UNIVERSIDAD NACIONAL

JOSÉ FAUSTINO SÁNCHEZ CARRIÓN

FACULTAD DE CIENCIAS E INGENIERÍA

INSTITUTO DE INVESTIGACIÓN
TEXTO UNIVERSITARIO

ESTADÍSTICA DESCRIPTIVA

AUTORES

Mg. MIRTHA SUSSAN TREJO LÓPEZ
Mg. YOLANDA MARIANELA CASTAÑEDA CARRIÓN
Mo. COSME ULISES VALDERDE FLORES

COLABORADOR
Ing. CARLOS MANUEL CRUZ CASTAÑEDA

Huacho, Diciembre 2010

INTRODUCCIÓN

En la mayoría de las Facultades de la Universidad se imparten cursos de Estadística y con carácter obligatorio. La comprensión de las ideas estadísticas y su posible aplicación es en la actualidad esencial en el ser humano y que debe ser entendida como una herramienta útil para la toma de decisiones en la vida cotidiana, laboral y empresarial.

Nuestro propósito principal al realizar este texto es lograr una exposición cuidadosa y legible donde los alumnos descubrirán por qué y en que sentido cada método de análisis de datos, en particular, es apropiado en cada situación específica.

Este texto está orientado a servir de base para la formación científica del alumno que requieran del conocimiento de las diversas técnicas de la Estadística.

Se motiva a los alumnos a estudiar información reciente e ideas nuevas, mediante una exposición clara con aplicaciones prácticas en áreas seleccionadas del conocimiento y ejemplos que se utiliza para transmitir las ideas esenciales de la Estadística Descriptiva.

Cada tema está motivado con una aplicación práctica seguida de una generalización o regla, luego se resuelve una aplicación para ampliar las explicaciones del método elegido.

En el desarrollo del texto los fundamentos matemáticos son de completo entendimiento, es preferible entender los conceptos y principios estadísticos, que memorizar una lista de fórmulas y términos sin ser capaz de aplicarlos o de comprender sus aplicaciones.

Los Autores

ÍNDICE GENERAL

INTRODUCCIÓN

CAPÍTULO I. MEDIDAS DE TENDENCIA CENTRAL

1.1. Definición 4

1.2. **La Media Aritmética** 6

- 1.2.1. Cálculo de la Media Aritmética 6
- 1.2.2. Propiedades de la Media Aritmética 10
- 1.2.3. Métodos abreviados de Cálculo de la Media Aritmética 12
- 1.2.4. Media Aritmética a partir de submuestras 14
- 1.2.5. Ventajas de la Media Aritmética 16
- 1.2.6. Desventajas de la Media Aritmética 16

1.3. Mediana 17

- 1.3.1. Cálculo de la Mediana 17
- 1.3.2. Propiedades de la Mediana 24
- 1.3.3. Ventajas de la Mediana 24
- 1.3.4. Desventajas de la Mediana 24

1.4. Moda 25

- 1.4.1. Cálculo de la Moda 25
- 1.4.2. Relación entre la Media, la Mediana y la Moda 28
- 1.4.3. Ventajas de la Moda 29
- 1.4.4. Desventajas de la Moda 29

1.5. La Media Geométrica 29

- 1.5.1. Cálculo de la Media Geométrica 30
- 1.5.2. Ventajas de la Media Geométrica 31
- 1.5.3. Desventajas de la Media Geométrica 31
- 1.5.4. Aplicaciones de la Media Geométrica 31

1.6. La Media Armónica 32

- 1.5.1. Cálculo de la Media Armónica 32
- 1.5.2. Ventajas de la Media Armónica 33
- 1.5.3. Desventajas de la Media Armónica 34
- 1.5.4. Aplicaciones de la Media Armónica 34

1.7. Cuartiles 34

- 1.5.1. Representación de los Cuartiles 34
- 1.5.2. Cálculo de los Cuartiles 35

1.8. Deciles º 38

- 1.5.1. Representación de los Deciles 38
- 1.5.2. Cálculo de los Deciles 39

1.9. Percentiles 39
1.5.1. Representación de los Percentiles 39
1.5.2. Cálculo de los Percentiles 39

CAPÍTULO II. MEDIDAS DE DISPERSIÓN

2.1. Importancia de las Medidas de Dispersión 43
2.2. Amplitud de Variación 44
2.2.1. Cálculo de la Amplitud de Variación 44
2.2.2. Ventajas de la Amplitud de Variación 45
2.2.3. Desventajas de la Amplitud de Variación 45
2.3. Desviación Media 45
2.3.1. Cálculo de la Desviación Media 45
2.3.2. Ventajas y desventajas de la Desviación Media 47
2.4. Varianza 48
2.4.1. Cálculo de la Varianza 48
2.4.2. Propiedades de la Varianza 50
2.4.3. Componentes de la Varianza 51
2.4.4. Métodos Abreviados de Cálculo de la Varianza 53
2.5. La Desviación Estándar 54
2.5.1. Cálculo de la Desviación Estándar 55
2.5.2. Propiedades de la Desviación Estándar 56
2.5.3. Ventaja de la Desviación Estándar 56
2.5.4. Desventajas de la Desviación Estándar 57
2.5.5. Aplicaciones de la Desviación Estándar 57

CAPÍTULO III. MEDIDAS DE ASIMETRÍA

3.1. Simetría 60
3.2. Asimetría Positiva 60
3.3. Asimetría Negativa 61
3.4. Coeficiente de Asimetría 61
3.5. Interpretación de los Coeficientes de Asimetría 61

CAPÍTULO IV. MEDIDAS DE APUNTAMIENTO

4.1. Coeficiente de Kurtosis 64

BIBLIOGRAFÍA

1. ÁVILA ACOSTA, Roberto (2005) Estadística Elemental. Estudios y Ediciones R.A. Lima Perú.
2. ELORZA, Haroldo (2008) Estadística para Ciencias Sociales del Comportamiento y de la Salud. CENGACE. 3ª Edición.
3. HOPKINS / GLASS (2005) Estadística Básica para las Ciencias Sociales y del Comportamiento. Prentice Hall Hispano americana S. A. México. 3ª Edición.
4. MASON / LIND / MARCHAL (2007) Estadística para Administración y Economía. Alfa omega Grupo Editor, S.A. de C. V. 1ºª Edición.
5. MITACC MEZA, Máximo (2006) Tópicos de Estadística Descriptiva y Probabilidad. Editorial San Marcos Lima -Perú.
6. MOYA CALDERÓN, Rufino (2010) Estadística Descriptiva Editorial San Marcos Lima Perú.
7. PÉREZ LEGOAS, Luís (2008) Estadística Básica para Ciencias Sociales y Educación Editorial San Marcos Lima Perú.
8. VÉLIZ CAPUÑAY, Carlos (2007) Estadística. Aplicaciones. Impreso en PERÚ OFFSET.
9. WEIMER Richard C. (2006) Estadística. Compañía Editorial S.A de C. V. México

CAPÍTULO I

MEDIDAS DE TENDENCIA CENTRAL

1.1. Definición

Las medidas de tendencia central son aquellas cuyos valores tienden a ocupar posiciones centrales o intermedias entre el menor y el mayor valor del conjunto de datos, a partir de la cual se encuentran estas medidas, brindando de alguna forma, información sobre el centro de la distribución.

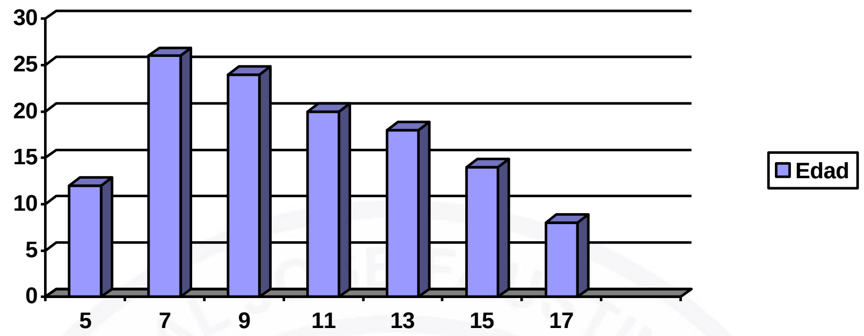
Supongamos que se eligen tres estudiantes de niveles primaria, secundaria y superior, cuya distribución por edades ha permitido obtener los histogramas del Gráfico 1,1(a), (b) y (c) observando un desplazamiento de las distribuciones hacia la derecha a medida que se avanza el nivel educativo de cada grupo; la diferencia entre los tres histogramas es sólo el cambio de posición o localización a lo largo del eje horizontal o valor de la variable (edad).

En cada uno de los tres casos los valores observados se agrupan alrededor de cierto "valor central" o "valor medio". Estos valores centrales, por su desplazamiento en el mismo sentido y magnitud, se considera como números que describen la posición de la distribución de frecuencias, y se definen como medidas de posición o de tendencia central.

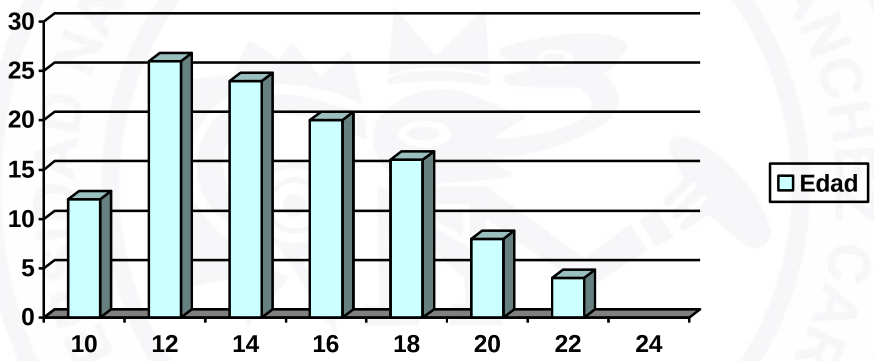
Las medidas de tendencia central más importantes y muy usadas son: la media aritmética, la mediana, la moda, la media geométrica, la media armónica, los cuantiles, los deciles y los percentiles.

Gráfico Nº 1.1

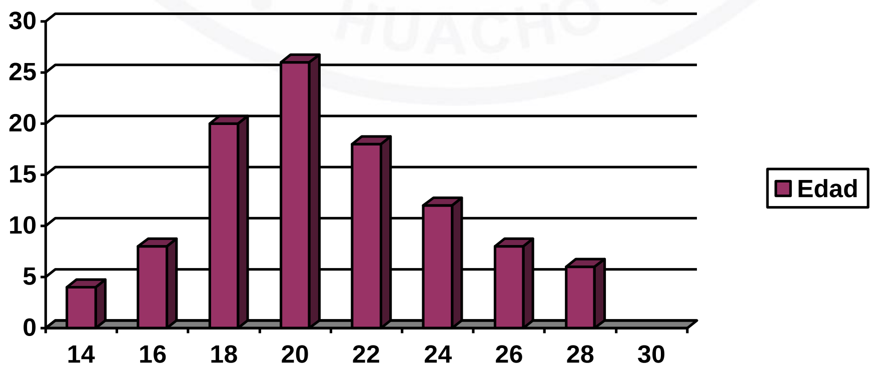
DISTRIBUCIÓN DE TRES CONJUNTOS DE ESTUDIANTES EN NIVELES PRIMARIA, SECUNDARIA Y SUPERIOR, SEGÚN LA EDAD



(a) Nivel Primaria



(b) Nivel Secundaria



(c) Nivel Superior

1.2. La Media Aritmética

Es la medida de tendencia central más conocida, familiar a todos nosotros, y de mayor uso, también fácil de calcular, ya sea de datos tabulados como no tabulados.

Cuando se habla de "media" en la práctica se entiende "media aritmética" en este sentido hablamos de salario medio, número medio de accidentes, rendimiento medio de un obrero, etc.

Para una variable x_i la media se simboliza como $\overline{x}$, M(x)

1.2.1. Cálculo de la Media Aritmética

a) La Media de Datos no Agrupados

Para la media de datos no agrupados es el cociente de la suma de valores de la variable (x_i) dividido entre el número de observaciones (n), se denomina Media Aritmética Simple.

$$M(x) = \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Ejemplo. Supongamos que los puntajes obtenidos en 5 exámenes parciales de estadística son $x_1 = 13$; $x_2 = 10$; $x_3 = 14$; $x_4 = 11$; $x_5 = 10$. ¿Cuál es la nota promedio?

Solución.

La nota promedio será

$$M(x) = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{13+10+14+11+10}{5} = \frac{58}{5} = 11{,}6$$

M(x) = 12 puntos

Ejemplo. Una persona que trabaja en forma independiente gana el primer mes S/ 1 500; el segundo mes S/ 1 200 y el tercer mes S/ 1 800. ¿Cuánto gana en promedio mensual?.

Solución.

En este caso $x_1 = 1500$; $x_2 = 1200$; $x_3 = 1800$

$$M(x) = \frac{\sum_{i=1}^{3} x_i}{3} = \frac{x_1 + x_2 + x_3}{3} = \frac{1500+1200+1800}{3} = \frac{4500}{3}$$

M(x) = S/. 1 500 por mes

b) La media de Datos Agrupados en tablas de Frecuencia

Los datos que usamos frecuentemente se agrupan y se presentan en forma de distribución con intervalos.

En general, es imposible recuperar los datos brutos originales, y se tiene que trabajar con datos elaborados por alguna identidad. De esta forma si queremos un valor típico que representa los datos, debemos estimarlos basándonos en la distribución de frecuencias; pero también se presentan que los datos se agrupan en tablas sin intervalos y se denomina **Media Aritmética Ponderada**, $M(x) = \frac{\sum_{i=1}^{m} x_i . n_i}{n}$ es un buen indicador para calcular costos medios de producción, promediar números índices y tasas de cambio.

b.1. Tablas sin intervalo

Ejemplo.

Durante 30 días se ha observado el número de pasajeros que viajan de Lima a Iquitos, siendo estos los resultados siguientes:

36	30	34	38	39	40	30	35	36	37
30	39	38	31	37	32	39	38	30	38
38	39	40	40	38	33	31	35	35	32

Determinar la media aritmética.

Solución

Tabla 1. 1

CÁLCULO DE LA MEDIA O NÚMERO PROMEDIO DE PERSONAS QUE VIAJAN DE LIMA A IQUITOS POR DÍA

x_I	n_i	$x_I n_i$
30	4	120
31	2	62
32	2	64
33	1	33
34	1	34
35	3	105
36	2	72
37	2	74
38	6	228
39	4	156
40	3	120
TOTAL	n = 30	1068

Aplicando la fórmula se tiene:

$$M(x)=\frac{\sum_{i=1}^{11} x_i.n_i}{30}=\frac{1068}{30}=35{,}6$$

M(x) = 36 pasajeros diarios.

Ejemplo.

Se realiza una encuesta a un conjunto de 40 familias sobre el número de hijos. Los resultados se muestran a continuación:

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

Determinar la media aritmética.

Solución

Tabla 1. 2.

CÁLCULO DE LA MEDIA O NÚMERO DE HIJOS POR FAMILIA

x_I	n_i	$x_I n_i$
0	4	0
1	5	5
2	9	18
3	6	18
4	8	32
5	8	40
TOTAL	n = 40	113

Usando la fórmula se tiene:

$$M(x)=\frac{\sum_{i=1}^{6} x_i.n_i}{40}=\frac{113}{40}=2{,}8$$

M(x) = 3 hijos por familia.

b.2. Tablas con intervalos

Ejemplo.

