

Medidas de dispersión o variabilidad

Tema 5

Profesor Tevni Grajales G.

A dos grupos diferentes de estudiantes se les preguntó cuánto deseaban pagar como cuotas de graduación.

En ambos casos el promedio o media de lo que estaban dispuestos a pagar fue el mismo 25 pesos pero en uno de los casos, cuando se informó la decisión hubo una fuerte reacción negativa de parte de un alto porcentaje de los alumnos, mientras que en el otro grupo la decisión fue aceptada con mucho agrado y buena disposición. ¿Qué pudo haber causado la diferencia en las respuestas finales de los grupos, si se siguió el mismo procedimiento para la toma de la decisión? Aunque no pretendemos en este caso dar una respuesta definitiva, a continuación una información muy valiosa que podría contribuir a explicar lo que sucedió.

A continuación se presenta la cantidad que cada persona propuso pagar según el grupo a que pertenecía.

Grupo A: 75, 50, 35, 24, 16, 12, 11, 10, 10, 7; Grupo B: 30, 20, 15, 35, 25, 28, 22, 26, 24.

El primer paso que podemos dar consiste en ordenar los valores propuestos por cada persona en cada grupo. La siguiente tabla presenta los valores ordenados.

GRUPO A	GRUPO B
75	35
50	30
35	28
24	26
16	25
12	24
11	22
10	20
10	15
7	Total = 225
Total = 250	

Medidas de tendencia central

Grupo A	Grupo B
N = 10	N = 9

Suma de valores = 250		Suma de valores = 225
Media = 25		Media = 25
Mediana = 14		Mediana = 25
Moda = 10		Moda = 15*

Medidas de variabilidad

Las medidas de variabilidad que podemos observar en estas distribuciones son: la amplitud, la desviación de la media, la variancia y la desviación típica

La Amplitud

1. La amplitud total, rango o gama que es la diferencia entre el valor más elevado y el valor más bajo en la ordenación.

Para el grupo A: $75 - 7 = 68$

Para el grupo B: $35 - 15 = 20$

En esta observación comparamos el rango de grupo A con el del grupo B y reconocemos que existe una mayor diferencia en los valores escogidos por los miembros del grupo A. Esto no nos dice si esta diferencia está distribuida de una manera equilibrada o si se trata de algunas posiciones extremas.

2. La amplitud intercuartil que es la diferencia entre el valor del tercer cuartil menos el del primer cuartil.

Para el grupo A: $38.75 - 10 = 28.75$

Para el grupo B: $29 - 21 = 8$

La distancia intercuartil elimina el efecto de las personas que se ubicaron en los extremos superior e inferior de la distribución y nos permite ver que en efecto el grupo A está constituido por personas que difieren mucho más que los del grupo B en sus preferencias.

La desviación de la media

La desviación de una puntuación es su distancia desde la media de la distribución y se representa con una x minúscula. De manera que para calcular la desviación de una puntuación respecto a la media de la distribución, se resta el valor de la media del valor observado. $x = (X - \bar{X})$.

Las puntuaciones que se encuentran sobre la media en una ordenación poseen un valor positivo, mientras que aquellas bajo la media lo tienen negativo. Para determinar la desviación desde la media (DM) se suman los valores absolutos de las x 's correspondientes a cada caso y se divide entre el total de casos.

GRUPO A	x	GRUPO B	x
75	$75 - 25 = 50$	35	$35 - 25 = 10$
50	$50 - 25 = 25$	30	$30 - 25 = 5$
35	$35 - 25 = 10$	28	$28 - 25 = 3$

24	$24 - 25 = 1$	26	$26 - 25 = 1$
16	$16 - 25 = -9$	25	$25 - 25 = 0$
12	$12 - 25 = -13$	24	$24 - 25 = -1$
11	$11 - 25 = -14$	22	$22 - 25 = -3$
10	$10 - 25 = -15$	20	$20 - 25 = -5$
10	$10 - 25 = -15$	15	$15 - 25 = -10$
7	$7 - 25 = -18$	Total = 225	

Total = 250

Para el grupo A la desviación de la media resulta de sumar los valores absolutos de $x = 175$ y dividirlo entre 10 que son los casos (N) dando como resultado $DM = 17.5$. La misma medida para el grupo B los valores absolutos de x suman 28 y los casos (N)son 9 dando como resultado $DM = 3.11$.

La variancia o varianza

Se trata de la suma de las desviaciones al cuadrado, consideradas desde la media, de las puntuaciones individuales, dividida por $N - 1$. La varianza se representa por una letra s al cuadrado. Las desviaciones son sometidas al cuadrado para eliminar los signos, lo cual se justifica dado que lo importante es la cantidad de la desviación y no la dirección de la misma.

