¿Qué significa las Medidas de Tendencia Central para distribuciones de frecuencia?

¿Qué es coeficiente de variación y para qué sirve?

¿Qué es un intervalo de confianza?

¿Qué es una estimación por Intervalos de Confianza?

Comparación de la variación en diferentes poblaciones.

Anteriormente afirmamos que, ya que las unidades de la desviación estándar son las mismas que las unidades de los datos originales, es más fácil comprender la desviación estándar que la varianza. Sin embargo, esta misma propiedad dificulta comparar la variación de valores que se tomaron de distintas poblaciones. El coeficiente de variación resuelve tal desventaja.

Definición

Coeficiente de variación o **CV** de un conjunto de datos muestrales o poblacionales, expresado como porcentaje, describe la desviación estándar relativa a la media, y está dada de la siguiente forma:

Muestra

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Población

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

EJEMPLO

Estatura y peso de hombres Si utilizamos los datos muestrales estatura y peso de 40 hombres, incluidos en el conjunto de datos, obtendremos los estadísticos que aparecen en la siguiente tabla. Calcule el coeficiente de variación de las estaturas, después el coeficiente de variación de los pesos; finalmente, compare los dos resultados.

	Media ($\bar{x}$)	Desviación estándar (s)
Estatura	68.34 in	3.02 in
Peso	172.55 lb	26.33 lb

SOLUCIÓN

Debido a que tenemos estadísticos muestrales, los dos coeficientes de variación se obtienen de la siguiente manera:

$$\text{Estaturas:} \qquad CV = \frac{s}{\bar{x}} \cdot 100\% = \frac{3.02 \text{ in}}{68.34 \text{ in}} \cdot 100\% = 4.42\%$$

$$\text{Pesos:} \qquad CV = \frac{s}{\bar{x}} \cdot 100\% = \frac{26.33 \text{ lb}}{172.55 \text{ lb}} \cdot 100\% = 15.26\%$$

Aun cuando la diferencia en unidades imposibilita la comparación de la desviación estándar de $\mathbf{3.02}$ pulgadas, con la desviación estándar de 26.33 libras, es posible comparar los coeficientes de variación, que carecen de unidades. Se observa que las estaturas $(\boldsymbol{con\ CV = 4.42}\%)$ tienen una variación considerablemente menor que los pesos $(\boldsymbol{con\ CV = 15.26}\%)$. Lo anterior tiene sentido, ya que, por lo general, vemos que los pesos de los hombres varían mucho más que sus estaturas. Por ejemplo, es muy raro encontrar un adulto que mida el doble que otro, pero es mucho más común ver a uno que pese el doble que otro.

Media y desviación estándar de datos agrupados.

En la mayoría de los casos las medidas de ubicación, como la media, y las medidas de dispersión, como la desviación estándar, se determinan utilizando valores individuales. Los paquetes de software de estadística facilitan el cálculo de estos valores, incluso en el caso de conjuntos grandes de datos. Sin embargo, algunas veces sólo se cuenta con la distribución de frecuencias y se desea calcular la media o la desviación estándar. En la siguiente explicación se le enseñará cómo calcular la media y la desviación estándar a partir de datos organizados en una distribución de frecuencias. Hay que insistir en que una media o una desviación estándar de datos agrupados es una estimación de los valores reales correspondientes.

Media aritmética.

Para aproximar la media aritmética de datos organizados en una distribución de frecuencia, comience suponiendo que las observaciones en cada clase se representan a través del punto medio de la clase. La media de una muestra de datos organizados en una distribución de frecuencias se calcula de la siguiente manera:

media aritmética para datos agrupados

$$\bar{x} = \frac{\sum fM}{n}$$

donde:

$\bar{x}$ designa la media muestral.

M es el punto medio de cada clase.

f es la frecuencia en cada clase.

fM es la frecuencia en cada clase multiplicada por el punto medio de la clase. $\sum fM$es la suma de estos productos.

n es el número total de frecuencias.

Ejemplo

Los cálculos de la media aritmética de datos agrupados en una distribución de frecuencias que aparecen en seguida se basan en los datos de las ganancias de Applewood Auto Group. Determine la ganancia media aritmética por vehículo.