El jefe de control de calidad de SIDERPERÚ revela los pesos en kilogramos (kg) de 50 lingotes de acero.

Peso (kg)	Nº de lingotes
90 – 94	10
94 – 98	5
98 – 102	15
102 – 106	12
106 – 110	8

Obtener la media aritmética.

Solución.

Tabla 1.3

CÁLCULO DEL PESO PROMEDIO DE 50 LINGOTES DE ACERO

Peso	Nº de lingote n_i	Marca de clase x_i	Pesos Ponderado $x_i.n_i$
90 – 94	10	92	920
94 – 98	5	96	480
98 – 102	15	100	1500
102 – 106	12	104	1248
106 – 110	8	108	864
Total	n = 50	m = 5	5012

Utilizando la fórmula se tiene:

$$M(x)=\frac{\sum_{i=1}^{5} x_i.n_i}{50}=\frac{5012}{50}=100{,}24$$

M(x) = 100 kilogramos por lingote..

Ejemplo.

El jefe de remuneraciones de una empresa revela los sueldos mensuales en nuevos soles de los trabajadores

Sueldos mensuales	Nº de trabajadores
800 – 1000	20
1000 – 1200	10
1200 – 1400	15
1400 – 1600	17
1600 – 1800	18

Determinar la media aritmética.

Solución.

Tabla 1,4.

CÁLCULO DEL SUELDO MENSUAL PROMEDIO DE 80 TRABAJADORES

Sueldos mensuales	Nº de trabajadores n_i	Marca de clase x_i	Sueldos mensuales ponderados $x_i.n_i$
800 – 1000	20	900	18 000
1000 – 1200	10	1100	11 000
1200 – 1400	15	1300	19 500
1400 – 1600	17	1500	25 500
1600 – 1800	18	1700	30600
Total	n = 80	m = 5	104 600

Por la fórmula se tiene:

$$M(x) = \frac{\sum_{i=1}^{5} x_i.n_i}{80} = \frac{104600}{80} = 1307{,}50$$

M(x) = 1307,50 por trabajador.

1.2.2. Propiedades de la Media Aritmética

a. La media aritmética de una constante es igual a la constante misma

M(k) = k

$$M(k) = \frac{\sum_{i=1}^{n} k}{n} = \frac{k+k+...+k}{n} = \frac{nk}{n} = k$$

b. La media del producto de una constante por una variable, es igual al producto de la constante por la media de la variable.

M(kx) = k M(x); donde k es una constante

$$M(kx) = \frac{\sum_{i=1}^{n} kx_i}{n} = \frac{kx_1 + kx_2 + ... + kx_n}{n} = \frac{k(x_1 + x_2 + ... + x_n)}{n} = \frac{k.M(x)}{n}$$

c. La media de la suma de dos o más variables es igual a la suma de las medias de cada una de las variables.

M(x + y) = M(x) + M (y)

$$M(x+y) = \frac{\sum (x_i + y_i)}{n} = \frac{(x_1 + y_1) + ... + (x_n + y_n)}{n} = \frac{\sum x_i}{n} + \frac{\sum y_i}{n} = M(x) + M(y)$$

d. La media de una variable más una constante es igual a la media de la variable más la constante.

M(x + k) = M(x) + k

$$M(x+k) = \frac{\sum(x+k)}{n} = \frac{\sum x}{n} + \frac{\sum k}{n} = M(x) + k$$

e. La suma de las desviaciones con respecto a la media aritmética es igual a cero, cualquiera que sea la distribución.

$$\sum_{i=1}^{m}(x_i - \bar{x})n_i = 0$$

$$\sum_{i=1}^{m}(x_i - \bar{x})n_i = \sum(x_i n_i - \bar{x}n_i) = \sum x_i n_i - \bar{x}\sum n_i = \sum x_i n_i - n\bar{x} = 0$$

$$\sum_{i=1}^{m}(x_i - \bar{x})n_i = n\bar{x} - n\bar{x} = 0$$

Ejemplo.

Para los 50 lingotes representados en la Tabla 1,3 se tiene la comprobación numérica:

Tabla 1,5

Intervalos	x_i	$x_i - \bar{x}$	n_i	$(x_i - \bar{x})n_i$
90 – 94	92	92-100,24	10	- 82,4
94 – 98	96	96-100,24	5	- 21.2
98 – 102	100	100-100,24	15	- 3,6
102 – 106	104	104-100,24	12	45,12
106 – 110	108	108-100,24	8	62,08
			50	0
				$\sum(x_i - \bar{x})n_i$

La media aritmética se puede utilizar para estimar la cantidad total de una población.

Total = N $\bar{x}$

Donde: N = Tamaño de la población. $\bar{x}$ = Media aritmética de la muestra.

Ejemplo.

Determinar la media del conjunto de observaciones $x_1, x_2, ..., x_6$ a las cuales se les ha restado 4, resultando los valores 3, 1, 0, 2, 4 y 5.

Solución.

Se tiene $y_1 = 3$, $y_2 = 1$, $y_3 = 0$, $y_4 = 2$, $y_5 = 4$, $y_6 = 5$

La ecuación de transformación está dado por:

$y_i = x_i - 4$; i = 1, 2, 3, 4, 5, 6

Luego:

$y = M(y) = M(x) - 4$ donde M(x) = M(y) + 4

Pero $M(y) = \frac{1}{6}\sum_{i=1}^{6} y_i = \frac{1}{6}(3+1+0+2+4+5) = \frac{5}{2}$

Finalmente, $M(x) = M(y) + 4 = \frac{5}{2} + 4 = \frac{13}{2}$

Ejemplo.

Considérese que los lingotes de acero son vendidos a razón de 8 unidades monetarias (u.m) por kilogramo y que la empresa paga 0,80 unidades monetarias por gastos de manejo, por cada lingote vendido. Calcular la cantidad promedio recibida por lingote.

Solución

La ecuación de transformación, que representa la cantidad recibida por lingotes es $z_i = 8x_i - 0{,}8$ con i = 1,2,…,5 usando los datos tabulados. Es decir, x_i es la marca de clase del intervalo i.

Por tanto, M (z) = 8 M(x) – 0,8 = 8 (100,24) – 0,8 = 801,12 u.m / lingote

Ejemplo

Un estudio económico muestra que el gasto en alimentación de una familia de cuatro personas es en promedio S/ 1 200 mensuales, determinar la media del gasto diario en alimentos.

Solución

Sean x_i = gasto mensual de la familia. $\bar{x} = S/1200$

y_i = gasto diario de la familia

La relación entre x_i e y_i es x_i = 30 y_i . Por lo tanto, el gasto de alimentos de la familia de cuatro personas es en promedio

$$\bar{y} = \frac{\bar{x}}{30} = \frac{1200}{30} = 40$$ Nuevos soles diarios.

1.2.3. Métodos abreviados de Cálculo de la Media Aritmética

Método 1

Se trata de reducir la magnitud de la variable, transformando las marcas de clases $x_1, x_2, \ldots, x_n$ en desviaciones d_i, con i = 1, 2, …, m respecto de un origen de trabajo arbitrariamente elegido, es decir;

$d_i = y_i - O_t$ para i = 1, 2, …; m

Luego, se determina la media aritmética de la nueva variable desviación “d”.

Criterios para elegir el Origen de Trabajo

Se consideran los siguientes criterios:

a. Si m es impar, O_i debe ser el valor central de las x_i .

b. Si m es par, O_i puede ser uno de los valores centrales el de mayor frecuencia

Entonces el método consiste en:

1. Elegir un origen de trabajo O_i .
2. Determinar las desviaciones de las x_i con i = 1, 2, …, m respecto a O_i ;

   $d_i = x_i - O_t$
3. Se determina la media aritmética de la variable desviación “d”

$$M(d) = \frac{1}{n}\sum_{i=1}^{m} n_i.d_i = \frac{1}{n}\sum_{i=1}^{m} n_i.x_i - \frac{O_t}{n}\sum_{i=1}^{m} n_i = M(x) - O_t$$

De donde $M(x) = M(d) - O_t$

Ejemplo

Determinar la media aritmética por el método abreviado para los datos agrupados en la Tabla 1.3

Solución

Tabla Nº 1.6

x_i	n_i	$d_i = x_i - O_t$	$n_i.d_i$
92	10	-8	-80
96	5	-4	-20
100	15	0	0
104	12	4	48
108	8	8	64
	50		12

Elegimos O_t $\rightarrow$ **100**

$$M(d) = \frac{1}{50}\sum_{i=1}^{5} n_i.d_i = \frac{12}{50} = 0{,}24$$

Por lo tanto,

$$M(x) = M(d) + O_t = 0{,}24 + 100 = 100{,}24$$

Método 2. Método Codificado

Si la amplitud es constante, es decir, $x_2^{'} - x_1^{'} = x_3^{'} - x_2^{'} = ... = x_m^{'} - x_{m-1}^{'} = c$ en el caso descrito cuando las x_i están espaciadas a una distancia constante. Entonces:

1. Se elige el origen de trabajo O_t siguiendo los criterios mencionados.

2. Se toma la amplitud constante (común) de los intervalos como unidad para medir las desviaciones, respecto al origen de trabajo O_t. Es decir, $\mu_i = \frac{x_i - O_t}{c}$, i = 1, 2, 3, …, m
3. Se calcula la media aritmética de la variable "μ".

$$M(\mu) = \frac{1}{n}\sum_{i=1}^{m} n_i.d_i = \frac{1}{nc}\sum_{i=1}^{m} n_i(y_i - O_t) = \frac{1}{c}(M(x) - \frac{O_t}{n}\sum_{i=1}^{m} n_i = \frac{1}{c}\left(M(x) - O_t\right)$$

De donde M(x) = c M (μ) + O_t

Ejemplo

Los sueldos mensuales de 80 trabajadores representados en la Tabla 1. 4, determinar la media utilizando el método codificado:

1. Se construye la tabla de frecuencia absoluta

Tabla 1.7

x_i	n_i	$\mu_i = \frac{(x_i - 1300)}{200}$	$n_i \cdot \mu_i$
900	20	-2	- 40
1100	10	-1	- 10
1300	15	0	0
1500	17	1	17
1700	18	2	36
	80		3

2. Como m = 5, elegimos $O_t \rightarrow$ 1300, el valor central de los x_i .

3. La diferencia entre las x_i es constante e igual a 200. Entonces $\mu_i = \frac{x_i - 1300}{200}$

4 .Cálculo de la media de la variable "μ".

$$M(\mu) = \frac{1}{80}\sum_{i=1}^{5} \mu_i.n_i = \frac{3}{80} = 0{,}0375$$

Por tanto:

M(x) = c M (μ) + O_t = 200 (0,0375) + 1 300 = S/ 1 307, 50

1.2.4. Media aritmética a partir de sub muestras

Este procedimiento de cálculo se justifica por que en muchas situaciones hay la necesidad de obtener muestras diferentes de una población, ya sea por razones técnicas o por disponibilidad de información. Un ejemplo del primer caso:

Se desea estimar el diámetro medio de las tuercas producidas por una máquina es lógico que tenemos que registrar observaciones en días diferentes.

Ejemplo del segundo caso: se desea estimar el promedio del rendimiento académico de la población estudiantil del sistema universitario peruano; tomaremos una muestra de la universidad A, otra de la B, etc; y luego obtendremos el promedio deseado.

Supongamos que de una población (o de dos poblaciones diferentes) se obtienen dos muestras de tamaño n_1 y n_2 respectivamente. Sean $\overline{x}_1$ y $\overline{x}_2$ las medias aritméticas de las muestras, entonces la media asociada a las n_1 + n_2 observaciones esta dada por:

$$\overline{x} = \frac{n_1\overline{x}_1 + n_2.\overline{x}_2}{n_1 + n_2}$$

Ejemplo

Un examen fue rendido por 80 alumnos de una sección A y 90 de una sección B. En la sección A se obtuvo un promedio de 11.8 y en sección B un promedio de 12.5. Obtener la media aritmética de las notas de todos los estudiantes que rindieron el examen.

Solución

Sea n_A = número de alumnos de la sección A

n_B = número de alumnos de la sección B

Sean $\overline{x}_A$ = media aritmética de las notas de la sección A

$\overline{x}_B$ = media aritmética de las notas de la sección B

Es decir, n_A = 80; n_B = 90; $\overline{x}_a$ = 11,8; $\overline{x}_B$ = 12,5

Sea $\overline{x}$ = la media aritmética de las notas de todos los alumnos que rindieron la prueba.

Entonces;

$$\overline{x} = \frac{n_A\overline{x}_A + n_B.\overline{x}_B}{n_A + n_B} = \frac{80(11{,}8) + 90(12{,}5)}{80 + 90} = 12{,}17$$

Es decir, el promedio de las dos secciones es $\overline{x}$ = 12,17

Ejemplo.

La empresa A tiene 100 empleados, con un sueldo promedio mensual de S/ 1 200. La empresa B tiene 200 empleados con un sueldo promedio mensual de S/ 1 000.

a) ¿Cuál es el sueldo promedio mensual de las dos empresas en conjunto?

b) Si a las dos empresas se agrega una tercera con 50 empleados y un sueldo promedio mensual de S/. 900, ¿cuál es el sueldo promedio para las tres empresas en conjunto?

Solución

a) n_A = 1 00; $\overline{x}_A$ = S/. 1 200; n_B = 200; $\overline{x}_B$ = S/. 1 000

Si $\overline{x}_{A+B}$ Es el promedio mensual de las dos empresas en conjunto. Entonces,

$$\bar{x}_{A+B} = \frac{n_A \bar{x}_A + n_B \bar{x}_B}{n_A + n_B} = \frac{100(1200) + 200(1000)}{100 + 200} = S/.1066,67$$

b) $n_{A+B} = 300$; $\bar{x}_{A+B} = 1066,67$; $n_c = 50$; $\bar{x}_c = 900$

Entonces,

$$\bar{x}_{(A+B)+C} = \frac{300(1066,67) + 50(900)}{300 + 50} = \frac{365001}{350} = S/.1042,86$$

1.2.5. Ventajas de la Media Aritmética

1. Es un concepto familiar a la mayoría de las personas
2. Es una medida que puede ser calculada y única. Esta unicidad se debe a que cada conjunto de datos tiene una y sólo una media
3. Se toma en cuenta cada observación del conjunto de datos
4. Es una medida que se determina con mayor certeza que otras características de un conjunto de datos

1.2.6. Desventajas de la Media Aritmética

1**.** La media puede verse afectada por valores extremos que no sean representativos del resto de las observaciones. Por ello, cuando se está utilizando esta medida en un análisis, se advierte la representatividad de los extremos y la influencia que éstos tiene sobre el resultado.

Ejemplo

Supongamos que los sueldos mensuales (en nuevos soles) de los trabajadores de una empresa es como sigue:

Tabla 1.8

Cargo	Nº trabajadores	Sueldos mensuales (S/.)
Gerente General	1	7 500
Administrador	1	5 200
Contador	1	3 000
Empleado	20	1 600
Obrero calificado	15	800
Obrero semicalificado	22	600

a) Determinar el sueldo promedio mensual por trabajador de la empresa
b) ¿Será representativo el sueldo promedio del conjunto de trabajadores?
c) ¿Cuál sería el procedimiento adecuado para un análisis de datos?

