Medidas de dispersión

Estadística Descriptiva

Medidas de dispersión

Las medidas de dispersión cuantifican la variabilidad (el grado de separación) que presenta un conjunto de datos. Mientras mayor sea el valor de la medida usada, mayor será la distancia de los datos respecto al centro. Entre las medidas de dispersión usadas comúnmente, se encuentran:

- Rango.
- Varianza.
- Desviación estándar o típica.
- Coeficiente de variación.

Rango o amplitud

La dispersión puede medirse en términos de la diferencia entre los dos valores extremos del conjunto de datos. De esta forma, el rango se define como la diferencia entre el máximo y el mínimo valor de la distribución.

$$\text{Rango} = x_{max} - x_{min}$$

Observación: Esta medida de dispersión tiene como inconveniente ser poco representativa, cuando existen valores extremos atípicos . Por esta razón puede ser más conveniente usar el rango intercuartil, que se define como la diferencia entre el tercer y el primer cuartil.

$$\text{Rango Intercuartil} = Q_3 - Q_1 = P_{75} - P_{25}$$

(cuartil 3 menos cuartil 1, o bien percentil 75 menos percentil 25)

El rango intercuartil corresponde a la amplitud o longitud del intervalo donde se concentra el 50% central de los datos. Indica que tan concentrado o dispersos están los datos centrales.

Medidas de dispersión

Varianza

La varianza se define como el promedio de las desviaciones (distancias) cuadráticas de las observaciones respecto del promedio. Mide el grado de dispersión de los valores de la variable respecto a la media aritmética.

Observación: Esta medida de dispersión tiene como inconveniente que las unidades de la variable están al cuadrado.

Las siguientes fórmulas se utilizan para calcular la varianza, pero nosotros utilizaremos la planilla de cálculo Excel para obtener este valor.

Varianza poblacional: $$\sigma^2 = \frac{\sum \left(x_i - \bar{x}\right)^2}{n} = \overline{X^2} - \left(\overline{X}\right)^2$$

Varianza muestral: $$S^2 = \frac{\sum \left(x_i - \bar{x}\right)^2}{n-1}$$

Desviación estándar

La desviación estándar se define como la raíz cuadrada de la varianza. Como la varianza tiene como unidad de medida el cuadrado de la unidad de la variable, se hace difícil de interpretar. Al tomar la raíz cuadrada, se facilita la interpretación.

Las siguientes fórmulas se utilizan para calcular la desviación estándar, pero nosotros utilizaremos la planilla de cálculo Excel para obtener este valor.

Desviación estándar poblacional: $\sigma = \sqrt{\sigma^2} = \sqrt{\dfrac{\sum \left(x_i - \overline{x}\right)^2}{n}} = \sqrt{\overline{X^2} - \left(\overline{X}\right)^2}$

Desviación estándar muestral: $S = \sqrt{S^2} = \sqrt{\dfrac{\sum \left(x_i - \overline{x}\right)^2}{n-1}}$

Observación: Las medidas de dispersión vistas hasta ahora no deben ser usadas para comparar entre dos variables con unidades distintas (por ejemplo, peso y estatura de los estudiantes de Duoc) o cuando queremos remover el efecto de la media (por ejemplo, el peso de los elefantes varía más en términos absolutos que el peso de los grillos, ¿pero en términos relativos?).

Medidas de dispersión

El coeficiente de variación se define como la razón entre la desviación estándar y la media aritmética. Normalmente se expresa como porcentaje, al multiplicarlo por 100.

Observación: Esta medida de dispersión tiene como ventaja que no depende de las unidades de medidas de la variable, lo cual permite comparar la variabilidad de datos de variables con distintas unidades.

La planilla de cálculo Excel no tiene una fórmula directa para calcular el coeficiente de variación, por lo que se debe aplicar la siguiente fórmula.

Coeficiente de variación: $$CV = \frac{\sigma}{\overline{\overline{X}}} = \frac{S}{\overline{\overline{X}}}$$

→ Desviación estándar

→ Media aritmética

Medidas de dispersión

Coeficiente de Variación

0% ←——————————————→ 100%

Datos menos dispersos
(más homogéneos)

Datos más dispersos
(más heterogéneos)

Homogéneo: Uniforme, semejante, similar, idéntico.

Heterogéneo: Diverso, variado, mezclado, distinto.

Observación: En la mayoría de las distribuciones de datos el coeficiente de variación toma valores desde 0% al 100%.

Ejemplo 1: Medidas de dispersión con Excel

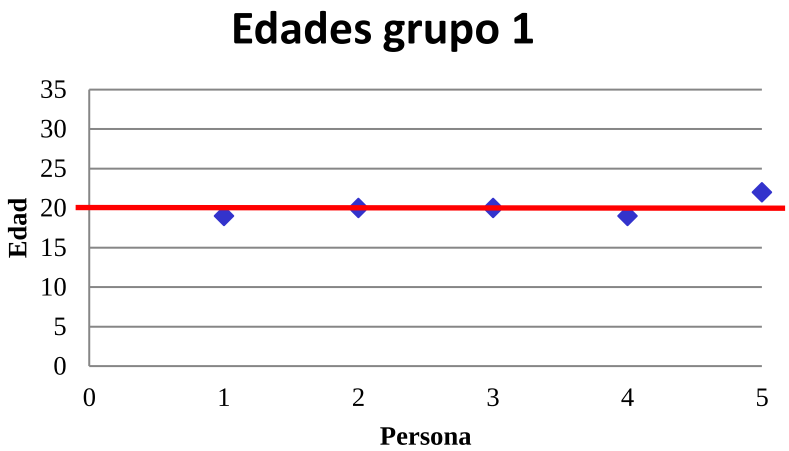
Los siguientes datos representan las edades en años de dos grupos diferentes de personas, en las cuales se calculó el promedio. Determine el grupo de personas que presenta un comportamiento más homogéneo en sus edades.

