ESTADISTICA

Medidas de dispersión o variabilidad

Medidas de dispersión o variabilidad

La dispersión o variación de los datos nos indica cuan esparcidos se encuentran estos. En cambio las medidas de tendencia central nos permiten conocer cuáles son los valores alrededor de los cuales se concentran los datos de una muestra o población.

Analizamos las mismas a través del siguiente ejemplo:

Las edades de los hermanos de Andrés son 4, 6, 8, 10, en cambio las edades de los primos son 1,3,5,19 años. Calculamos la media de las edades de ambos grupos

- Media de los hermanos: $\mu = (4+6+8+10)/4=28/4= 7$ años
- Media de los primos: $\mu = (1+3+5+19)/4=28/4= 7$ años

Si bien es cierto que la media de ambos grupos es 7 años, este valor es mas representativo en las edades de los hermanos de Andrés que la de sus primos, pues algunas edades de sus primos como 1, 3 y 19 se encuentran muy alejados del valor medio 7 años. En este caso se habla de dispersión o desviación de los valores.

En conclusión: Podemos decir que para caracterizar una distribución de datos no basta con conocer las medidas de centralización. Es preciso, además, indicar el grado de dispersión de los datos.

Las medidas de dispersión ayudan a las medidas de tendencia central a dar una mejor descripción de la globalidad de los datos, midiendo de que forma se alejan o se acercan al valor central.

A continuación veremos algunos de las más utilizadas, considerando siempre primero datos no agrupados

Rango R :

El rango también conocido como amplitud, es la diferencia entre el valor mas alto y el mas bajo registrados. Se representa por la letra R. cuanto mayor es la amplitud, mayor es la variación o dispersión de los datos

Ejemplos:

En las edades de los hermanos de Andrés es: R= 10-4 = 6

En las edades de los primos de Andrés es: R= 19-1= 18

Esta medida, el rango, indica que en las edades de los hermanos de Andrés existe menor dispersión que en las edades de los primos

Desviación Individual:

La desviación es el valor absoluto de las diferencias entre los valores individuales observados y la media.

Simbolizamos: D = I Xi - μ I   Donde: Xi es el valor de la variable

μ es el valor de la media

Ejemplo:

Tomamos la siguiente serie de números: 4,5,6,6,7,7,8 y calculamos las desviaciones:

Primero hallamos la media: μ= (4+5+6+6+7+7+8)/ 7= 6,1

Las desviaciones serán:

I 4-6,1I = 2,1

I 5-6,1I = 1,1

I6 -6,1I = 0,1

I 6-6,1I = 0,1

I 7-6,1I = 0,9

I 7-6,1I = 0,9

I 8-6,1I = 1,9

Las medidas de dispersión más utilizadas son: la desviación media, la deviación típica o estándar y la varianza, y la relativa que es el coeficiente de variación.

VARIANZA (S^2)

La varianza de un conjunto de valores es la media del cuadrado delas desviaciones.

La varianza siempre será mayor que cero, cuanto más se aproxima a cero, más concentrados están los valores alrededor de la media, mientras que, cuanto mayor sea la varianza, mas dispersos están. Se representa con S^2 si corresponde a los datos obtenidos por muestreo y con σ^2 (sigma al cuadrado) si corresponde a la población.

Su fórmula es la siguiente:

$$S^2 = \frac{\sum_{i=1}^{i=n}(Xi - \mu)^2}{n}$$

Donde :

n: es la cantidad de datos

Xi: es cada uno de los valores

μ : es la media de los datos

$S^2 = \frac{\sum_{i=1}^{i=m}(Xi - \mu)^2 * fi}{n}$ es otra forma de escribir la formula anterior, y se utiliza para datos agrupados.

Donde :

n: es la cantidad de datos o suma de frecuencia absolutas

Xi: es cada uno de los valores o la marca de clase de los intervalos

μ: es la media de los datos

fi: la frecuencia absoluta

m: la cantidad de intervalos

Ejemplo:

a)si tomamos los valores 4, 5,6,6,7,7,8, , como hemos visto la media de estos valores es μ=6,1

Aplicamos la fórmula de la varianza: $S^2 = \frac{\sum_{i=1}^{\mathrm{i=n}} (Xi - \mu)^2}{n}$

$$S^2 = \frac{(4-6,1)^2 +(5-6,1)^2+(6 -6,1)^2+(6-6,1)^2+(7-6,1)^2+(7-6,1)^2+(8-6,1)^2}{7} =$$

$$S^2 = \frac{10,87}{7} =1,6$$

b) Cuando los valores están ordenados en tablas de frecuencias, podemos calcular la varianza aplicando la segunda fórmula:

Número de personas de una manzana por domicilio				
Número de personas Xi	Fi	Xi * fi	$(Xi - \mu)^2$	$(Xi - \mu)^2 * fi$
1	2			
2	5			
3	8			
4	10			
5	15			
6	12			
7	8			
8	6			
9	6			
10	4			
11	2			
12	2			
Totales	80			

c) Tomamos del ejemplo anterior de datos agrupados y la completamos de la siguiente manera:

Cantidad Vendida (intervalos de clases)	Marca de clases Xi	Fi	Xi * fi	$(Xi - \mu)^2$	$(Xi - \mu)^2 * fi$
8-10	9	2			
11-13	12	10			
14-16	15	13			
17-19	18	3			
20-22	21	2			
Totales		30			

DESVIACIÓN TIPICA O ESTANDAR:

La desviación típica o estándar es la raíz cuadrada de la varianza. Se representa con la letra S en las muestras y σ en la población.