Solución

a) $\overline{X} = \dfrac{7500(1)+5200(1)+300(1)+1600(20)+800(15)+600(22)}{60} = S/.1215$ por trabajador

b) No es representativo, por que hay sólo 3 personas con sueldo que hacen crecer el promedio

c) Un procedimiento adecuado podría ser estratificar previamente los datos originales en tres categorías; los Directivos: Gerente General, Administrador y Contador. Empleados y obreros: Calificados y no calificados.

Luego, realizar los cálculos de las medias aritméticas en forma separada para cada categoría.

2. El cálculo de la media aritmética es tedioso por que se usan todas las observaciones en los cálculos (a menos que se use un método corto de datos agrupados para aproximar la media)
3. La media presenta dificultades en el cálculo cuando las tablas de frecuencia tiene: a) el primer intervalo sin indicar el límite inferior; b) el último intervalo donde no se conocen el límite superior; c) una combinación de los dos casos.

(a)	(b)	(c)
Menos de 200	500 – 1000	Menos de 1000
200 – 400	1000- 1500	1000 – 2000
400 – 600	1500 – 2000	2000 - 3000
600 – 800	2000 – 2500	3000 – 4000
800 – 1000	más de 2500	más de 4000

1.3. Mediana

Dado un conjunto de n observaciones $x_1, x_2, ..., x_n$ de la variable o característica x, se define la mediana de este conjunto de valores, como aquel valor que no es superado ni supera a más de la mitad de la n observaciones, arregladas en orden de magnitud creciente o decreciente. La mediana se denota por M_e ó $M_e(x)$.

1.3.1. Cálculo de la Mediana

La mediana para datos no agrupados

Se debe distinguir dos situaciones: número de observaciones impar y número de observaciones par.

a) **Si el número de observaciones n es impar**, basta con ordenar los datos en forma creciente $x_{(1)}, x_{(2)}, ..., x_{(n)}$. Es decir, $x_{(1)}$ denota a la observación más pequeña, $x_{(2)}$

denota a la segunda observación menor,…; $x_{(n)}$ denota a la observación mayor y toma como valor de la mediana la observación que ocupa la posición central. Es decir: $M_e = x_{\left(\frac{n+1}{2}\right)}$

Ejemplo

Considere los pesos en kilogramos (kg) de 7 adultos:

60 80 70 55 75 68 50.

Determinar el peso mediano

Solución

1) Ordenar los datos en forma creciente

50	55	60	68	70	75	80
$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$
$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$

2) Como n = 7, un número impar de observaciones, la mediana es el valor de la observación que ocupa la posición central, es decir $M_e = x_{\left(\frac{7+1}{2}\right)} = x_{(4)} = 68$

Interpretando este valor significaría que el 50 % de los adultos pesan 68 kg o menos y el otro 50% más de 68 kg

b) Si el número de observaciones n es par, en este caso, después de ordenar el conjunto de observaciones, existe dos valores centrales $x_{\left(\frac{n}{2}\right)}$ y $x_{\left(\frac{n}{2}+1\right)}$, tomado como mediana la semi suma de ambos valores.

$$M_e = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

Ejemplo

Un gerente revela la producción de automóviles en 14 días

3	7	12	5	18	6	2
21	13	4	15	16	17	8

Calcular la producción mediana

Solución

1. Ordenar los datos en forma creciente

2	3	4	5	6	7	8
$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$
$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$
12	13	15	16	17	18	21
$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$
$x_{(8)}$	$x_{(9)}$	$x_{(10)}$	$x_{(11)}$	$x_{(12)}$	$x_{(13)}$	$x_{(14)}$

2. Como n = 14, un número par de observaciones, la mediana es el promedio de las dos observaciones centrales $x_{\left(\frac{14}{2}\right)} = x_{(7)} = 8$ y $x_{\left(\frac{14}{2}+1\right)} = x_{(8)} = 12$

$$\text{Es decir, } M_e = \frac{x_{(7)} + x_{(8)}}{2} = \frac{8+12}{2} = 10$$

Interpretando este valor significa que el 50% de los días se fabrican 10 automóviles o menos y el otro 50% más de 10 automóviles.

La mediana para datos agrupados

Para calcular la mediana a partir de tablas de frecuencia, debe determinarse las frecuencias absolutas acumuladas N_i, que permiten conocer hasta que valor de la variable o intervalo se tiene acumulado el 50% de n. Primeramente se calcula $\frac{n}{2}$ este valor se compara con los N_i, donde se presentan los casos siguientes:

i) La mediana en tablas sin intervalos

i.1) Si $\frac{n}{2}$ no coincide con algún N_i se tiene:

$$N_{i-1} < \frac{n}{2} < N_i$$

Entonces $M_e = Y_i$, éste es el valor de la variable asociada a N_i.

Ejemplo

Un taller de imprenta revela los errores de las páginas de un texto universitario.

Tabla 1.9

Distribución de 50 páginas según número de errores

Y_i	n_i
2	10
3	15
4	8
5	2
6	5

Determinar el error mediano

Solución

1) **Determinar la columna de frecuencias absolutas acumulada** N_i

Y_i	n_i	N_i
2	10	10
3	15	25
4	8	33
5	2	35
6	5	40
	n = 40	

2) Como n = 40; $\frac{n}{2} = \frac{40}{2} = 20$ entonces la menor frecuencia absoluta acumulada que a 20 es $N_2 = 25 > 20$. Luego, i = 2; i -1 = 1

3) $\frac{n}{2} = 20 > N_{j-1} = N_1 = 10$, Entonces la mediana es $M_e = Y_2 = 3$

Este valor se interpreta que el 50 % de las páginas tienen 3 errores o menos y el otro 50 % más de tres errores.

i.2) Si $\frac{n}{2}$ coincide con algún N_i se tiene que:

$N_{i-1} = \frac{n}{2} < N_i$. Entonces $M_e = \frac{1}{2}(Y_{i-1} - Y_i)$

Ejemplo

Un profesor analiza las calificaciones obtenidas por los alumnos en el curso de Estadística del Semestre Académico 20010 – I.

Tabla 1.10

Distribución de los alumnos según calificación obtenida en Estadística Semestre Académico 2010 – I

Y_i	n_i
9	8
10	4
11	6
12	3
13	7
14	5
15	3

Solución

1) Determinar la columna de frecuencias absolutas acumuladas N_I

Y_i	n_i	N_i
9	8	8
10	4	12
11	6	**18**
12	3	**21**
13	7	28
14	5	33
15	3	36
	n =36	

2) Como n = 36; $\frac{n}{2}=\frac{36}{2}=18$ coincide con $N_3=18$, por lo tanto $N_3=N_{i-1}=18$, Luego $N_i=21$. De estos resultados se deduce que $Y_i=Y_4=12$ y $Y_{i-1}=Y_3=11$ de donde la mediana es: $M_e=\frac{1}{2}(Y_{i-1}-Y_i)=\frac{1}{2}(12+11)=11{,}5$ puntos.

Es decir, que el 50 % de los estudiantes tienen calificaciones iguales o menos que 11,5 puntos y el otro 50 % más de 11,5 puntos

ii) La mediana en tablas de intervalos

ii.2) Si $\frac{n}{2}$ no coincide con algún N_i se tiene $N_{i-1}<\frac{n}{2}<N_i$. Entonces

$$M_e=L_i+\left[\frac{\frac{n}{2}-N_{i-1}}{n_i}\right].c$$

donde L_i = límite inferior del intervalo que contiene la mediana

n = Número total de datos u observaciones

N_{i-1} = Frecuencia absoluta acumulada, anterior a la clase que contiene la mediana

n_i = Frecuencia absoluta simple del intervalo que contiene la mediana

c = Amplitud del intervalo de la clase mediano.

Ejemplo

Una librería revela los precios (S/.) de los textos universitarios

Tabla 1.11

Distribución de los precios (S/.) de los textos universitarios

Intervalos	n_i
50 -70	20
70 – 90	30
90 – 110	25
110 -130	10
130 – 150	18
150 -170	17

Hallar el precio mediano.

Solución

1. Determinar la columna de frecuencias absolutas acumulada N_i

Intervalos	n_i	N_i
50 -70	20	20
70 – 90	30	**50**
90 – 110	**25**	**75**
110 -130	10	85
130 – 150	18	103
150 -170	17	120
	n = 120	

2.Calcular $\frac{n}{2}=\frac{120}{2}=60$ observando este valor no coincide con algún N_i se encuentra entre 50 y 75 es decir 50 < 60 < 75 de donde tiene $N_{i-1}<\frac{n}{2}<N_i$ de donde $N_{i-1}=50$ y N_i = 75.

3. Ubicar el intervalo mediano. En este caso $N_i = N_3 = 75$ que corresponde al tercer intervalo; $L_i = 90$; $N_{i-1}=50$; $n_i = 25$; C = 110 – 90 = 20

4. Reemplazando los valores obtenidos en la fórmula de M_e

$$M_e = 90 + \left[\frac{\frac{120}{2}-50}{25}\right]..20 = 98$$

Significa que el 50 % del total de textos universitarios, es decir de 60 de ellos tienen precios iguales o inferiores a S/: 98, en tanto que los 60 textos restantes tienen precios superiores a S/: 98.

ii.2) Si $\frac{n}{2}$ coincide con algún N_i se tiene que $N_{i-1} < \frac{n}{2} < N_i$ donde $L_{i-1} - L_i$ .es el intervalo que corresponde a N_i.

Ejemplo

Una evaluación médica revela los pesos (Kg) de los alumnos

Tabla 1.12

Distribución de 60 alumnos según sus pesos (Kg)

Intervalos	n_i
45 – 50	14
50 – 55	16
55 – 60	12
60 – 65	10
65 - 70	6
70 - 75	2

Calcular el peso mediano.

Solución

1. Determinar la columna de frecuencias absolutas acumuladas

Intervalos	n_i	N_i
45 – 50	14	14
50 – 55	16	30
55 – 60	12	42
60 – 65	10	52
65 - 70	6	58
70 - 75	2	60
	60	

2. Calcular $\frac{n}{2} = \frac{60}{2} = 30$, coincide con $N_2 = 30$ este valor es $N_{i-1} = 30$.

   Como N_i = 42. L_i = 55; $n_i = 12$; C = 60 – 55 = 5

3. Reemplazando los valores obtenido en la fórmula de M_e

$$M_e = 55 + \left[\frac{\frac{60}{2} - 30}{12} \right] ..5 = 55$$

Significa que 30 alumnos pesan 55 Kg o menos y los 30 alumnos restantes pesan más de 55 Kg.

1.3.2. Propiedades de la mediana

1. La mediana es un valor adecuado cuando se utiliza para describir distribuciones cuyos valores centrales están muy próximos.
2. Como la mediana depende del número de valores observados, entonces está afectada por las observaciones y no por el tamaño de los valores extremos.
3. La suma $\sum_{i=1}^{n} |x_i - c|$ es mínima cuando c es la mediana de $x_1, x_2, .., x_N$.

   Esto es, $c = M_e$

1.3.3. Ventajas de la mediana

1. Es fácil de entender y puede ser calculada a partir de cualquier clase de datos, aún para datos agrupados en clase abiertas en los extremos
2. Es el valor más representativo de un conjunto de datos que otros promedios, debido a su independencia, de sus valores extremos
3. Se puede encontrar la mediana inclusive datos cualitativos ordinal

1.3.4. Desventajas de la mediana

1. Organiza los datos antes de relacionar cualquier tipo de cálculo para determinar la mediana. Se requiere de mayor tiempo cuando el conjunto de datos tiene muchos elementos
2. Ciertos procedimientos estadísticos son mucho más complejos que aquellos que utilizan la media
3. No es adecuado a manipulaciones algebraicas.

Ejemplo

Una agencia comercial afirma que el salario promedio mensual pagado a sus empleados profesionales es de S/. 3 500, esto sugiere que dicha agencia paga bien. Sin embargo, un análisis posterior indicó que se trata de una pequeña empresa, que trabajan 4 jóvenes profesionales con haberes mensuales de S/. 1 500 cada uno y el gerente general con un haber de S/: 6 000 mensuales. ¿Usted puede seguir afirmando que la firma paga bien?

Solución

Evidentemente no. Por que la media aritmética es afectada por el valor extremo de S/: 6 000. En cambio si calculamos la mediana de los datos, resulta:

$$\begin{array}{ccccc} 1500 & 1500 & 1500 & 1500 & 6000 \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ x_{(1)} & x_{(2)} & x_{(3)} & x_{(4)} & x_{(5)} \end{array}$$

$M_e = x_{\left(\frac{n+1}{2}\right)} = x_{(3)} = S/.1500$ que es un valor representativo en el sentido que localiza mejor el "centro" del conjunto de datos, que la media aritmética.

Ejemplo.

Se tienen cinco tirajes de impresión. Los resultados se pueden clasificar de acuerdo a la claridad de la imagen. Los resultados organizados desde el mejor hasta el peor es: extremadamente claro, muy claro; claro; sensiblemente borroso y muy borroso. De modo que, la mediana es la clasificación que ocupa la posición central $\frac{n+1}{2} = \frac{5+1}{2} = 3$, ósea la tercera clasificación ($M_e = claro$).

1.4. **Moda**

Es un concepto muy simple, con el mismo sentido que se da en el lenguaje común. Es el valor que ocurre con mayor frecuencia, es decir, el valor de la variable más frecuente. La moda puede no existir, e inclusive no ser única en el caso de existir. De otra forma, ésta medida está constituida por la categoría de la distribución o representada por el intervalo que tiene la mayor frecuencia absoluta simple (n_i). La moda también se denomina Modo, Valor Modal o Promedio Típico, se simboliza con Mo ó Md.

1.4.1. Cálculo de la Moda

La moda para datos no agrupados

Es la mayor repetición de un conjunto de datos u observaciones. Si el conjunto de datos tiene una sola moda se denomina unimodal, tiene dos modas, se llama bimodal. En general, si tiene más de dos modas se denomina multimodal.

Ejemplo.

Consideremos los pesos (en kilos) de 9 adultos

82, 65, 59, 74, 60, 67, 71, 73, y 70

Estas 9 observaciones no tienen moda

Ejemplo

Consideremos la distribución de los peso de 14 adultos

63, 67, 70, 69, 81, 57, 63,
73, 68, 63, 71, 71, 71, 83

Solución

El valor 63 y 71 ocurren 3 veces y el resto ocurren una vez cada uno. Luego, la moda de estas observaciones es $Mo. = 63$ y 71 kilos (bimodal)

Interpretación este valor significa que el peso más frecuente en los adultos es 63 y 71 kilos.

La moda para datos agrupados

Cuando los datos están tabulados, la clase que tiene la mayor frecuencia será la que contiene a la moda, y se llama clase modal.

i) Tablas sin intervalos

Ejemplo

Se realizó una encuesta de la distribución de familias, según el número de personas

Tabla 1,13

Distribución de las familias, según el número de personas, en el Centro Poblado de Renovación, 2009

Número de personas x_i	Número de familias n_i
2	16
3	24
4	52
5	76
6	40
7 y más	12
Total	n = 220

Calcular el tamaño modal.

