1.- Las calificaciones de 50 alumnos en Matemáticas han sido las siguientes:

5, 2, 4, 9, 7, 4, 5, 6, 5, 7, 7, 5, 5, 2, 10, 5, 6, 5, 4, 5, 8, 8, 4, 0, 8, 4, 8, 6, 6, 3, 6, 7, 6, 6, 7, 6, 7, 3, 5, 6, 9, 6, 1, 4, 6, 3, 5, 5, 6, 7.

Calcular: (2.5 puntos)

a) La **moda, mediana y media**.
b) **Varianza y desviación estándar.**

En la tabla se muestran todos los cálculos en forma columnar

x_i	f_i	Fi	$f_i x_i$	$(xi-\bar{x})^2$	$f_i(xi-\bar{x})^2$
0	1	1	0	30,0304	30,0304
1	1	2	1	20,0704	20,0704
2	2	4	4	12,1104	24,2208
3	3	7	9	6,1504	18,4512
4	6	13	24	2,1904	13,1424
5	11	24	55	0,2304	2,5344
6	12	36	72	0,2704	3,2448
7	7	43	49	2,3104	16,1728
8	4	47	32	6,3504	25,4016
9	2	49	18	12,3904	24,7808
10	1	50	10	20,4304	20,4304
Suma	50		274		198,48

De allí que:

a)

La moda es el dato que se presenta con mayor frecuencia y es el número 6 cuya frecuencia es 12.

Mediana: Ya que el número de datos es $n = 50$ la medianaes el promedio de las datos que se encuentra en las posiciones $\frac{n}{2} = \frac{50}{2} = 25$ y 26

Observando las frecuencias acumuladas F_i vemos que la mediana es el número 6 que ocupa las posiciones 25 y 26 (hasta la 36) en la lista ordenada.

La media a partir de datos ya calculados en la tabla es:

$$\bar{x} = \frac{\sum f_i x_i}{n} = \frac{274}{50} = 5{,}48$$

b) La varianza calculada a partir de la tabla como

$$s^2 = \frac{\sum f_i(x_i - \bar{x})^2}{n} = \frac{198{,}48}{50} = 3{,}9696$$

La desviación estándar es:

$$s = \sqrt{s^2} = \sqrt{3{,}9696} = 1{,}99$$

2.- Las calificaciones de los alumnos en un examen de Estadística han sido: 6, 4, 4, 3, 6, 10, 1, 0, 2, 6, 6, 8, 5 (5.5 puntos)

a) a. Calcular la media aritmética, la moda y la mediana.
b) b. Si usted fuese un líder estudiantil, ¿qué medida de centralidad escogería para argumentar la buena "calidad"del grupo? Por qué?
c) c. Si usted fuese el profesor de la materia, ¿qué medida de centralidad escogería para argumentar la pésima "calidad"del grupo? Por qué?
d) d. Si usted fuese un observador imparcial, ¿qué podría decir sobre el nivel del grupo? Por qué?

En la tabla se muestran todos los cálculos en forma columnar

x_i	f_i	Fi	$f_i x_i$
0	1	1	0
1	1	2	1
2	1	3	2
3	1	4	3
4	2	6	8
5	1	7	5
6	4	11	24
8	1	12	8
10	1	13	10
Suma	13		61

a)

<u>Media ariмética</u>

$$\bar{x} = \frac{\sum f_i x_i}{n} = \frac{61}{13} = 4{,}69$$

<u>Moda</u>

La moda es 6 que es el dato con mayor frecuencia (4)

Mediana

Ya que el número de datos es $n = 13$, impar, la mediana es el dato que ocupa la posición el $\frac{n+1}{2} = \frac{14}{2} = 7$.