GRUPO A	$(X - X) = x$	x		GRUPO B	$(X - X) = x$	x
75	$75 - 25 = 50$	2 500		35	$35 - 25 = 10$	100
50	$50 - 25 = 25$	625		30	$30 - 25 = 5$	25
35	$35 - 25 = 10$	100		28	$28 - 25 = 3$	9
24	$24 - 25 = 1$	1		26	$26 - 25 = 1$	1
16	$16 - 25 = -9$	81		25	$25 - 25 = 0$	0
12	$12 - 25 = -13$	169		24	$24 - 25 = -1$	1
11	$11 - 25 = -14$	196		22	$22 - 25 = -3$	9
10	$10 - 25 = -15$	225		20	$20 - 25 = -5$	25
10	$10 - 25 = -15$	225		15	$15 - 25 = -10$	100
7	$7 - 25 = -18$	324		Total = 225		$\Sigma x = 270$
Total = 250		$\Sigma x = 4 446$				

La varianza para el grupo A resulta de dividir la suma de los cuadrados de x (S_x) entre $(N - 1)$, es decir entre $(10 - 1)$ lo cual es 9 y da como resultado 494. La misma operación, con el grupo B, conduce a

dividir 270 entre 8 lo cual da como resultado 33.75.

Cuando se trabaja con datos agrupados en intervalos de clase se identifica el punto medio de cada intervalo (X_i) según aparece en la columna 3 del ejemplo y se multiplica por la frecuencia en cada intervalo (f_i) que son los valores de la columna 2 dando como resultado la columna 4 ($X_i f_i$). Además se eleva al cuadrado el valor de los puntos medios de cada intervalo (X_i) y se multiplica por la frecuencia en cada intervalo (f_i) lo cual resulta en nuestro ejemplo en la columna 5.

Para calcular la varianza de la distribución se procede de la siguiente manera:

Se eleva al cuadrado la suma de la columna 4 ($X_i f_i$) y se divide entre el total de casos (total de la columna 2). Una vez hecha esta operación su resultado se resta de total que resulta en la columna 5 (suma de X_i al cuadrado por f_i) y la diferencia se divide entre el número de casos menos uno. Dando como resultado la varianza.

Ejemplo:

Un consultor estudió una muestra de 40 tiendas de una gran ciudad e investigó, entre otras cosas, el monto anual gastado en publicidad a través de la radio por esas tiendas. La distribución de frecuencias siguiente resume la información recogida. Se calcula de varianza y la desviación estándar.

Miles de pesos	Números de tiendas (f)	X_i	$X_i f_i$	X_i al cuadrado por f_i
10 - 19	4	15	60	900
20 - 29	14	25	350	8 750
30 - 39	8	35	280	9 800
40 - 49	6	45	270	12 150
50 - 59	3	55	165	9 075
60 - 79	4	70	280	19 600
80 - 99	1	90	90	8 100
Totales	40		1 495	68 375

La varianza resulta de elevar 1 495 al cuadrado y dividirlo entre 40 = 55 875,6. Este resultado se resta de la suma de $X_i f_i = 68 375$ resultando 12 499,4 lo cual se dividen entre el total de casos menos uno (40-1) es decir 39. Esto nos da la varianza de 320,5.

La desviación típica o estándar

Es la raíz cuadrada de la variancia (es decir de la media del cuadrado de las desviaciones) y representa lo que podríamos denominar un promedio de las distancias que separan a todos los valores en la distribución, respecto a su media. Esta medida permite identificar cuántas veces el promedio de la distancia se ubica cada valor respecto a su media.En el caso del ejemplo inmediato anterior es la raíz cuadrada de 320,5 = 17,9.

Ejercicio:

En un estudio realizado el 3 de Febrero de 1990 por la Comisión Centroamericana de Salud y Bienestar Social se registraron los pesos de los nacidos en diversos hospitales del área, escogidos al azar. A continuación los datos obtenidos en el estudio:

- República de Panamá, los niños pesaron en libras 3.5, 4.8, 5.3, 2.8, 6, 4.1 y las niñas pesaron 4, 4.2, 2.8, 6.3, 5.5, 7, 5.5, 5.5
- En Costa Rica los niños pesaron 4.6, 5.4, 5.2, 3.6, 5.5, 7, 5.5, y las niñas pesaron 4.7, 5.2, 6.3, 3.5, 4.2
- En Nicaragua los niños pesaron 4.5, 5.3, 4.1, 3.5, 3.3, 4.1, 5.8 y las niñas pesaron 3.8, 4.2, 5.5, 6.2, 4.1, 5.3, 6.3.

1. Elabora un cuadro estadístico que resuma toda la información.
2. Haz un gráfico que presente en forma clara y precisa los aspectos más relevantes de los resultados.
3. Determina cuál es la proporción que corresponde a cada nacionalidad en la muestra.
4. Señala en un gráfico cuál es el porcentaje de varones que pesan más de 4.5 libras.
5. Determina la moda, mediana, media, rango, amplitud intercuartil, varianza y desviación típica de la muestra.