Solución

Se observa que la mayor frecuencia es $n_4 = 76$ que corresponde al cuarto valor de la variable $x_4 = 5$, esto indica que el valor modal es 5. Luego $Mo = 5$. El resultado significa que es más frecuente hallar familias integradas por 5 personas cada una.

ii) Tablas con intervalos

Cuando se tiene una tabla con intervalos, al igual que la mediana se determina el intervalo modal que corresponde a la mayor frecuencia absoluta simple (n_i) y se utiliza la fórmula siguiente:

$$Mo = L_i + \left[\frac{(n_i - n_{i-1})}{(n_i - n_{i-1}) + (n_i - n_{i+1})} \right] .c$$

Donde:

L_i = Límite inferior que contiene la moda

n_i = Frecuencia absoluta simple de la clase que contiene la moda

n_{i-1} = Frecuencia absoluta simple anterior a la clase que contiene la moda

n_{i+1} = Frecuencia absoluta simple posterior a la clase que contiene la moda

c = Amplitud del intervalo de la clase modal

Ejemplo

Considerar la Tabla 1.11 y calcular el precio modal de los textos universitarios.

Intervalos	n_i
50 -70	20
70 – 90	30
90 – 110	25
110 -130	10
130 – 150	18
150 -170	17

Solución

1. Identificar la mayor frecuencia absoluta simple (n_i), y los valores n_{i-1}; n_{i+1}. En la tabla, corresponde n_i = n_2 = 30, luego la frecuencia absoluta simple anterior es $n_1 = 20$, y la posterior a n_2 es $n_3 = 25$ de donde se deduce que $n_{i-1} = 20$; $n_{i+1} = 25$
2. Ubicar el intervalo modal

   Aquí el intervalo que corresponde a n_i = n_2 = 30, corresponde al segundo intervalo, $L_i = 70$; c = 90 – 70 = 20
3. Reemplazando los valores obtenido en la fórmula de Mo.

$$Mo = 70 + \left[\frac{(30-20)}{(30-20)+30-35)} \right].20 = 83{,}33$$

Significa que el precio más frecuente en los 120 textos universitarios es S/. 83,33 o que también que la mayoría de los textos universitarios tienen precio de S/. 83,33.

Ejemplo

Considerar la Tabla 1.12 y calcular el peso modal de los alumnos

Intervalos	n_i
45 – 50	14
50 – 55	16
55 – 60	12
60 – 65	10
65 - 70	6
70 - 75	2

Solución

1. Identificar la mayor frecuencia absoluta simple (n_i), y los valores n_{i-1} ; n_{i+1} . En la tabla corresponde $n_i = n_2 = 16$, luego la frecuencia absoluta simple es $n_1 = 14$, y la posterior a n_2 es $n_3 = 12$ deduciéndose que $n_{i-1} = 14$ y $n_{i+1} = 12$

2. Ubicar el intervalo modal

   El intervalo que corresponde a $n_i = n_2 = 16$ que corresponde al segundo intervalo;

   $L_i = 50$, c = 55 – 50 = 5

3. Reemplazando los valores obtenidos en la fórmula de Mo

$$Mo = 50_i + \left[\frac{(16-14)}{(16-14)+(16-12)} \right] .5 = 51{,}66$$

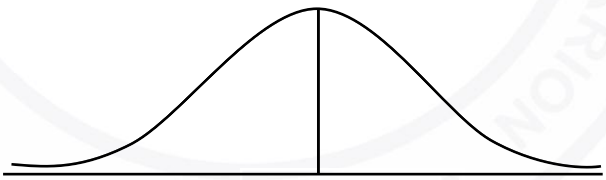
Significa que el peso más frecuente en los 60 alumnos es 51 kilos con 660 gramos o también que la mayoría de los alumnos pesan 51 kilos con 660 gramos

1.4. 2. Relación entre la Media, la Mediana y la Moda

1. En una distribución de frecuencias simétrica cuya representación gráfica es acampanada, y además unimodal, la media, la mediana y moda se localizan en el centro y siempre son iguales (Figura Nº 1.4). Es decir: $\bar{x} = M_e = Mo$

Figura Nº 1.4

Distribución simétrica



$\bar{x} = Me = Mo$

2. Si la distribución tiene la forma acampanada, es unimodal; pero no tiene simetría, las tres medidas toman valores diferentes, y la mediana generalmente está comprendida entre la moda y la media aritmética.

   i) Si la distribución es más alargada para valores grandes de la variable (asimetría a la derecha o positiva) entonces se tiene:

   $\bar{x} > M_e > Mo$ (Figura Nº 1.5)

   ii) Si la distribución es más alargada para valores pequeños de la variable (asimetría a la izquierda o negativa), entonces se tiene:

   $\bar{x} < M_e < Mo$ (Figura Nº 1.6)

Figura Nº 1.5

Asimetría positiva

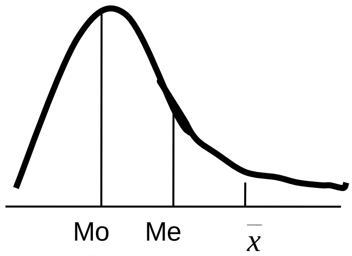
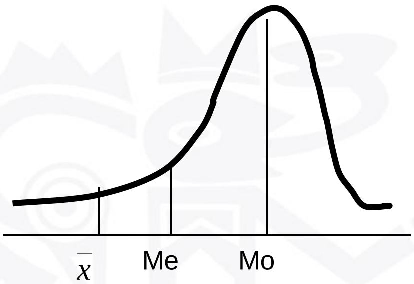


Figura Nº 1.6

Asimetría negativa



3. Si la distribución es moderadamente asimétrica y unimodal, se cumple aproximadamente la relación: $\bar{x} - Mo = 3(\bar{x} - Me)$

1.4.3. Ventajas de la moda

1. Es muy útil para datos cualitativos
2. No está afectada por valores extremos
3. Se puede calcular aún cuando una más de las clases sean abiertas en los extremos

1.4.4. Desventajas de la moda

1. No hay un valor modal, por que el conjunto de datos no contiene valores que se repitan más de una vez otras veces, cada valor es la moda, por que cada uno aparece el mismo número de veces.
2. Cuando el conjunto de observaciones contiene dos, tres o más modas, éstas son difíciles de interpretar y comparar.

1.5. Media Geométrica

La media geométrica o promedio geométrico se simboliza por Mg ó G. En muchas ocasiones se utiliza su transformación en el manejo estadístico de variables con distribución normal. Es relevante cuando varias cantidades son multiplicadas para producir un total.

1.5.1. Cálculo de la Media Geométrica

La media geométrica para datos no agrupados

La media geométrica simple "Mg" de n observaciones $x_1, x_2, x_3, ..., x_n$ positivas, está dada por la raíz enésima del producto de los n valores observados, es decir; $Mg = \sqrt[n]{x_1.x_2.,...,x_n}$ donde: Mg = Media Geométrica

n = Número total de datos u observaciones y $x_1, x_2, x_3, ..., x_n$ son los datos

Ejemplo

Las utilidades obtenidas por una compañía constructora en cuatro proyectos fueron 3, 2, 4 y 6 % respectivamente. ¿Cuál es la media geométrica de las ganancias?

Solución

$$Mg = \sqrt[4]{(3).(2).(4).(6)} = \sqrt[4]{144} = 3{,}46\%$$

La media geométrica de 3,46 da una cifra más conservadora para las utilidades por que no se ve tan afectada por los valores extremos. En realidad será igual o menor que la media aritmética.

Ejemplo

La producción anual de un producto en una fábrica, ha tenido el siguiente crecimiento:

2 006 a 2 007 la producción aumento de 16 00 a 2 000 (25 %)

2007 a 2008 la producción aumentó de 2000 a 3000 (50%)

2008 a 2009 la producción aumentó de 3000 a 6600 (120 %) cada año el aumento fue de 25%, 50% y 120 % respectivamente. ¿Cuál fue el aumento promedio anual de la producción?

Solución

Por los incrementos se deduce que la producción tiene un comportamiento no lineal, entonces es recomendable determinar el promedio geométrico

$Mg = \sqrt[3]{(25).(50).(120)} = \sqrt[3]{150000} = 53{,}1$ que expresa que la producción tiene un incremento anual promedio de 53 %.

La media geométrica para datos agrupados

La media geométrica ponderada, es la raíz enésima del producto de los puntos medios elevados a sus respectivas frecuencias absolutas simples. Es decir:

$$Mg = \sqrt[n]{x_1^{n_1}.x_2^{n_2}.,...,x_n^{n_m}}$$

Donde:

Mg = Media Geométrica

n = Número total de datos u observaciones

x_i = Puntos medios

n_i =Frecuencias absolutas simples

También se puede expresar en la forma:

$$\log Mg = \frac{1}{n}\left[n_1 \log x_1 + n_2 \log x_2 + \ldots + n_m \log x_m\right]$$

Ejemplo

Consideremos un conjunto de datos distribuidos en una tabla de frecuencias y hallar la media geométrica.

Tabla Nº 1.14

x_i	n_i	$\log x_i$	$n_i \log x_i$
15	**5**	**1,1761**	**5,8805**
25	**19**	**1,3979**	**26,5601**
35	**24**	**1,5441**	**37,0584**
45	**13**	**1,6532**	**21,4916**
55	**4**	**1,7404**	**6,9616**
	n = 65		**97,9522**

$$\log Mg = \frac{1}{65}\left[5\log 15 + 19\log 25 + 24\log 35 + 13\log 45 + 4\log 55\right] = \frac{97{,}9522}{65} = 1{,}5069$$

$$Mg = anti\log(1{,}5069) = 32{,}13$$

1.5.2. Ventajas de la Media Geométrica

1. Se toman en cuenta todos los valores de la variable
2. Los valores extremos tienen menos influencia que en la media aritmética por estar definida a través de productos en vez de sumas
3. Es más representativa que la media aritmética cuando la variable evoluciona de forma acumulativa con efectos multiplicativos
4. Está definida en forma objetiva y es única, si existe

1.5.3. Desventajas de la Media Geométrica

1. Su cálculo es muy tedioso
2. Está limitada por valores positivos para que pueda ser interpretada
3. Si algún valor de la variable es cero, entonces la media geométrica será cero
4. Si aparece algún valor negativo, la media geométrica tomará un valor imaginario
5. Su significado estadístico es menos intuitivo que la media aritmética
6. No puede ser calculada en distribuciones con clases abiertas
7. Se presta a manipulaciones algebraicas posteriores

1.5.4. Aplicaciones de la Media Geométrica

Se aplica mayormente en las series cronológicas que siguen una tendencia exponencial, se hace indispensable su uso, si se desea calcular valores intermedios, es decir, si se quiere interpolar linealmente.

Se utiliza para determinar el incremento porcentual promedio en ventas, producción u otras actividades o series económicas de un periodo a otro y para promediar porcentajes, índices y tasas de cambio.

1.6. Media Armónica

Es un promedio muy útil en conjuntos de números que se definen en relación con alguna unidad se representa por H ó M_h

1.6.1. Cálculo de la Media Armónica

La Media Armónica para datos no agrupados

La Media Armónica simple "M_h" de n términos no nulos $x_1, x_2, ..., x_n$ es el recíproco de la media aritmética de los recíprocos de los términos considerados. Es decir:

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

Donde:

M_h = Media armónica

n = Número total de términos

x_i = Valores de los términos

Ejemplo

Un equipo de trabajadores textiles tiene que producir 180 metros de casimir; de los cuales elaboran los primeros 90 metros con una productividad de 15 metros diarios, y los 90 metros restantes lo hacen a razón de 20 metros por día. ¿Cuál es la productividad diaria durante todo el trabajo?

Solución

Si se considera la media aritmética, resulta:

$$\bar{x} = \frac{15+20}{2} = 17{,}5 \text{ metros diarios}$$

Por su parte, para elaborar los primeros 90 metros se necesitan 6 días, y para los siguientes 90 metros sólo 4,5 días. Si la productividad promedio diaria es de 17,5 metros, en los 10.5 días se fabricarán (10,5) (15) = 183,75 metros que es superior a los 80 metros considerados, lo que es inconsistente. Sin embargo, cuando se utiliza la media armónica, los resultados son consistentes.

$$M_h = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{2}{\frac{1}{15}+\frac{1}{20}} = 17{,}14 \text{ metros diarios.}$$

Por lo tanto, trabajando los 10,5 con una productividad diaria de 17,14 metros, se logra producir (10,5) (17,14) = 180 metros.

Ejemplo

Cinco secretarias tienen las siguientes velocidades 40, 36, 28 y 35 palabras por minuto para escribir. Si cada una de ellas escribe un mismo texto. Calcular la media armónica.

Solución

$$M_h = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{5}{\frac{1}{40}+\frac{1}{36}+\frac{1}{28}+\frac{1}{40}+\frac{1}{35}} = 35{,}196 = 35 \text{ palabras por minuto.}$$

La Media Armónica para datos agrupados

La media armónica ponderada se define por:

$$M_h = \frac{n}{\sum_{i=1}^{m}\frac{n_i}{x_i}}$$

Donde:

M_h = Media armónica

n = Número total de términos

m = Número de clases

x_i = Marca de clases

n_i = Frecuencia absoluta simple

Ejemplo

Considérese un conjunto de datos distribuidos en una tabla de frecuencias y determinar la media armónica.

Tabla Nº 1.15

Intervalos	n_i	x_i
20 – 40	15	30
40 – 60	25	50
60 – 80	10	70
80 – 100	40	90
100 - 120	30	110
	n = 120	

Solución

$$M_h = \frac{n}{\sum_{i=1}^{m} \frac{n_i}{x_i}} = \frac{120}{\frac{15}{30}+\frac{25}{50}+\frac{10}{70}+\frac{40}{90}+\frac{30}{110}} = 54{,}71$$

1.6.2. Ventajas de la Media Armónica

1. Está definida en forma objetiva y es única
2. Su cálculo es sencillo
3. Intervienen todos los valores de la distribución

1.6.3. Desventajas de la Media Armónica

1. Se presta a manipulaciones algebraicas posteriores
2. No es posible calcularla cuando existen valores iguales a cero
3. Esta afectada por los valores extremos. Pero da a los valores grandes un peso menor que el de la media geométrica; mientras que a los valores pequeños; le da un peso mayor que el que da la media aritmética y la media geométrica

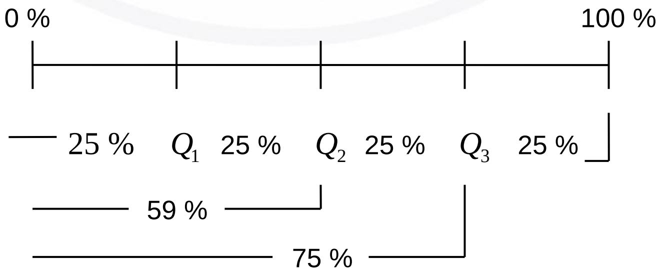
1.6.4. Aplicaciones de la Media Armónica

1. La media armónica se aplica cuando se tiene términos para cuyos recíprocos se quiere calcularse.
2. Se utiliza para promediar velocidades, rendimientos, productividades, etc en los que hay que combinar una serie de conceptos, tales como: “entidades de producción” (recorridos, fincas, empresas, secciones, etc), “recursos producidos” por cada entidad, “total de recursos”, “ritmos de producción” de cada identidad que se expresa en producto obtenido por unidad de producción y unidades de producción de cada entidad por su ritmo de producción

1.7. Cuartiles

Son valores que dividen un conjunto ordenado en cuatro partes iguales y se denota por Q_i para i = 1, 2, 3

Figura Nº 1.7



1.7.1. Representación de los Cuartiles

Primer cuartil: Q_1

Es el tal que el 25 % de las observaciones son menores o iguales y el

75 % mayores.

Segundo cuartil: Q_2

Es la mediana, tal que el 50 % de las observaciones son menores o iguales y el otro 50 % mayor.