Observando la columna de las frecuencias acumuladas concluimos que la mediana es el número 5.

b)

Para estudiar la validez de la media o de la varianza estudiaremos la desviación estándar completando la tabla

x_i	f_i	Fi	$f_i x_i$	$(xi-\bar{x})^2$	$f_i(xi-\bar{x})^2$
0	1	1	0	22,02	22,02
1	1	2	1	13,63	13,63
2	1	3	2	7,25	7,25
3	1	4	3	2,86	2,86
4	2	6	8	0,48	0,96
5	1	7	5	0,09	0,09
6	4	11	24	1,71	6,84
8	1	12	8	10,94	10,94
10	1	13	10	28,17	28,17
Suma	13		61		92,77

Calculamos la varianza

$$s^2 = \frac{\sum f_i(x_i - \bar{x})^2}{n} = \frac{92,77}{13} = 7,14$$

La desviación estándar es:

$$s = \sqrt{s^2} = \sqrt{7,14} = 2,67$$

El coeficiente de variación es

$$CV = \frac{s}{\bar{x}} = \frac{2,67}{4,69} = 0,57$$

Como el coeficiente de variación es alto, la media no es un buen representante de las calificaciones y es mas baja que la mediana que es 5 y está más cercana a la moda que es 6. Además la mediana es más alta justificando la buena calidad del grupo. Por supuesto que la moda de valor 6 sería la que más le convendría al estudiante.

c) El profesor para justificar la "mala calidad" del grupo escogería la media (4,69)

d) Hay dificultades para el observador imparcial dada la dispersión de los datos en los extremos que podríamos llamar atípicos y al alto coeficiente de variación. En esos casos es preferible recurrir a la mediana (5)

4.- Se ha aplicado un test a los empleados de una fábrica, obteniéndose la siguiente tabla:

Puntuación	f_i
[38, 44)	7
[44, 50)	8
[50, 56)	15
[56, 62)	25
[62, 68)	18
[68, 74)	9
[74, 80)	6

Calcular: (7 puntos)

a) La **moda, mediana y media**. Interprete los resultados
b) **Desviación estándar.** Interprete los resultados
c) Cuartil 3, decil 5 y percentil 58. Interprete los resultados

Primero consignamos los datos es un tabla incluyendo algunos cálculos

i	Li	Ls	Xi	fi	Fi	fi*Xi
1	38	44	41	7	7	287
2	44	50	47	8	15	376
3	50	56	53	15	30	795
4	56	62	59	25	55	1475
5	62	68	65	18	73	1170
6	68	74	71	9	82	639
7	74	80	77	6	88	462
			Suma	88		5204

Li: Límite inferior de la clase

Ls: Límite superior de la clase

Xi: Marca de clase o punto medio

fi: Frecuencia de la clase

Fi: Frecuencia acumulada

a)

Moda:

La clase modal es la clase [56,62] con la mayor frecuencia que es 25

Para calcular la Moda a partir de la tabla

i	Li	Ls	Xi	fi	Fi	fi*Xi
1	38	44	41	7	7	287
2	44	50	47	8	15	376
3	50	56	53	15	30	795
4	56	62	59	25	55	1475
5	62	68	65	18	73	1170
6	68	74	71	9	82	639
7	74	80	77	6	88	462
			Suma	88		5204

Usamos la fórmula:

$$M_o = L_i + IC\left[\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})}\right]$$

Donde:

$L_i = 56$ Es el límite inferior de la clase de la moda

$IC = 6$ Intervalo de clases

$f_i = 25$ Frecuencia absoluta de la clase de la moda

$f_{i-1} = 15$ Frecuencia absoluta de la clase anterior a la clase de la moda

$f_{i+1} = 18$ Frecuencia absoluta de la clase posterior a la clase de la moda

$$M_o = L_i + IC\left[\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})}\right]$$

$$M_o = 56 + 6\times\left[\frac{25-15}{(25-15)+(25-18)}\right] = 56 + 6 \times \left[\frac{10}{10+7}\right] = 56 + 6 \times \left[\frac{10}{10+7}\right] = 59{,}53$$

Mediana:

La posición de la mediana por ser un número de datos impar estaría entre las posiciones

$$\frac{n}{2} = \frac{88}{2} = 44 \text{ y } 45.$$

Mirando las frecuencias acumuladas vemos que la clase [56,62] contiene los datos en las posiciones 25 a 55 y por lo tanto esa es la clase de la mediana.