Es la medida de dispersión más utilizada e importante porque se emplea para comprobar la confianza que merecen algunas medidas estadísticas

Su fórmula es: $S = \sqrt{varianza}$, que se utiliza también para datos agrupados

Ejemplos:

La desviación típica del ejemplo número de personas de una manzana por domicilio es:

$S = \sqrt{6,85}$ = 2,6

Analizamos: las dos medidas que resumen cuantitativamente el resultado del número de personas de una manzana por domicilio son la media y la desviación típica o estándar:

$\mu = 5{,}8 \quad S = 2{,}6$

la distribución de los resultados se considera normal, si un 68% del total de los puntajes se halla comprendido entre los valores tomados como parámetro:

$\mu - S \qquad y \qquad \mu + S$

en nuestro ejemplo , el 68% de las personas deberían estar entre:

$5{,}8 - 2{,}6 = 3{,}2 \qquad\qquad y \qquad\qquad 5{,}8 + 2{,}6 = 8{,}4$

Verificamos:

El 68% de 80 = 54 personas aproximadamente, deberían pertenecer a domicilios de entre 3 y 8 personas. Por consiguiente, la distribución puede considerarse normal, en sentido de habitual, usual, no supone ni malo ni bueno.

COEFICIENTEDE VARIACION:

Indica la relación entre la desviación estándar o típica y la media. Puede expresarse porcentualmente multiplicando por 100 el cociente entre la desviación típica y la media.

Es una medida objetiva de la dispersión de un conjunto de datos. Se utiliza normalmente para comparar 2 poblaciones o muestras. Tienen que ser las mismas unidades de medidas, si están en metros tienen que ser en metros, no puede ser uno en metros y otro en kilogramos.

La muestra que tenga un coeficiente de variación menor es más homogénea o menos dispersa.

Su formula :

$CV = \frac{S}{\mu} * 100\% = \cdots\%$ Donde: S: desviación típica

μ: media

Ejemplos:

a) Tomamos los resultados obtenidos en el ejemplo " número de personas de una manzana por domicilio"

Como: μ= 5,8 y S= 2,6

CV= (2,6/ 5,8) * 100= 44,8%

b) La media aritmética y la desviación típica de las estaturas de dos equipos de futbol figuran en la tabla siguiente:

Equipos	Desviacion Tipica S	Media μ	Coef. Variacion CV
A	3,45	174,3	CV=3,45/174,3*100= 1,9%
B	6,72	174,3	CV=6,72/174,3 *100= 3,9

Determinemos en cuál de los dos equipos la distribución es más homogénea.

El equipo A tiene CV= 1,9%, menor que el equipo B que tiene CV= 3,9%, por lo tanto A tiene las estaturas más homogéneas.

Ejercicios de Aplicación

1) Las temperaturas máximas registradas en la ciudad en una semana fueron

34ºC 32ºC 33º 34ºC 35ºC 35ºC 36ºC

Calcula la media de las temperaturas máximas registradas y la varianza.

2) La siguiente tabla es una distribución de las cantidades de pantalones vaqueros pedidas por un cliente a una fabrica que lo produce

Talla (intervalos de clases)	Marca de clases Xi	Nº de pedidos Fi	Xi * fi	I Xi - µ I²	I Xi - µ I² * fi
32-36		6			
37-41		10			
42-46		16			
47-51		12			
52-56		6			
Totales		50			

Calcular: La media y la desviación típica

3) La siguiente tabla ilustra la existencia de respiradores a la fecha (5/08/2004) en los hospitales de salud pública:

Hospital	Nº de respiradores X_i
Pediátrico Acosta Ñu	10
Barrio Obrero	4
Materno San Pablo	9
Medicina Tropical	7
Boettner	4
Emergencias Médicas	4
Trinidad	4
Del Quemado	3
Loma Pytá	0
Mariano Roque Alonso	0
Totales	

Calcular: a) la media, b) la varianza, c) la desviación típica o estándar,

4) Una prueba de 5 items fue aplicada a un curso de 42 alumnos. El número de respuestas correctas se registra en la siguiente tabla:

Nº de respuestas Xi	Fi	Xi*fi	$Xi^2 * fi$
0	1		
1	2		
2	8		
3	14		
4	12		
5	5		
total	42		

Calcular: a) la media, b) la varianza, c) la desviación típica o estándar, d) el coeficiente de variación.

5) Una empresa de nuestro medio ha registrado la siguiente variación del dólar compra el ultimo día de cada mes, en el primer semestre del año 2004:

Hallar:

- La media
- La varianza
- La desviación típica
- El coeficiente de variación

Enero	febrero	Marzo	Abril	mayo	Junio
6200	6030	5860	5725	5970	5890

6) La tasa de hemoglobina en g/100ml, encontrada en 560 personas normales es como sigue:

Hallar:

- La media
- La varianza
- La desviación típica
- El coeficiente de variación

Hemoglobina (Intervalos de clases)	Marca de clases Xi	Fi			
12,5-13,5		5			
13,5-14,5		34			
14,5-15,5		134			
15,5-16,5		213			
16,5-17,5		129			
17,5-18-5		39			
18,5-19,5		6			
Totales		560			