Tercer cuartil: Q_3

Es el valor tal que 75 % de las observaciones son menores o iguales y los 25 % mayores.

1.7.2. Cálculo de los Cuartiles

Los cuartiles para datos no agrupados

$$Q_1 = \frac{n+1}{4}$$ **observación ordenada**

$$Q_2 = \frac{n+1}{2}$$ **observación ordenada**

$$Q_3 = \frac{3(n+1)}{4}$$ **observación ordenada**

Ejemplo

Al examinar los registros de facturación de una empresa editora con ventas a crédito, el editor toma una muestra de 11 facturas no pagadas. Las sumas que se adeudan en miles de nuevos soles son: 4, 18, 11, 7, 7, 10, 21, 5, 33, 9 y 12. Calcular los cuartiles.

Solución.

1. **Ordenar los datos**

$$\begin{array}{ccccccccccc} 4 & 5 & 7 & 7 & 9 & 10 & 11 & 12 & 18 & 21 & 33 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ x_{(1)} & x_{(2)} & x_{(3)} & x_{(4)} & x_{(5)} & x_{(6)} & x_{(7)} & x_{(8)} & x_{(9)} & x_{(10)} & x_{(11)} \end{array}$$

2. **Como n = 11, entonces**

El primer cuartil es $Q_1 = \frac{11+1}{4} = x_{(3)} = 7$. Significa que el 25 % del total de facturas no pagadas adeudan S/. 7 000 ó menos y el 75 % restante adeudan más de S/. 7 000.

Segundo cuartil $Q_2 = \frac{n+1}{2} = \frac{11+1}{2} = x_{(6)} = 10$. Significa que el 50 % del total de facturas no pagadas adeudan S/ 10 000 ó menos y el 50 % restante adeudan más de S/. 10 000.

Tercer cuartil $Q_3 = \frac{3(n+1)}{4} = \frac{3(11+1)}{4} = x_{(9)} = 18$.

Significa que el 75 % del total de facturas no pagadas adeudan S/. 18 000 ó menos y el 25 % restante adeudan más de S/. 18 000.

Ejemplo

El jefe de control de carreteras revela la velocidad (Km / h) de 12 vehículos seleccionados aleatoriamente.

50, 70, 110, 80, 60, 45, 38, 95, 60, 72, 40, 55

Determinar los cuartiles.

Solución

1. Ordenar los datos

$$\begin{array}{cccccccccccc} 38 & 40 & 45 & 50 & 55 & 60 & 60 & 70 & 72 & 80 & 95 & 110 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ x_{(1)} & x_{(2)} & x_{(3)} & x_{(4)} & x_{(5)} & x_{(6)} & x_{(7)} & x_{(8)} & x_{(9)} & x_{(10)} & x_{(11)} & x_{(12)} \end{array}$$

2. Como n = 12, entonces:

El primer cuartil es $Q_1 = \frac{12+1}{4} = 3{,}25$. Significa que el valor de Q_1 es el tercer dato mas 25 % de la diferencia entre los valores de las observaciones tercero y cuarto. Así el valor de la tercera observación es 7 y el del cuarto es 9, entonces:

$Q_1 = 45 + (45-50)(0{,}25) = 45 + 1{,}25 = 46{,}25 Km/h.$.

Significa que el 25 % del total de vehículos, es decir, 3 de ellos tienen velocidades menores o iguales a 46, 25 Km/h y los 9 restantes o sea el 75 % de vehículos, tienen velocidades mayores a 46,25 Km/h.

Segundo cuartil $Q_2 = \frac{n+1}{2} = \frac{12+1}{2} = 6{,}5$. Evidentemente el segundo cuartil coincide con la mediana, es decir $Q_2 = M_e$.

$$Q_2 = \frac{x_{(6)} + x_{(7)}}{2} = \frac{60+60}{2} = 60 \text{ Km / h}$$

Significa el 50 % del total de vehículos, es decir, 6 de ellos tienen velocidades **menores o iguales a 60 Km /h y los 6 restantes, o sea el 75 % de vehículos, tienen velocidades mayores a 60 Km / h.**

Tercer cuartil $Q_3 = \frac{3(n+1)}{4} = \frac{3(12+1)}{4} = \frac{39}{4} = 9{,}75$

Significa que el valor de Q_3 es el del noveno dato más 75 % de la diferencia entre los valores de las observaciones novena y décima. Puesto que la novena observación es 72 y la décima 80, entonces:

$$Q_3 = 72 + (80-72)(0{,}75) = 78 \text{ Km / h}$$

Significa el 75 % del total de vehículos, es decir, 9 de ellos tienen velocidades menores o iguales a 78 Km/h y los tres restantes o sea el 25 % de vehículos, tienen velocidades mayores a 78 Km/h.

Los cuartiles para datos agrupados

$$Q_1 = L_i + \left[\frac{\frac{n}{4} - N_{i-1}}{n_i} \right].c$$

$$Q_2 = M_e = L_i + \left[\frac{\frac{n}{2} - N_{i-1}}{n_i} \right].c$$

$$Q_3 = L_i + \left[\frac{\frac{3n}{4} - N_{i-1}}{n_i} \right].c$$

Ejemplo

Considerar los pesos en kilogramos (Kg) de un grupo de 60 alumnos mostrados en una tabla de distribución de frecuencia

Tabla Nº 1.16

Intervalos	n_i
40 – 45	10
45 – 50	12
50 – 55	20
55 – 60	8
60 -65	4
65 - 70	6

Hallar los cuartIles.

Solución

1 Construir la columna de frecuencias absolutas acumuladas

Intervalos	n_i	N_i
40 – 45	10	10
45 – 50	12	22
50 – 55	20	42
55 – 60	8	50
60 -65	4	54

65 - 70	6	60
	n = 60	

2 Primer cuartil

Analizar $\frac{n}{4}=\frac{60}{4}=15$. $L_i=45$; $n_i=12$; $N_{i-1}=10$; c = 5. Reemplazando tenemos:

$Q_1=45+\left(\frac{15-10}{12}\right).5=47{,}08$. Significa que del total de alumnos, el 25 % , es decir 15 de ellos, tienen pesos inferiores o menores a 47,08 Kg, y el 75 % restante tienen pesos superiores a 47, 08 Kg.

Segundo Cuartil

Analizar $\frac{n}{2}=\frac{60}{2}=30$. $L_i=50$; $n_i=20$; $N_{i-1}=22$; c = 5. Reemplazando tenemos:

$Q_2=50+\left(\frac{30-22}{20}\right).5=52$. Significa que del total de alumnos, el 50 %, es decir, 30 de ellos tienen pesos inferiores o menores de 52 Kg y el 50 % restante tienen pesos superiores a 52 Kg.

Tercer cuartil

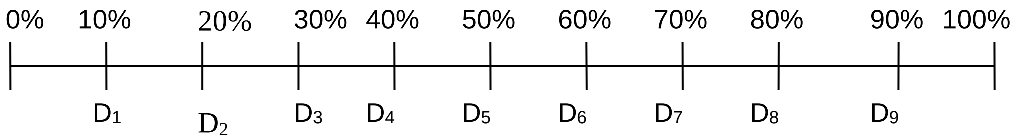
Analizar $3\frac{n}{4}=\frac{3(60)}{4}=45$. $L_i=55$; $n_i=8$; $N_{i-1}=42$; c = 5. Reemplazando tenemos:

$Q_3=55+\left(\frac{45-42}{8}\right).5=56{,}875$. Significa que del total de alumnos, el 75 %, es decir, 45 de ellos, tienen pesos o inferiores o menores 56, 875 Kg, y el 25 % restante tienen pesos superiores a 56, 875 Kg.

1.8. Deciles

Son valores que dividen un conjunto ordenado en 10 partes iguales y se denota por D_i, I = 1, 2,…, 9

Figura Nº 1.8



1.8.1. Representación de los Deciles

Primer Decil: D_1

Es el valor tal que el 10 % de las observaciones son menores o iguales y el 90 % mayor

Segundo Decil: D_2

Es el valor tal que el 20 % de las observaciones son menores o iguales y el 80 % mayor.

Así sucesivamente.

Noveno Decil: D_9

Es el valor tal que el 90 % de las observaciones son menores o iguales y el 10 % mayor.

1.8.2. Cálculo de los Deciles

Los deciles para datos no agrupados

$$D_i = \frac{i(n+1)}{10}\text{; i = 1, 2, .., 9}$$

Los deciles para datos agrupados

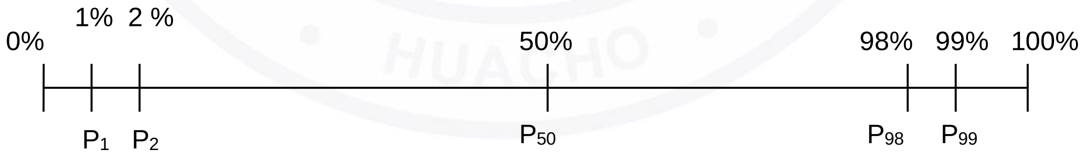
$$D_i = L_i + \left(\frac{\frac{in}{10} - N_{i-1}}{n_i} \right).c \qquad \text{i = 1, 2,.., 9}$$

1.9. Percentiles

Son valores que dividen un conjunto ordenado en 100 partes iguales y se denota por

P_i i = 1, 2,…, 99

Figura Nº 1. 9



1.9.1. Representación de los Percentiles

Primer percentil: P_1

Es el valor tal que el 1 % de las observaciones son menores iguales y el 99% mayor.

Segundo Percentil: P_2

Es el valor tal que el 2 % de las observaciones son menores o iguales y el 98 % mayor.

Así, sucesivamente.

Noventa y nueve avo Percentil. P_{99}

Es el valor tal que el 99% de las observaciones son menores o iguales y el 1 % mayor.

1.9.2. Cálculo de los Percentiles

Los percentiles para datos agrupados

$$P_i = L_i + \left(\frac{\frac{in}{100} - N_{i-1}}{n_i} \right) .c \quad \textbf{para i = 1, 2, ….,99}$$

Ejemplo

Una empresa decide hacer un reajuste entre sus empleados. La clasificación se lleva a cabo mediante la aplicación de un test que presenta las siguientes puntuaciones.

Tabla Nº 1.17

Puntuaciones	Nº de empleados
10 – 30	94
30 – 50	140
50 – 70	160
70 – 90	98
90 - 110	8

Determinar D_2 **y** P_{72}

Solución

1. Construimos la columna de frecuencias absolutas acumuladas

Intervalos	n_i	N_i
10 – 30	94	94
30 – 50	140	234
50 – 70	160	394
70 – 90	98	492
90 - 110	8	500
	n =500	

2. Segundo decil

Analizar $\frac{2n}{10} = \frac{2(500)}{10} = 100$. $L_i = 30$.; $n_i = 140$; $N_{i-1} = 94$; c = 20

Reemplazando tenemos:

$D_2 = 30 + \left(\frac{100 - 94}{140} \right).20 = 30{,}86$. Significa que, el 20 % es decir, 100 de ellos tienen calificaciones menores o iguales a 30,86 puntos y el 80 restante tienen calificaciones mayores a 30,86 puntos.

Setenta y dos avo Percentil

Analizar $\frac{72n}{100} = \frac{72(500)}{100} = 360$. $L_i = 50$.; $n_i = 160$; $N_{i-1} = 234$; c = 20

Reemplazando tenemos:

$$P_{72} = 50 + \left(\frac{360 - 234}{160} \right).20 = 65{,}75$$

Significa que del total de empleados, el 72 % es decir, 360 de ellos tienen calificaciones menores o iguales a 65,75 puntos y el 28 % restante tienen calificaciones mayores a 65,75 puntos.

Ejercicios Propuestos

1. Los siguientes datos representan el número de interrupciones por día de trabajo debido a fallas mecánicas en una planta procesadora de alimentos:

   3, 4, 1, 3, 6, 5, 6, 3, 2, 3

   Calcular la media, la mediana y l a moda
2. La media mínima para aprobar una asignatura es 11. Si un estudiante obtiene las notas 13,5; 14; 9,5; 12; 8,5; 8, 11,5; 10 en los exámenes de la asignatura. ¿El estudiante fue aprobado?
3. El Ministerio de Transportes y Comunicaciones revela la distribución del número de accidentes por día, en cierta autopista

Nº de accidentes	0	1	2	3	4
Nº de días	10	15	10	2	3

   a) Determinar la media b) Calcular la mediana c) Hallar la moda
4. El jefe de control de calidad de una empresa ha clasificado un lote de 80 artículos con una distribución de 6 clases y una amplitud de 5 unidades. Si las frecuencias correspondientes son: 6, 12, 24, 18, 13 y 7 siendo la cuarta marca de clase igual a 35. Determinar la mediana y la moda de la distribución.
5. En un examen participaron cuatro grupos A, B, C, y D, con total de 100 alumnos. Habiendo obtenido un promedio general de 72 puntos. Los puntajes medios de los grupos A, B y C fueron 75, 62 80 puntos respectivamente. Los archivos del grupo D se extraviaron; pero se sabe que los cursos A y B eran el 40 % y el 25 % del total respectivamente, y que en el grupo C habían 15 alumnos más que en el grupo D. En base a la información anterior, determinar la nota media de la sección D.

6. Un proyecto minero tiene cuatro zonas mineralizadas. La zona I tiene 3 millones de T. M con una ley 1,8 % por T.M; la zona II, tiene 5 millones de T. M. con una ley de 2,5 % por T.M; la zona III, 7 millones de T. M con una ley de 2,0 % por T. M con una ley de 3,0 % por T. M. Determinar la ley promedio por T.M. del proyecto minero.
7. Un investigador recibe las siguientes respuestas a un enunciado en una encuesta de evaluación: discrepa fuertemente, discrepa ligeramente, discrepa un poco, concuerda, fuertemente. ¿Cuál es la mediana de las 5 respuestas?
8. Una organización de cadena de restaurantes tiene 7 de ellos en cierta carretera. Diariamente cada restaurante debe buscar alimentos frescos de un almacén central. Además el servicio requiere dos camiones cada día, entonces para minimizar la distancia total de los 7 restaurantes al almacén central,¿dónde debe estar situado el almacén?

Restaurantes	Distancia en millas del restaurante al almacén
A	12
B	40
C	76
D	96
E	124
F	132
G	148

9. Una compañía maneja una pequeña refinería que vende gasolina al por mayor, a minoristas independiente. Las ventas de la semana pasada fueron las siguientes.

Galones de gasolina (en miles)	Número de operaciones
00 – 10	10
10 – 20	20
20 - 30	30
30 – 40	25
40 – 50	15
50 – 60	10
60 – 70	5
70 - 80	5

a) Determinar la media de los galones vendidos la semana pasada
b) Hallar la mediana y moda
c) Calcular Q_1; D_5; P_{38} y varianza

CAPÍTULO II

MEDIDAS DE DISPERSIÓN

Llamadas también medidas de variabilidad, cuantifican la separación, la dispersión, la variabilidad de los valores de la distribución respecto al valor central.

Por ejemplo, si estás de paseo en un Parque Nacional y te encuentras frente a un río que cruzar, le preguntas a quien conoce más, que profundidad tiene ese río. Supongamos que hay un guardián del Parque y éste te dice que el río tiene profundidad promedio de un metro. ¿Lo cruzarías sin información adicional?.