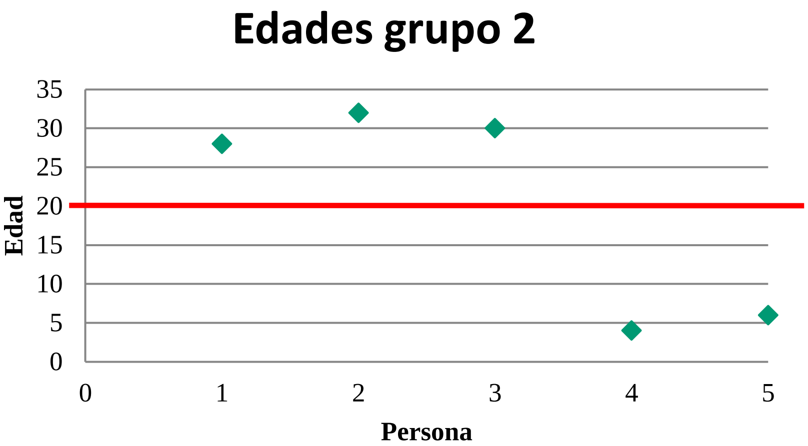
Grupo 1	Grupo 2
19 – 20 – 20 – 19 – 22	28 – 32 – 30 – 4 – 6

Media: 20	Media: 20

Ejemplo 1: Medidas de dispersión con Excel

Respuesta





Como se puede observar en la gráfica de cada grupo de personas, las medidas de tendencia central (en este caso la media) no reflejan la variabilidad o dispersión del grupo de datos.

Es por esto que debemos utilizar otro indicador para medir la variabilidad de los datos. Este indicador puede ser el coeficiente de variación.

Ejemplo 1: Medidas de dispersión con Excel

Respuesta

Con la función "Análisis de datos" y usando "Estadística descriptiva", podemos realizar un resumen de estadísticas, para calcular el coeficiente de variación y determinar cual de las muestras presenta un comportamiento más homogéneo en sus edades.

Edades del Primer Grupo	
Media	20
Error típico	0,547722558
Mediana	20
Moda	19
Desviación estándar	1,224744871
Varianza de la muestra	1,5
Curtosis	2
Coeficiente de asimetría	1,360827635
Rango	3
Mínimo	19
Máximo	22
Suma	100
Cuenta	5

Coeficiente de Variación = 1,224744871/20

= 0,061

= 6,1%

Edades del Segundo Grupo	
Media	20
Error típico	6,164414003
Mediana	28
Moda	#N/A
Desviación estándar	13,78404875
Varianza de la muestra	190
Curtosis	-3,194459834
Coeficiente de asimetría	-0,572744408
Rango	28
Mínimo	4
Máximo	32
Suma	100
Cuenta	5

Coeficiente de Variación = 13,78404875/20

= 0,689

= 68,9%

Ejemplo 1: Medidas de dispersión con Excel

Respuesta

El coeficiente de variación de la edad del primer y segundo grupo, son respectivamente 6,1% y 68,9%, por lo que el primer grupo presenta un comportamiento más homogéneo con respecto a la edad, ya que su coeficiente de variación es menor.

Ejemplo 2: Medidas de dispersión con Excel

Se tienen los registros del sueldo en miles de pesos y los años de antigüedad de 30 operarios de una fábrica. Determine con respecto a qué variable los operarios presentan menor variabilidad. Justifique su respuesta.

Sueldos en m$

396	351	427
296	360	338
385	400	317
348	367	346
405	361	392
367	411	492
496	359	292
372	455	400
483	433	362
309	435	378

Años de antigüedad

12	5	6
11	5	3
9	7	4
6	5	6
8	7	8
7	8	1
5	10	7
9	12	10
8	8	3
5	8	5

Ejemplo 2: Medidas de dispersión con Excel

Respuesta

Con la función "Análisis de datos" y usando "Estadística descriptiva", podemos realizar un resumen de estadísticas, pero primero debemos presentar los datos en una columna para cada variable y así determinar cual de los conjuntos de datos presentan menor variabilidad.

Sueldos en m$	
Media	384,4333333
Error típico	9,768902494
Mediana	375
Moda	367
Desviación estándar	53,50648258
Varianza de la muestra	2862,943678
Curtosis	-0,114155291
Coeficiente de asimetría	0,427487903
Rango	204
Mínimo	292
Máximo	496
Suma	11533
Cuenta	30

Coeficiente de Variación = 53,50648/384,43

= 0,139

= 13,9%

Años de antigüedad	
Media	7,266666667
Error típico	0,454690329
Mediana	7
Moda	8
Desviación estándar	2,490441497
Varianza de la muestra	6,202298851
Curtosis	-0,617307685
Coeficiente de asimetría	0,247648118
Rango	9
Mínimo	3
Máximo	12
Suma	218
Cuenta	30

Coeficiente de Variación = 2,49044/7,26

= 0,343

= 34,3%

Ejemplo 2: Medidas de dispersión con Excel

Respuesta

El coeficiente de variación del sueldo de los operarios es 13,9% y el de la antigüedad es 34,3%, por lo que el sueldo presenta una menor variabilidad, ya que su coeficiente de variación es menor.

Resumen

- Indicadores Estadísticos (Estadígrafos)
  - Posición
    - Tendencia Central
      - •Media aritmética
      - •Moda
      - •Mediana
    - Tendencia No Central
      - •Percentiles
  - Dispersión
    - •Rango
    - •Varianza
    - •Desviación estándar
    - •Coeficiente de variación