Ganancia	Frecuencia
$[200 - 600)$	8
$[600 - 1000)$	11
$[1000 - 1400)$	23
$[1400 - 1800)$	38
$[1800 - 2200)$	45
$[2200 - 2600)$	32
$[2600 - 3000)$	19
$[3000 - 3400)$	4
Total	180

Solución

Ganancia	Frecuencia	Punto medio (M)	fM
$[200 - 600)$	8	$\frac{200+600}{2} = 400$	$8(400) = 3200$
$[600 - 1000)$	11	$\frac{600+1000}{2} = 800$	$11(800) = 8800$
$[1000 - 1400)$	23	1200	27600
$[1400 - 1800)$	38	1600	60800
$[1800 - 2200)$	45	2000	90000
$[2200 - 2600)$	32	2400	76800
$[2600 - 3000)$	19	2800	53200
$[3000 - 3400)$	4	3200	12800
Total	180		$\sum fM = 333200$

Por tanto, la Media aritmética para datos agrupados es

$\overline{x} = \frac{\sum fM}{n} \rightarrow \overline{x} = \frac{333200}{180} = \mathbf{1851.11}$, Así, se concluye que la ganancia media por vehículo es de aproximadamente $1 851.

Desviación estándar Para calcular la desviación estándar de datos agrupados en una distribución de frecuencias, necesita ajustar ligeramente la fórmula. Pondere cada una de las diferencias cuadradas por el número de frecuencias en cada clase. La fórmula es:

Desviación estándar para datos agrupados

$$s = \sqrt{\frac{\sum f(M-\overline{x})^2}{n-1}}$$

Ganancia	Frecuencia	Punto medio (M)	fM	$(M-\overline{x})$	$(M-\overline{x})^2$	$f(M-\overline{x})^2$
$[200-600)$	8	$\frac{200+600}{2} = 400$	$8(400) = 3200$	$400 - 1\,851 = -1451$	$(-1451)^2 = 2105401$	16 843 208
$[600-1000)$	11	$\frac{600+1000}{2} = 800$	$11(800) = 8800$	-1051	$(-1051)^2 = 1104601$	12 150 611
$[1000-1400)$	23	1200	27600	-650	423801	9 747 423
$[1400-1800)$	38	1600	60800	-251	63001	2 394 038
$[1800-2200)$	45	2000	90000	149	22201	999 045
$[2200-2600)$	32	2400	76800	549	301401	9 644 832
$[2600-3000)$	19	2800	53200	949	900601	17 111 419
$[3000-3400)$	4	3200	12800	1349	1819801	7 279 204
Total	180		$\sum fM = 333200$			$\sum f(M-\overline{x})^2 =$ 76 169 780

$$s = \sqrt{\frac{\sum f(M-\overline{x})^2}{n-1}}$$

$$s = \sqrt{\frac{76\,169\,780}{180-1}} \Longrightarrow s = \sqrt{425529.4973} = 652.33$$

Por tanto, la desviación estándar es $s = 652.33$

Recuerde que:

La desviación estándar es la medida de dispersión más común, que indica qué tan dispersos están los datos con respecto a la media. Mientras mayor sea la desviación estándar, mayor será la dispersión de los datos.

Realiza la siguiente actividad.

El IRS (Internal Revenue Service) estaba interesado en el número de formas fiscales individuales que preparan las pequeñas empresas de contabilidad. El IRS tomó una muestra aleatoria de 50 empresas de contabilidad pública con 10 o más empleados que operan en la zona de Dallas-Fort Worth.

La siguiente tabla de frecuencias muestra los resultados del estudio. **Calcule la media y la desviación estándar.**

Número de clientes	Frecuencia
20 a 30	1
30 a 40	15
40 a 50	22
50 a 60	8
60 a 70	4

a) **¿Qué nombre recibe la tabla?**
b) **La media.**
c) **Desviación estándar.**

Los gastos en publicidad constituyen un elemento significativo del costo de los artículos vendidos. En seguida aparece una distribución de frecuencias que muestra los gastos en publicidad de 60 compañías manufactureras ubicadas en el suroeste de Estados Unidos. **Calcule la media y la desviación estándar de los gastos en publicidad.**

Gastos en publicidad ($ millones)	Número de compañías
25 a 35	5
35 a 45	10
45 a 55	21
55 a 65	16
65 a 75	8
Total	60