Probablemente no. Tú querías saber acerca de la variación de la profundidad. Si la profundidad máxima es de 1,5 metros y la mínima de 0,5 metros podría ser que te animarías a cruzarlo; pero ¿qué pasaría si averiguas que la profundidad del río va de 0,1 metros (o sea 10 centímetros en la orilla a 1,90 metros)?. Entonces quizás note decidirías a cruzarlo.

Antes de decidirte a cruzar el río desearías tener información sobre la profundidad típica y sobre la variación de la profundidad del río.

2.1 Importancia de las medidas de dispersión

- Proporcionan información adicional que permite juzgar la confiabilidad de la medida de tendencia central. Si los datos se encuentran ampliamente dispersos, la posición central es menos representativa de los datos se compara dispersiones de diferentes muestras
- Si no se desea tener una amplia dispersión de valores con respecto al centro de la distribución o esto presenta riesgos inaceptables, necesitamos tener habilidad de reconocerlo y evitar escoger distribuciones que tenga las dispersiones más grandes. Supongamos por ejemplo, que te interesa evaluar dos programas distintos de prevención de drogas. En el programa A, los jóvenes son egresados, en promedio a

los 18 meses; en el programa B los jóvenes también son egresados a los 18 meses en promedio. Si comparas estas medidas, podrías concluir, que ambos programas tienen desempeños similares. Sin embargo, al observar 3 años de registro de cada joven egresado, te das cuenta que tu conclusión no es del todo correcta. El programa A egresa como mínimo a los jóvenes a los 16 meses y como máximo a los 20. El progre B egresa como mínimo a los 11 meses y como máximo a los 25 meses. En el primer caso se tiene un programa más regular y quizá más confiable que en el segundo caso.

2.2 Amplitud de Variación

Se le conoce como Rango o Recorrido, es la diferencia entre los dos valores extremos que toma la variable. Es la medida de dispersión más sencilla y la que proporciona menos información. Además, esta información puede ser errónea, pues el hecho de que no influyan más de dos valores del total del conjunto de observaciones puede provocar una deformación de la realidad. Se simboliza por A.V.

2.2.1. Cálculo de la Amplitud de Variación

La amplitud de variación para datos no agrupados

Es la diferencia entre el valor mayor y el valor menor en un conjunto de datos

A.V = Valor mayor – Valor menor

Ejemplo

Se tienen las edades de cinco estudiantes universitarios

23 18 27 34 25

Calcular la amplitud de variación.

Solución

A.V = 34 – 18 = 16 años

Ejemplo

En un hospital el pulso de cada paciente se mide tres veces al día y que los registros de dos pacientes muestran:

Paciente 1:	73	77	74
Paciente 2:	64	90	73

¿Cuál es la amplitud en pulsaciones para cada paciente?

Solución

Para el paciente 1: A.V = 77 – 73 = 4

Para el paciente 2: A.V = 90 – 64 = 26

La amplitud de variación para datos agrupados

Es la diferencia entre el límite superior del último intervalo clase y el límite inferior del primer intervalo.

A.V = Límite superior de la clase m –límite inferior de la clase 1

Ejemplo

Un profesor revela las notas finales de un grupo de 50 alumnos en el curso de Cálculo de Probabilidades del V Ciclo.

Intervalos	2 - 4	4 - 6	6 - 8	8 - 10	10 - 12	12 - 14
Número de alumnos	10	2	3	14	12	4

Calcular la amplitud de variación

Solución

A. V = 14 – 2 = 12 puntos

2.2.2. Ventajas de la Amplitud de Variación

Es una medida simple para calcular y fácil de entender.

2.2.3. Desventajas de la Amplitud de Variación

- **En su cálculo sólo intervienen dos elementos del conjunto**
- **Al aumentar el número de observaciones, puede esperarse que aumente la variabilidad.**
- **No es una medida adecuada para comparar la variabilidad de dos grupos de observaciones, a menos que estos sean del mismo tamaño**

2.3. Desviación Media

Denominada Desviación Promedio, es la media aritmética de los valores absolutos de las desviaciones respecto a la media. Se simboliza por D.M.

2.3.1. Cálculo de la desviación Media

La desviación media para datos no agrupados

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

Donde:

x_i = Valor de cada observación o elemento

$\bar{x}$ = Media aritmética

n = Número total de observaciones

Ejemplo

Tres alumnos son evaluados en una competencia sus conocimientos en 10 materias, cada una sustentada con 10 preguntas. La idea del concurso es encontrar al alumno más idóneo para representar al colegio en un concurso a nivel nacional. El número de preguntas buenas se muestran a continuación:

Materia	Carlos	Julio	Javier
1	2	7	5
2	9	2	6
3	10	2	5
4	2	6	5
5	3	6	5

6	1	3	5
7	9	6	4
8	9	7	5
9	1	6	6
10	4	5	4

Solución

Lo primero que analizaremos es la media de los puntajes para cada uno de los alumnos, con el fin de determinar el alumno con mayor promedio de preguntas buenas.

Las medias para los resultados coinciden los tres alumnos responden en promedio 5 preguntas correctas por prueba. ¿Cuál sería entonces el indicador diferenciador entre los alumnos?

Completemos el análisis anterior calculando la desviación media:

Carlos:

$$D.M = \frac{2|2-5| + 3|9-5| + |10-5| + |3-5| + 2|1-5| + |4-5|}{10} = \frac{34}{10} = 3{,}4$$

Julio:

$$D.M = \frac{2|7-5| + 2|2-5| + 4|6-5| + |3-5| + |5-5|}{10} = \frac{16}{10} = 1{,}6$$

Javier:

$$D.M = \frac{6|5-5| + 2|6-5| + 2|4-5|}{10} = \frac{4}{10} = 0{,}4$$

Carlos muestra una desviación media de 3,4 indicando que los datos se alejan en promedio de la media en 3,4 preguntas buenas.

Julio disminuye su variación 1,6, siendo Javier el que menos variación presenta con 0,4 preguntas tanto por arriba como por debajo de la media aritmética. Sen recomienda al colegio elegir como ganador en este caso a Javier, presenta resultados más constantes que los otros dos alumnos.. Javier en promedio acierta 5 preguntas buenas con una variación muy baja.

La Desviación media para datos agrupados

$$D.M = \frac{\sum_{i=1}^{m} n_i |x_i - \bar{x}|}{n}$$

Donde:

n_i = Frecuencias absolutas simples

x_i = Marca de clase o puntos medios

$\bar{x}$ = Media Aritmética

n = Número total de observaciones

Ejemplo

Una máquina dispensadora de gaseosas está programada para llenar un envase con 350 centímetros cúbico (cc) de un refresco. A partir de una muestra de prueba realizada sobre 30 envases se realizó la siguiente distribución:

Tabla Nº 2.2

Intervalos	n_i
130 – 140	2
140 – 150	5
150 -160	14
160 – 170	4
170 – 180	4
180 - 190	1

Calcular e interpretar la desviación media

Solución

1) Se determina la marca de clase de cada intervalo
2) Se halla la media aritmética de los datos agrupados.
3) Se obtiene las desviaciones de las marcas de clase con respecto a la media
4) Se toma el valor absoluto de las desviaciones calculadas en (2) y se multiplican por sus respectivas frecuencias absolutas simples
5) Se suman los productos calculados en (3) y se divide por el número de observaciones.

Todos los pasos están reunidos en la tabla siguiente

Intervalos	n_i	x_i	$n_i \cdot x_i$	$x_i - \bar{x}$	$n_i \lvert x_i - \bar{x} \rvert$
130 – 140	2	135	270	22	44
140 – 150	5	145	725	12	60
150 -160	14	155	2170	2	28
160 – 170	4	165	660	8	32
170 – 180	4	175	700	18	72
180 - 190	1	185	185	28	28
Total	n = 30		4710		264

La desviación media es 8,8 cc. Concluimos que con datos suministrados de una muestra, el dispensador llenó los 30 envases con un promedio de 157 cc con una desviación media de 8, 8 cc. La desviación media describe un rango de dispersión promedio de llenado del dispensador ubicándolo entre 148, 2 cc y 165,8 cc.

2.3.2. Ventajas y Desventajas de la Desviación Media

a. **La desviación media es una mejor medida de la dispersión que la amplitud de variación porque toma en cuenta todas las observaciones en consideración. Pondera cada elemento e indica que tan lejos, en promedio, se encuentra cada observación de la media. Es menos** sensible a los valores extremos de los datos. Si es muy alta, indica gran dispersión; si es muy baja refleja un gran agrupamiento y que los valores son parecidos entre sí.

b. Promedia los valores absolutos de las desviaciones, esto es, que no reconoce el signo de las desviaciones.

2.4. Varianza

Es una medida estadística que mide la dispersión de los valores respecto a un valor central (media) es decir, la media de las diferencias cuadráticas de las puntuaciones respecto a su media aritmética se representa por σ^2, s^2, V(x).

2.4.1. Cálculo de la Varianza

La varianza para datos no agrupados

$$\sigma^2 = \frac{\sum_{i=1}^{m}\left(x_i - \bar{x}\right)^2}{n} \qquad \textbf{(Poblacional)}$$

$$s^2 = \frac{\sum_{i=1}^{m}\left(x_i - \bar{x}\right)^2}{n-1} \qquad \textbf{(Muestral)}$$

Donde:

x_i = Valor de cada observación o elemento

$\bar{x}$ = Media Aritmética

n = Número total de observaciones

Ejemplo

Las frecuencias cardiacas (pulsaciones por minuto) de 5 niños son:

130 132 127 129 132

Determinar la varianza de la frecuencia cardiaca.

Solución.

1) Se determina la media aritmética de la muestra
2) Se toma la diferencia entre cada observación y la media aritmética
3) Se eleva al cuadrado estas desviaciones
4) Se suman los cuadrados de las desviaciones
5) La suma se divide por n - 1

Entonces:

$$\bar{x} = \frac{130+132+127+129+132}{5} = \frac{650}{5} = 130$$

En seguida:

$$s^2 = \frac{(130-130)^2+(132-130)^2+(127-130)^2+(129-130)^2+(132-130)^2}{4} = \frac{18}{4} = 4{,}5$$

s^2 = 4,5 pulsaciones por minuto

La varianza para datos agrupados

$$\sigma^2 = \frac{\sum_{i=1}^{m} n_i.(x_i-\bar{x})^2}{n} \qquad \textbf{Poblacional}$$

$$s^2 = \frac{\sum_{i=1}^{m} n_i.(x_i-\bar{x})^2}{n-1} \qquad \textbf{Muestral}$$

Donde:

x_i = Marca de clase o punto medio

n_i = Frecuencia absoluta simple

$\bar{x}$ = Media Aritmética

n = Número total de observaciones

Ejemplo

El director de una institución educativa muestra la distribución de los pesos (Kg) de 300 escolares:

Tabla Nº 2.3

Peso (Kg)	20 - 25	25 - 30	30 - 35	35 – 40	40 - 45	45 - 50	50 - 55
Nº escolares	25	30	45	35	60	55	50

Calcular la varianza de los pesos.

Solución

1) Se determina la marca de clase de cada intervalo
2) Se encuentra la media aritmética de los datos agrupados
3) Se halla la diferencia entre cada marca de clase y la media aritmética
4) Se eleva al cuadrado las desviaciones y se multiplica por sus respectivas frecuencias absolutas simples
5) Se suman y dividen por n

Intervalo	x_i	n_i	$n_i \cdot x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$n_i (x_i - \bar{x})^2$
20 - 25	22,5	25	562,5	- 17,3	299,29	7482,25
25 - 30	27,5	30	825,0	- 12, 3	151,29	4538,7

30 - 35	32,5	45	1462,5	- 7,3	53,29	2398,05
35 – 40	37,5	35	1312,5	- 2,3	5,29	185,15
40 - 45	42,5	60	2550,0	2,7	7,29	437,4
45 - 50	47,5	55	2612,5	7,7	59,29	3260,95
50 - 55	52,5	50	2625,0	12,7	161,29	8064,5
Totales		n = 300	11950,0			26367,00

Entonces:

$$\bar{x} = \frac{11950}{300} = 39{,}8$$

Luego,

$$\sigma^2 = \frac{26367}{300} = 87{,}89 \textbf{ Kg}$$

2.4.2. Propiedades de la varianza

1. Para cualquier distribución la varianza es siempre una cantidad no negativa $\sigma^2 \geq 0$; $s^2 \geq 0$; esto es evidente, puesto que todas las observaciones positivas o negativas al elevarse al cuadrado se hacen positivas.
2. Si el valor de las observaciones son iguales, entonces la varianza es cero. En este caso las observaciones se confunden en un punto, la media es el mismo punto y la desviación es cero

$$\sigma^2 = \frac{0}{n} = 0\,;$$

3. La varianza de una constante es cero

V(k) = 0 k es una constante

$$V(k) = M\left\{[k - M(k)]^2\right\} = M\left\{[k-k]^2\right\} = M(o) = 0$$

4. La varianza del producto de una constante por una variable, es igual al cuadrado de la constante por la varianza de la variable.

$$V(k.Y) = k^2 V(Y)$$

$$V(k.Y) = M.\left\{[kY - M(k.Y)]^2\right\} = M\left\{[kY - kM(Y)]^2\right\}$$

$$V(k.Y) = M.\left\{k^2[Y - M(Y)]^2\right\} = k^2 M\left\{[Y - M(Y)]^2\right\}$$

$$V(k.Y) = k^2 V(Y)$$

5. La varianza de la suma de una variable más una constante es igual a la varianza de la variable.

$$V(Y+k) = V(Y)$$

$$V(Y+k) = M.\left\{[(Y+k) - M(Y+k)]^2\right\} = M\left\{[(Y+k) - M(Y) - k^2]\right\}$$

$$V(Y+k)=M.\{[Y-M(Y)]^2\}=V(Y)$$
$$V(Y+k)=V(Y)$$

Ejemplo

Se duplican los sueldos de los 80 trabajadores. ¿Cuál es la nueva varianza y el sueldo promedio?. Sabe que: $\overline{Y}=1215$ y $V(Y)=1460$.

Solución

Al duplicarse los sueldos k = 2

$$V(2Y)=2^2V(Y)=4(1460)=S/.5840$$ es la nueva varianza

El nuevo sueldo promedio.

$$M(2Y)=2M(Y)=2(1215)=S7.2430$$

Ejemplo.

Si a cada uno de los 80 trabajadores se incrementa su sueldo en 60 soles mensuales. ¿Cuál sería la nueva varianza y el sueldo promedio?

Solución

Si K = 60, entonces $V(Y+60)=V(Y)=1460$, es decir un incremento constante a cada elemento no altera la dispersión de la distribución. Por otra parte, el nuevo sueldo promedio quedaría incrementado en 60 soles.

$$M(Y+60)=M(Y)+60=1215+60=S/.1275$$

2.4.3. Componentes de la Varianza

Si un conjunto de datos se divide en subconjuntos, categorías o estratos es posible descomponer la varianza en dos componentes.

Supongamos que un conjunto de datos ha sido dividido en L estratos o subconjuntos, cada estrato tendrá un tamaño (n_r), su respectiva media aritmética $(\overline{x_r})$ y varianza (s_r^2), valores que expresa la importancia de cada uno de los estratos en el total del conjunto.