i	Li	Ls	Xi	fi	Fi	fi*Xi
1	38	44	41	7	7	287
2	44	50	47	8	15	376
3	50	56	53	15	30	795
4	56	62	59	25	55	1475
5	62	68	65	18	73	1170
6	68	74	71	9	82	639
7	74	80	77	6	88	462
			Suma	88		5204

A partir de estos datos se calcula la mediana así:

Mediana = L_i +IC $\frac{\left(\frac{N}{2}-F_{i-1}\right)}{f_i}$

Donde

$L_i = 56$ Es el límite inferior de la clase de la mediana

$IC = 6$ Intervalo de clases

F_{i-1} = 30 Frecuencia acumulada de la clase anterior

f_i = 25 Frecuencia de la clase de la mediana

De acuerdo a la fórmula

Mediana = L_i +IC $\frac{\left(\frac{N}{2}-F_{i-1}\right)}{f_i} = 56 + 6 \times \frac{\left(\frac{88}{2}-30\right)}{25}$ = 59,36

Media

Se calcula con la fórmula

$$\bar{x} = \frac{\sum X_i f_i}{n} = \frac{5204}{88} = 59,14$$

Interpretación

Moda = 59,53 , Mediana = 59,36 , Media = 59,14

La medida de tendencia central más utilizada es la media que es 59,14 que al ser tan cercana a la Moda y la mediana es en este caso un excelente indicador de las puntuaciones del grupo

b)

Desviación estándar

Primero ampliamos los cálculos en la tabla:

i	Li	Ls	Xi	fi	Fi	fi*Xi	$(Xi-\bar{x})^2$	$f_i(xi-\bar{x})^2$
1	38	44	41	7	7	287	328,93	2302,49
2	44	50	47	8	15	376	147,29	1178,33
3	50	56	53	15	30	795	37,65	564,82
4	56	62	59	25	55	1475	0,02	0,46
5	62	68	65	18	73	1170	34,38	618,88
6	68	74	71	9	82	639	140,75	1266,71
7	74	80	77	6	88	462	319,11	1914,66
			Suma	88		5204	1008,13	7846,36

Luego calculamos la varianza con la fórmula

$$s^2 = \frac{\sum f_i(X_i-\bar{x})^2}{n} = \frac{7846,36}{88} = 89,16$$

Y luego la desviación estándar

$$s = \sqrt{s^2} = \sqrt{89,16} = 9,44$$

El coeficiente de variación mide con mayor precisión la dispersión de los datos referente a la media.

$$CV = \frac{s}{\bar{x}} = \frac{9,44}{59,14} = 0,16$$

Este coeficiente de variación entre bajo y moderado indica que la media es un excelente indicador ya que un alto porcentaje de las puntuaciones está cerca de la media.

C)

Cuartil 3

Formula Qi, para serie de Datos agrupados:

$$Q_i = L_i + \frac{\frac{3i}{4} - f_{a-1}}{f_i} * I_c$$

Donde:

L_i =Límite inferior de la clase que lo contiene.

$P_i = \frac{3n}{4}$ Valor que representa la posición del cuartil. i =1, 2,3

f_i = La frecuencia de la clase que contiene la medida solicitada.

f_{a-1}= Frecuencia acumulada anterior a la que contiene la medida solicitada.

I_c = Intervalo de clase.

Otra manera de verlo es partir de que todas las medidas no son sino caso particulares del percentil, ya que el primer cuartil es el 25% percentil y el tercer cuartil 75% percentil.

En este caso $P_3 = \frac{3\times 88}{4} = 66$

La clase que contiene elementos en posición 66, tomando en cuenta las frecuencias acumuladas de la tabla es la clase [62,68)

Por lo tanto

$$Q_3 = 62 + \frac{66 - 55}{18} \times 6 = 65{,}66$$

Indicando que el 75% de los datos están por debajo de una puntuación de 65,66 y solo el 25% han obtenido una puntuación superior a 65,66.

