación es el grado de concordancia de las posiciones relativas de los datos de dos variables. En consecuencia el coeficiente de correlación de Pearson opera con puntuaciones tipificadas (que miden posiciones relativas) y se define:

$$r_{xy} = \frac{\sum z_x z_y}{N}$$

El fundamento del coeficiente de Pearson es el siguiente:

Cuanto más intensa sea la concordancia (en sentido directo o inverso) de las posiciones relativas de los datos en las dos variables, el producto del numerador toma mayor valor (en sentido absoluto). Si la concordancia es exacta, el numerador es igual a N (o a -N), y el índice toma un valor igual a 1 (o -1).

EJEMPLO (Máxima covariación positiva)

X	Y	z_x	z_y
2	4	-1.41	-1.41
3	5	-0.71	-0.71
4	6	-0.00	-0.00
5	7	0.71	0.71
6	8	1.41	1.41

Observa que los datos tipificados (expresados como puntuaciones z) en las dos columnas de la derecha tienen los mismos valores en ambas variables, dado que las posiciones relativas son las mismas en las variables X e Y.

Si obtenemos los productos de los valores tipificados para cada caso, el resultado es:

X	Y	z_x	z_y	$z_x z_y$
2	4	-1.41	-1.41	2
3	5	-0.71	-0.71	0.5
4	6	0.00	0.00	0
5	7	0.71	0.71	0.5
6	8	1.41	1.41	2
				$\Sigma = 5$

El cociente de dividir la suma de productos (5) por N (hay que tener en cuenta que N es el número de casos, NO el número de datos) es igual a 1:

$$= \frac{\sum z_x z_y}{N} = \frac{5}{5} = 1$$

EJEMPLO 2 (COVARIACIÓN POSITIVA DE ALTA INTENSIDAD)

X	Y	z_x	z_y	$z_x z_y$
2	5	-1.41	-0.71	1
3	4	-0.71	-1.41	1
4	6	0.00	0.00	0
5	8	0.71	1.41	1
6	7	1.41	0.71	1
				$\Sigma = 4$

Por lo consiguiente sería:

$$= \frac{\sum z_x z_y}{N} = \frac{4}{5} = 0.8$$

EJEMPLO 3 (Ausencia de covariación)

X	Y	z_x	z_y	$z_x z_y$
2	7	-1.41	0.71	-1
3	8	-0.71	1.41	-1
4	6	0.00	0.00	0
5	4	0.71	-1.41	-1
6	5	1.41	-0.71	-1
				$\Sigma = -4$

$$= \frac{\sum z_x z_y}{N} = \frac{-4}{5} = -0.8$$

EJEMPLO 4 (Covariación negativa de alta intensidad)

X	Y	z_x	z_y	$z_x z_y$
2	7	-1.41	0.71	-1
3	8	-0.71	1.41	-1
4	6	0.00	0.00	0
5	4	0.71	-1.41	-1
6	5	1.41	-0.71	-1
				$\Sigma = -4$

$$= \frac{\sum z_x z_y}{N} = \frac{-4}{5} = -0.8$$

EJEMPLO 5 (Máxima covariación negativa)

X	Y	z_x	z_y	$z_x z_y$
2	8	-1.41	1.41	-2.0
3	7	-0.71	0.71	-0.5
4	6	0.00	0.00	0.0
5	5	0.71	-0.71	-0.5
6	4	1.41	-1.41	-2.0
				$\Sigma = -5$

$$= \frac{\sum z_x z_y}{N-1} = \frac{-5}{5} = -1$$

El valor de la correlación es igual a 1 o -1 si la covariación es de intensidad máxima, y se va acercando hacia el 0 cuanto más pequeña sea la intensidad de la covariación. Además, el índice tiene signo positivo cuando la covariación es directa y negativo cuando es inversa.

conclusión

Para concluir este tema se vio muchos tipos de subtemas en el los cuales fueron los cuartiles, cuantiles, el coeficiente de Pearson entre otros pero cada uno tiene su propia función en la estadística y en este ensayo se dan algunos ejemplos de cada tema que se tocó .

Es muy importante cada tema porque con ellos puedes hacer de la estadística un buen uso para saber el valor o datos que se dan en la sociedad.