La variabilidad total puede estar afectada por:

a) Las variaciones dentro de cada estrato, esta variación interna en cada estrato se llama la intravarianza (s_w^2),

b) Las varianzas entre los diferentes estratos se denomina intervarianza (s_b^2), De este, modo la varianza total se expresa como $s^2 = (s_w^2) + (s_b^2)$,

A. Intravarianza: $(s_w^2) = M.\ (s_r^2)$

Es el estadígrafo que expresa las variaciones dentro de los estratos, y se define como el promedio de las varianzas de los estratos

$$s_w^2=M(s_r^2)=\frac{\sum s_r^2 n_r}{n}$$

Donde:

s_w^2 = Varianza del estrato r

n_r = Tamaño de cada estrato

B. Intervarianza: $s_b^2 = V(\overline{x}_r)$

Es el estadígrafo que expresa la variación entre los estratos, se define como la varianza entre las medias de los estratos.

$$s_b^2 = V(\overline{Y}_r) = \frac{\sum(\overline{Y}_r - \overline{Y})^2 n_r}{n} = \frac{\sum \overline{Y}_r^2 n_r}{n} - \left(\frac{\sum \overline{Y}_n n_r}{n}\right)^2$$

Donde:

$\overline{Y}_r$ = Media aritmética del estrato r

$\overline{Y}$ = Media aritmética total

n_r = Tamaño del estrato r

Para estratos de tamaño n_1 y n_2 que tienen medias iguales y varianzas s_1^2 y s_2^2 respectivamente, la varianza está dada por: $s^2 = \dfrac{s_1^2.n_1 + s_2^2.n_2}{n}$

Ejemplo

En una empresa, los trabajadores están clasificados en Directivos, Empleados y Obreros, de los cuales se conocen los siguientes datos:

Tabla Nº 2.4

Categorías	Nº trabajadores n_r	Sueldo promedio $\overline{Y}_r$	Varianza (s_r^2)
Directivos	5	3200	12 000
Empleados	15	1500	6 000
Obreros	25	900	28 000

¿Cuál es la varianza para la totalidad de los trabajadores?

Solución

Categorías	n_r	$\overline{Y}_r$	s_r^2	$s_r^2.n_r$	$\overline{Y}_r.n_r$	$\overline{Y}_r^2.n_r$
Directivos	5	3200	12000	60000	16000	51200000
Empleados	15	1500	6000	900000	22500	33750000
Obreros	25	900	28000	700000	22500	20250000
Total	45			850000	61000	105200000

La varianza total es: $s^2 = (s_w^2) + (s_b^2)$,

$$s_w^2 = M(s_r^2) = \frac{\sum s_r^2.n_r}{n} = \frac{850000}{45} = 18888{,}88$$

$$s_b^2 = V(\overline{Y}_r) = \frac{\sum \overline{Y}_r^2 n_r}{n} - \left(\frac{\sum \overline{Y}_n n_r}{n}\right)^2 = \frac{105200000}{45} - \left(\frac{61000}{45}\right)^2 = 2319403{,}976$$

$$s^2 = (18888{,}88) + (2319403{,}976) = 2338292{,}956$$

Como $s_b^2 > s_w^2$, entonces se puede concluir que la variación o dispersión total se debe principalmente a la variación de los sueldos entre los estratos o categorías de trabajadores.

2.4.4. Métodos Abreviados de Cálculo de la Varianza

Método 1.

1. Se elige un origen de trabajo O_t que generalmente es el valor central x_i de mayor frecuencia absoluta simple
2. Se determina las desviaciones de las variables x_i, respecto al origen de trabajo, es decir: $d_i = x_i - O_i$; i = 1, 2, …, m
3. Se determinan las varianzas de "d"

$$s_x^2 = V(X) = \frac{1}{n-1}\left[\sum_{i=1}^{m} d_i^2.n_i - \frac{\left(\sum_{i=1}^{m} d_i n_i\right)^2}{n}\right]$$

Ejemplo

Para la distribución de frecuencias siguiente:

Tabla Nº 2.5

Intervalo	0 - 20	20 - 60	60 - 80	80 - 100	100 - 140
Frecuencia	20	15	10	10	5

Calcular la varianza.

Solución

1. Calculamos las marcas de clase de la distribución
2. Elegimos el origen de trabajo $O_t = 70$
3. Calculamos en la tabla todos los valores que se necesitan sustituir en la primera fórmula del Método 1.

Intervalo	x_i	n_i	$d_i = x_i - O_t$	$d_i \, . n_i$	$d_i^2.n_i$
0 - 20	10	20	-60	-1200	72000
20 - 60	40	15	-30	-450	13500
60 - 80	**70**	10	0	**- 1650 | 450**	0
80 - 100	90	10	20	200	4000
100 - 140	120	5	50	250	12500

Totales		60		1200	102000

$$s^2 = \frac{1}{59}\left[102000 - \frac{(-1200)^2}{60}\right] = \frac{1}{59}[102000 - 24000] = \frac{78000}{59} = 1322{,}034$$

Método 2. (Método Codificado o Tipificado)

Aplicable sólo para datos agrupados en clase de amplitud constante

1. Se elige un origen de trabajo O_t
2. Se transforma las observaciones mediante la expresión $\mu_i = \dfrac{x_i - O_t}{c}$

Donde:

x_i = Marca de clase

c = Amplitud de clase

3. Se determina la varianza

$$s^2 = V(\mu) = \frac{c^2}{n-1}\left[\sum_{i=1}^{m} n_i \mu_i^{\,2} - \frac{\left(\sum_{i=1}^{m} n_i \mu_i\right)^2}{n}\right]$$

Ejemplo

Para la distribución de frecuencias siguiente:

Tabla Nº 2.6

Intervalo	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
Frecuencia	10	15	20	10	5

Calcular la varianza

1) Hallemos las marcas de clase de la distribución
2) Se elige el origen de trabajo $O_t = 50$ y c = 20
3) Calcular en una tabla todos los valores que necesitamos sustituir en la fórmula del paso (3).

Intervalo	x_i	n_i	$\mu_i = \dfrac{x_i - 50}{20}$	$n_i\ \mu_i$	$n_i.\mu_i^2$
0 - 20	10	10	-2	-20	40
20 - 40	30	15	-1	-15	15
40 - 60	**50**	20	0	**- 35 | 20**	0
60 - 80	70	10	1	10	10
80 - 100	90	5	2	10	20
Totales		60		-15	85

$$s^2 = \frac{(20)^2}{59}\left[85 - \frac{(-15)^2}{60}\right] = \frac{400}{59}[85 - 3{,}75] = 550{,}847$$

2.5. La Desviación Estándar

Denominada también desviación típica es la raíz cuadrada positiva de la varianza, trata de la dispersión de los datos respecto al valor de la media, cuanto mayor sea su valor, más dispersos estarán los datos. Es una de las medidas de dispersión de mayor uso, en el cual las unidades de la variable ya no están elevadas al cuadrado. Se representa por s, σ; $D(x)$.

2.5.1. Cálculo de la desviación Estándar

La Desviación Estándar para datos no agrupados

$$\sigma = \sqrt{\frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{n}}$$ **(Poblacional)**

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

Donde:

x_i = Valor de cada observación

$\bar{x}$ = Media aritmética

n_i = Número total de observaciones

La desviación Estándar para datos agrupados

$$\sigma = \sqrt{\frac{\sum_{i=1}^{m}n._i(x_i - \bar{x})^2}{n}}$$ **(Poblacional)**

$$s = \sqrt{\frac{\sum_{i=1}^{m}n._i(x_i - \bar{x})^2}{n-1}}$$ **(Muestral)**

Donde:

x_i = Marca de clase o punto medio

$\bar{x}$ = Media aritmética

n_i = Frecuencia absoluta simple

n = Número total de observaciones

Ejemplo

Considerando los trabajadores sobre la distribución de sus sueldos en la Tabla Nº 2.4

Categorías	Media $\bar{y}$	Varianza s^2	Desviación Estándar s

Directivos	3200	12 000	110
Empleados	1500	6 000	77
Obreros	900	28 000	167

Se deduce que:

$$s(empleados) < s(directivos) < s(obreros)$$

Significa que los sueldos de los obreros presentan mayor dispersión que los sueldos de los directivos y éstos mayor dispersión que los sueldos de los empleados. Encontrar una mayor dispersión también significa que existe una mayor discrepancia entre los sueldos mayores y menores.

2.5.2. Propiedades de la Desviación Estándar

- La desviación estándar de un conjunto de observaciones $x_1, x_2, ..., x_n$, siempre es una cantidad no negativa o cero, en el caso que las puntuaciones sean iguales $D(x) = s = \sigma \geq 0$
- La desviación estándar de una constante es cero $D(k) = 0$
- Si cada valor de las observaciones $x_1, x_2, ..., x_n$, se le suma o resta una constante b, la desviación estándar del nuevo conjunto de valores $y_1, y_2, ..., y_n$ donde $y_i = x_i \pm b$ con i = 1, 2,.., n coincide con la desviación estándar del conjunto original. Es decir $D(y) = D(x \pm b) = D(x)$.
- Si cada valor de las observaciones $x_1, x_2, ..., x_n$, se multiplica por una constante k, la desviación estándar del nuevo conjunto de valores $y_1, y_2, ..., y_n$ donde $y_i = kx_i$ con i = 1, 2,.., n es igual a la desviación estándar del conjunto original multiplicado por el valor positivo (Valores absolutos) de dicha constante. Es decir, $D(y) = D(kx) = |k| D(x)$
- Si tenemos varias distribuciones con la misma media y conocemos sus respectivas desviaciones estándar se puede calcular la desviación estándar total. Si todas las muestras tienen el mismo tamaño:

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

$$s = \sqrt{\frac{s_1^2 + s_2^2 + ... + s_n^2}{n}}$$

Si las muestras tienen distinto tamaño:

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

$$s = \sqrt{\frac{k_1 s_1^2 + k_2 s_2^2 + ... + k_n s_n^2}{k_1 + k_2 + ... + k_n}}$$

2.5.3. Ventajas de la Desviación Estándar

1) La desviación estándar es sin duda, la medida de dispersión que posee una mayor estabilidad frente a las fluctuaciones de la muestra tomada.
2) Se basa en todos los valores de la variable, tanto en su magnitud como en su signo.
3) Su estudio es indispensable cuando se trata de interpretar datos en relación con la distribución normal

2.5.4. Desventajas de la Desviación Estándar

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

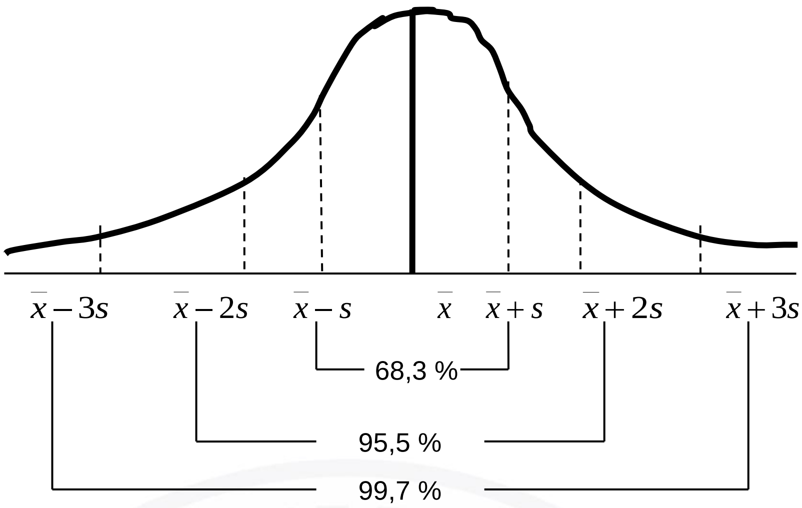
2.5.5. Aplicaciones de la Desviación Estándar

La varianza y la desviación estándar son las medidas de dispersión más frecuentemente usadas; sin embargo no tienen una interpretación muy clara, para poder superar esta dificultad, presentaremos una regla que puede describirse con precisión, la dispersión de una distribución simétrica, cuyo polígono tiene forma de campana, y que describe razonablemente bien la dispersión de otras distribuciones de datos que tienen esta forma. La frecuente ocurrencia de distribuciones acampanadas u otras formas similares a ésta en la naturaleza justifican la importancia práctica de la siguiente regla:

Dada una distribución de un conjunto de observaciones de tamaño n suficientemente grande que es simétrica y forma aproximadamente acampanada. Entonces, el 68,3 % de las observaciones estarán comprendidas dentro del intervalo $\bar{x} - s$ y $\bar{x} + s$.

El 95,5 % de las observaciones estarán comprendidas dentro del intervalo $\bar{x} - 2s$ y $\bar{x} + 2s$. El 99,7 % de las observaciones estarán comprendidas dentro del intervalo $\bar{x} - 3s$ y $\bar{x} + 3s$

Figura Nº 2.1



Ejemplo

Un fabricante recibe un pedido de cojinetes para una operación específica. Con el fin de verificar, si los diámetros de los cojinetes que fabrica cumple las especificaciones para tal operación, Se miden los diámetros de 50 cojinetes y se encuentra que la media es de 0,5354 cm y la desviación estándar de 0,00525 cm. Describir los datos

Solución

Aunque no tenemos información previa acerca de la distribución de los datos, hay una buena posibilidad que tenga la forma de campana o algo parecido, por lo tanto la regla empírica proporcionaría una buena descripción de los datos. Entonces:

$$\bar{x}-s=0{,}5354-0{,}00525=0{,}53015$$

$$\bar{x}+s=0{,}5354+0{,}00525=0{,}54065$$

$$\bar{x}-2s=0{,}5354-2(0{,}00525)=0{,}5249$$

$$\bar{x}+2s=0{,}5354+2(0{,}00525)=0{,}5459$$

$$\bar{x}-23=0{,}5354-3(0{,}00525)=0{,}51965$$

$$\bar{x}+23=0{,}5354+3(0{,}00525)=0{,}55115$$

De modo que, entre 0,53015 y 0,54065 estarán aproximadamente el 68,3% de los datos; entre 0,5249 y 0,5459 estarán aproximadamente el 95,5 % de los datos, entre 0,51965 y 0,55115 estarán aproximadamente el 99,7 % de los datos.

2.6. Coeficiente de Variación

Es una medida que se utiliza para comparar dos o más distribuciones cuando las unidades de la medida de las variables están expresadas en diferentes unidades o escalas de medida. Esta definida por la expresión:

$$C.V=\frac{Desviaciónestándar}{MediaAritmética}$$

$$C.V = \frac{s}{\bar{x}}$$

Comparando dos más distribuciones, es más homogénea o presenta menos dispersión aquella distribución que tiene el menor coeficiente de variación. En otras palabras, los datos son más heterogéneos cuando tienen mayor coeficiente de variación.

Ejemplo

En dos pruebas de conocimiento A y B, la prueba A se califico sobre 100 puntos, la media aritmética de las calificaciones fue de 72 puntos con una desviación estándar de 9 puntos. La prueba B se calificó sobre 80 puntos y los resultados dieron una media de 52 puntos con una desviación estándar de 6. Determinar en cuál de las dos pruebas hubo mayor variación

Solución

$$CV_A = (\frac{s_A}{\bar{x}_A}).100 = (\frac{9}{72}).100 = 12{,}5\%$$

$$CV_B = (\frac{s_B}{\bar{x}_B}).100 = (\frac{6}{52}).100 = 11{,}5\%$$

En conclusión, la prueba de conocimiento B tiene menor variación en los puntajes.

CAPITULO III

MEDIDAS DE ASIMETRÍA

Son indicadores que permiten establecer el grado de simetría o asimetría que presenta una distribución de frecuencia o variable aleatoria. La asimetría resulta útil en muchos campos, existen modelos que asumen una distribución normal, que tienen una asimetría cero.