<u>Decil 5</u>

Para datos agrupados los deciles se calculan mediante la fórmula:

$$D_k = L_k + \frac{K\left(\frac{n}{10}\right) - F_{k-1}}{f_k} * c$$

Donde:

K= 1, 2, 3,…9

L_k= Limite inferior de la clase de decil k

n= Número de datos

F_{k-1}= frecuencia acumulada de la clase que antecede a la clase del decil k

f_k= Frecuencia de la clase del decil k

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

Para determinar la clase del Decil 5 utilizamos la fórmula $P = K\left(\frac{n}{10}\right) = 5\left(\frac{88}{10}\right) = 44$

Por lo tanto la clase del decil 5 obtenida a partir de la observación de las frecuencias acumuladas es la clase [56,62). Luego

$$D_5 = 56 + \frac{5\left(\frac{88}{10}\right) - 30}{25} \times 6$$

$$D_5 = 59{,}36$$

Debajo del decil 5 se hallan el 50% de los datos o puntuaciones inferiores y sobre él se encuentran el 50% de los datos o puntuaciones superiores. Por lo tanto el decil 5 corresponde a la Mediana y por eso coincide con la calculada antes.

Percentil 58

Los percentiles son ciertos números que dividen la sucesión de datos ordenados en cien partes porcentualmente iguales. Estos son los 99 valores que dividen en cien partes iguales el conjunto de datos ordenados. Los percentiles se denotan P1, P2…, P99, leídos primer percentil,…, percentil 99.

Datos Agrupados

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

$$P_k = L_k + \frac{k\left(\frac{n}{100}\right) - F_{k-1}}{f_k} * c$$

Donde:

K= 1, 2, 3…99.

L_k = Limite real inferior de la clase del Percentil k.

n= Numero de datos.

F_{k-1}= Frecuencia acumulada de la clase que antecede a la clase del Percentil k.

f_k = Frecuencia de la clase del percentil k.

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

Para determinar la clase del Percentil 58 utilizamos su posición P con la fórmula

$$P = K\left(\frac{n}{100}\right) = 58\left(\frac{88}{100}\right) = 51{,}04$$

Por lo tanto la clase del Percentil 58 obtenida a partir de la observación de las frecuencias acumuladas es la clase [56,62) que contiene los datos entre las posiciones 31 a 55.

i	Li	Ls	Xi	fi	Fi	fi*Xi	$(Xi-\bar{x})^2$	$f_i(xi-\bar{x})^2$
1	38	44	41	7	7	287	328,93	2302,49
2	44	50	47	8	15	376	147,29	1178,33
3	50	56	53	15	30	795	37,65	564,82
4	56	62	59	25	55	1475	0,02	0,46
5	62	68	65	18	73	1170	34,38	618,88
6	68	74	71	9	82	639	140,75	1266,71
7	74	80	77	6	88	462	319,11	1914,66
			Suma	88		5204	1008,13	7846,36

Luego

$$P_{58} = L_k + \frac{k\left(\frac{n}{100}\right) - F_{k-1}}{f_k} * c = 56 + \frac{58\left(\frac{88}{100}\right) - 30}{25} \times 6 = 61{,}04$$

Indicando que el 58% de las puntuaciones más bajas es inferior a 61,04 y el 32% de las puntuaciones están sobre 61,04.

5.-Si se dice que la probabilidad de que llueva en un día cualquiera es de 0.25, ¿podemos afirmar que la probabilidad que llueva tres dias seguidos es de 0.75?. Argumente su respuesta.

Se supone que el hecho de que un día llueva es independiente del clima de otros días. Es decir se asume que los eventos "llover un día" son independientes. En tal caso como la probabilidad de lluvia cada día es 0,25, entonces la probabilidad de que llueva 3 días seguidos, hoy, mañana y pasado mañana es: 0,25 × 0,25 × 0,25 = 0,0156. Y por lo tanto no es 0,75.

Lic- José Barreto josearturobarreto@gmail.com +584241700032 +584163599615