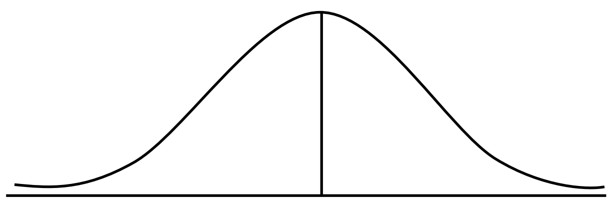
Pero en realidad, los valores no son nunca perfectamente simétricos y la asimetría de la distribución proporciona una idea sobre las desviaciones de la media son positivas o negativas.

3.1. Simetría

Cuando la media aritmética, mediana y moda son iguales

Figura Nº 3.1

Distribución simétrica



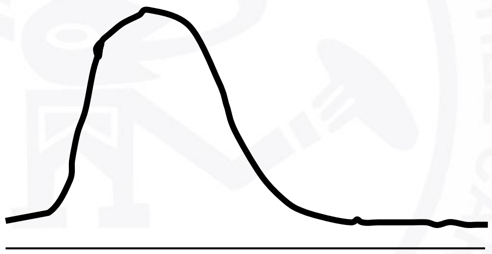
$$\bar{x} = Me = Mo$$

3.2. Asimetría positiva

Es aquella que la ramificación es más extendida hacia la derecha o hacia los valores grandes de la variable.

Figura Nº 3.2

Asimétrica positiva

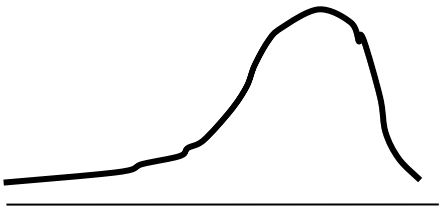


3.3. Asimetría negativa

Es aquella que la ramificación es más extendida hacia la izquierda o hacia los valores pequeños de la variable.

Figura Nº 3.3

Asimétrica negativa



3.4. Coeficiente de Asimetría

- **Coeficiente de Asimetría Karl Pearson**

$$AS_1 = \frac{\bar{x} - Mo}{S}$$

$$AS_2 = \frac{3(\bar{x} - Me)}{S}$$

Que constituyen respectivamente, el primer y segundo coeficiente de asimetría de Pearson

- **Coeficiente de Asimetría de Arthur Boeley**

$$AS = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

Que se llama "coeficiente cuartil de deformación"

3.5. Interpretación de los coeficientes de Asimetría

Si AS > 0, tiene asimetría positiva. La distribución extiende la cola hacia los valores grandes de la variable.

Si AS < 0, tiene asimetría negativa. La distribución extiende la cola hacia los valores pequeños de la variable.

Si AS = 0, La distribución es simétrica.

Por otra parte, también se deduce que hay asimetría cuando:

$Mo < Me < \bar{x}$ Asimetría positiva

$\bar{x} < Me < Mo$ Asimetría negativa

Ejemplo

Determinar el coeficiente de asimetría para la distribución siguiente

Intervalos	50 - 60	60 - 70	70 - 80	80 - 90	90 - 100
n_i	15	20	30	20	15

Solución

Para calcular la media, moda, cuartiles y desviación estándar construimos la siguiente tabla

Intervalos	n_i	x_i	$x_i\ n_i$	N_i	$\mu_i = \frac{x_i - 75}{10}$	$n_i\ \mu_i$	$n.\mu_i^2{}_i$
50 - 60	15	55	825	15	- 2	- 30	60
60 - 70	20	65	1300	35	- 1	- 20	20
70 - 80	**30**	**75**	**2250**	65	0	**0**	0
80 - 90	20	85	1700	85	1	20	20
90 - 100	15	95	1425	100	2	30	60
Total	100		7500			0	160

Media aritmética: $\bar{x} = \frac{7500}{100} = 75$

Moda: $Mo. = 70 + \left[\frac{(30-20)}{(30-20)+(30-20)} \right].10 = 75$

Cuartiles:

$$Q_1 = 60 + \left(\frac{25-15}{20} \right).10 = 65$$

$$Q_2 = 70 + \left(\frac{50-35}{30} \right).10 = 75$$

$$Q_3 = 80 + \left(\frac{75-65}{20} \right).10 = 85$$

Varianza:

$$s^2 = \frac{(10)^2}{99} \left[160 - \frac{0}{100} \right] = \frac{100}{99}[160] = 161{,}6$$

Desviación Estándar: $s = \sqrt{161{,}6} = 12{,}71$

Coeficiente de Asimetría de Karl Pearson

$$AS_1 = \frac{75-75}{12{,}71} = 0 \qquad AS_2 = \frac{3(75-75)}{12{,}71} = 0$$

Coeficiente de Asimetría de Arthur Bowley

$$AS = \frac{85-2(75)+65}{85-65} = 0$$

Por lo tanto, la distribución es simétrica

Ejemplo

En la distribución de los sueldos de 80 trabajadores, se conoce $\bar{x} = 1200$, $Me = 1195$; $Mo = 1150$, $Q_1 = 1200$; $Q_3 = 1300$, $S = 85{,}60$ ¿Cuál es la asimetría?

Solución

$$AS_1 = \frac{1200-1150}{85{,}60} = 0{,}58 > 0$$

$$AS_2 = \frac{3(1200-1195)}{85{,}60} = 0{,}058 > 0$$

$$AS = \frac{1300-2(1195)+1200}{1300-950} = 0{,}31 > 0$$

En cualquiera de las fórmulas AS es positivo, por lo tanto, tiene Asimetría Positiva, es decir; hay un ligero predominio de sueldos menores. Frecuentemente, la distribución de los salarios tiene asimetría positiva, porque existen muchos trabajadores que su sueldos son bajos y pocos trabajadores que ganan bien.

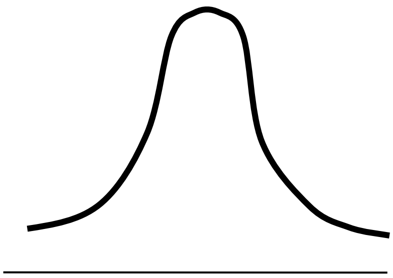
CAPITULO I V

MEDIDAS DE APUNTAMIENTO

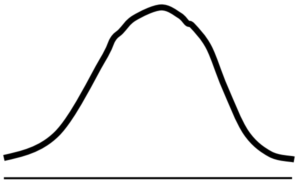
Es el grado de apuntamiento de una distribución, se analiza comparando la distribución con la forma de una curva normal o simétrica.

Si una distribución tiene relativamente un elevado pico o apuntamiento se llama Leptokúrtica; mientras que si es achatada se denomina Platikúrtica y si es moderadamente apuntada se dice que es Mesokúrtica.

Figura Nº 4.1



Leptokúrtica



Mesokúrtica

Platikúrtica

4.1. Coeficiente de Kurtosis

- **Kurtosis en función de momentos**

$$K_1 = \frac{m_4}{s^4}$$

Donde: $s^4 = \left(s^2\right)^2$

$$m_4 = \frac{\sum\left(x_i - \bar{x}\right)^2 . n_i}{n}$$

Interpretación

Si $K_1 > 3$, es Leptokúrtica. La distribución es más apuntada que la normal.

Si $K_1 = 0$, es Mesokúrtica. La distribución es moderadamente apuntada

Si $K_1 < 0$, es Platikúrtica. Su distribución es menos apuntada que la normal o sea achatada.

- **Kurtosis en función de cuantiles**

$$K_2 = \frac{Q_3 - Q_1}{2\left(P_{90} - P_{10}\right)}$$

Interpretación

Si $K_2 = 0{,}263$, es Mesokúrtica (Normal)

Si $K_2 < 0{,}263$, es Platikúrtica (achatada).

Si $K_2 > 0{,}263$, es Leptokúrtica (apuntada).

Ejemplo

Se han medido las pulsaciones de un equipo de atletas después de una carrera. Los datos obtenidos son.

Tabla Nº 4.1

Pulsaciones	70 -75	75 - 80	80 - 85	85 - 90	90 - 95	95 - 100
Nº de atletas	3	3	7	10	12	8

Analizar el tipo de apuntamiento de la distribución.

Solución

Intervalos	n_i	x_i	$x_i n_i$	$(x_i-\bar{x})^2$	$n_i(x_i-\bar{x})^2$	$(x_i-\bar{x})^4$	$n_i(x_i-\bar{x})^4$	N_i
70 -75	3	72,5	217,5	240,25	720,75	57720,0625	173160,1875	3
75 - 80	3	77,5	232,5	110,25	330,75	12155,0625	36465,1875	6
80 - 85	7	82,5	577,5	30,25	211,75	915,0625	6405,4375	13
85 - 90	10	87,5	875,5	0,25	2,50	0,0625	0,625	23
90 - 95	12	92,5	1110,0	20,25	243,00	410,0625	4920,75	35
95 - 100	8	97,5	780,0	90,25	722,00	8145,0625	65160,5	43
	43		3792,5		2230,75		2861112,6875	

Encontramos los datos siguientes:

$$\bar{x}=\frac{3792{,}5}{43}=88 \qquad s^2=\frac{2230{,}75}{42}=53{,}113$$

$$s^4=\left(s^2\right)^2=(53{,}113)^2=2820{,}991 \qquad Q_1=80+\left(\frac{10{,}75-6}{7}\right).5=83{,}393$$

$$Q_3=90+\left(\frac{32{,}25-23}{12}\right).5=93{,}854 \qquad P_{10}=75+\left(\frac{4{,}3-3}{3}\right).5=77{,}167$$

$$P_{90}=95+\left(\frac{38{,}7-35}{8}\right).5=97{,}313 \qquad m_4=\frac{286112{,}6875}{43}=6653{,}783$$

- **Coeficiente de Kurtosis en función de momentos**

$$k_1=\frac{m_4}{s^4}=\frac{6653{,}783}{2820{,}991}=2{,}36$$

Como $k_1=2{,}36<3$ la distribución es Platikúrtica su forma es achatada.

- **Coeficiente de Kurtosis en función de cuantiles**

$$k_2=\frac{Q_3-Q_1}{2\left(P_{90}-P_{10}\right)}=\frac{93{,}854-83{,}393}{2(97{,}313-77{,}167)}=0{,}26$$

Como $k_2=0{,}26<0{,}263$ la distribución es Platikúrtica su forma es achatada

Problemas Propuestos

1. En la siguiente distribución de frecuencias se dan los pesos de una muestra de 45 alumnos

Pesos en Kgr	40 – 45	45 - 50	50 - 55	55 - 60	60 - 65	65 -70
Nº de alumnos	4	10	15	8	5	3

   a) Determinar la varianza por el método abreviado
   b) Hallar el coeficiente de variación
   c) ¿La distribución es simétrica?
   d) ¿La distribución es mesokúrtica?

2. Un encargado de compras ha obtenido muestras de focos de luz de dos proveedores. En su propio laboratorio, ha probado ambas muestras con respecto a la duración de su vida útil, con los siguientes resultados**.**

Duración de vida útil (en horas)	Muestras de la	
	Empresa A	Empresa B
700 – 900	10	3
900 – 1100	16	42
1100 -1300	26	12
1300 -1500	8	7
1500 -1700	10	6
Total	70	70

   a) ¿Los focos de qué empresa tienen el mayor promedio en la duración de su vida útil?
   b) ¿Los focos de cuál de las empresas tienen mayor uniformidad?
   c) Calcular el coeficiente de variación. Comente los resultados.

3. Una publicidad sobre camiones enumera la siguiente distribución de kilómetros recorridos por galón de gasolina según reportes de los propietarios de esos vehículos.

Kms por galón	15 -18	18 - 21	21 -24	24 - 27	27 - 30	30 - 33	33 - 36
Porcentaje	6	10	16	24	14	18	12

   a) Determinar la desviación media y la desviación estándar.
   b) Hallar el coeficiente de variación
   c) Analizar la asimetría de la distribución
   d) Calcular los coeficientes de apuntamiento.

4. Las secciones A, B y C de la asignatura de Cálculo de Probabilidades rinden el mismo examen parcial. Los resultados obtenidos se registran en las siguientes tablas

Sección A	
x_i	n_i
2,5	3

7,5	8
12,5	22
17,5	30

Sección B	
Intervalos	$x_i\ n_i$
2 – 4	16
4 – 6	100
6 – 8	85
8 – 10	270
10 -12	55
12 -14	26

Sección C	
H_i	$h_i\ x_i^2$
0,10	2,5
0,20	10,0
0,80	86,4
1,00	45
Nº de alumnos en la sección C= 60	

a) El profesor de la sección A sostiene que la suya es mejor. ¿Es correcta la afirmación del profesor?. ¿Por qué?

b) ¿En cuál de las secciones las notas son más homogéneas?

5. La desviación horizontal respecto al blanco en pruebas de los cohetes es la siguiente (en m)

Intervalos	0 -10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
Frecuencias	50	20	15	35	60	20

Determinar los coeficientes de asimetría y apuntamiento

6. El coeficiente de variación de los sueldos de 200 trabajadores de una empresa es 0,57. Después de reajustar todos los sueldos en S/. 120, este coeficiente de variación,

es ahora de 0,50. Sin embargo, la gerencia fija un sueldo mínimo de S/. 550, lo que beneficia a 35 personas, que antes del reajuste ganaban menos de S/.480, con un sueldo medio de S/. 450 mensual. Determinar la cantidad de dinero que necesitará mensualmente la empresa para pagar los sueldos después de hacer efectivos los reajustes.

7. Se han elegido 150 productos para analizar sus pesos en gramos, los resultados están clasificados en la siguiente tabla.

Intervalos	2.00 – 2,04	2,04 -2,,08	2,08 – 2,12	2,12 -2, 16	216 – 2,20	2,20 – 2,24
n_i	12	20	38	50	16	14

a) Calcular la desviación media,, varianza, desviación estándar y coeficiente de variación.

b) Determinar los coeficientes de asimetría y apuntamiento

8. El contador de una cadena de supermercados desea analizar los saldos (en miles de S/.) de las cuentas de crédito.

Intervalos	0,2 – 0,4	0,4 – 0,6	0,8 -1,0	1,0 -1,2	1,2 – 1,4	1,4 – 1,6	1,6 – 1,8
n_i	20	10	15	5	30	25	15

a) Determinar la varianza y el coeficiente de variación

b) Hallar los coeficientes de asimetría y apuntamiento

RESUMEN

El presente trabajo de investigación titulado: "ESTADÍSTICA DESCRIPTIVA", tiene como finalidad poner al alcance de estudiantes de las diversas carreras profesionales de nuestra Universidad, un material bibliográfico, útil y práctico, que será utilizado en el transcurso de su formación profesional.

El contenido ha sido sistematizado con el propósito que estudiantes y lectores logren una mayor comprensión y aplicación de la Estadística, utilizando un vocabulario sencillo y preciso, reforzado con ejemplos y problemas prácticos facilitando el proceso de enseñanza aprendizaje.

El trabajo está estructurado en cuatro capítulos. El primer capítulo, dispone de medidas que representan valores centrales en los cuales se agrupan las observaciones y que son de utilidad en el análisis de una distribución y en la comparación de distribuciones. El segundo capítulo, trata de medir el grado de concentración de las observaciones alrededor de un valor central. El tercer capítulo, estudia las direcciones de la dispersión de los datos con respecto al centro de la distribución y el cuarto capítulo, analiza la distribución comparando con la forma de la curva normal y se puede medir en función de momentos o de cuantiles.