1.2 Medidas de variación: Rango, desviación estándar y coeficiente de variación

Medidas de Variación

Amplitud	Diferencia entre los valores mayor y menor de un conjunto de datos obtenidos en una medición.
Coeficiente de variación	Equivale a la desviación típica expresada en porcentaje respecto de la media aritmética. Es la desviación típica partido por la media aritmética.
Desviación estándar	Medida de la dispersión de una distribución de frecuencias respecto de su media. Equivale a la raíz cuadrada de la varianza. Se expresa como □ si corresponde a la población total o como **s** si corresponde a una muestra de la población
Rango	Medida equivalente a la amplitud
Valor Z	Medida del número de desviaciones estándar que un valor se aleja de la media Z= (x_i - X) / s o Z= (x_i - x) / n
Varianza	Medida de la variación de una serie de observaciones respecto de la media. Equivale a la dispersión respecto de la media en una serie de datos continuos. Su cálculo corresponde a: □ (x_i- x)^2/n si corresponde a la población total o sigma (x_i- X)^2/(n-1) si corresponde a una muestra de esa población, siendo **X** la media, **n** el tamaño de la población o de la muestra y $\mathbf{x_i}$ cada uno de los valores.

1.2.1 Varianza.

Existe otro mecanismo para solucionar el efecto de cancelación para entre diferencias positivas y negativas. Si elevamos al cuadrado cada diferencia antes de sumar, desaparece la cancelación:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

Esta fórmula tiene una desventaja, y es que sus unidades no son las mismas que las de las observaciones, ya que son unidades cuadradas.

Esta dificultad se soluciona, tomando la raíz cuadrada de la ecuación anterior:

Varianza

Es otra de las variaciones absolutas y la misma se define como el cuadrado de la desviación típica; viene expresada con las mismas letras de la desviación típica pero elevada al cuadrado, así S2 y s2. Las fórmulas para calcular la varianza son las mismas utilizadas por la desviación típica, exceptuando las respectivas raíces, las cuales desaparecen al estar elevados el primer miembro al cuadrado.

Varianza

Denotando por $x_1,\ldots,x_k$ los datos o las marcas de clase, llamaremos varianza a

$$s^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - a)^2 \cdot n_i = \frac{1}{n}\sum_{i=1}^{k} x_i^2 \cdot n_i - a^2$$

siendo *a* la media de la distribución.

Al valor

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - a)^2 \cdot n_i = \frac{n \cdot s^2}{n-1}$$

se le denomina *cuasivarzanza*.

VARIANZA.

Una medida de dispersión mucho más común, que se calcula al promediar los cuadrados de las desviaciones individuales a partir de la media, es la media de desviaciones cuadráticas o la varianza.

La varianza es una medida de dispersión promedia de un conjunto de datos. Para una población se construye al tomar la diferencia entre cada valor observado y la media poblacional, elevando al cuadrado cada una de estas desviaciones y luego hallando la media aritmética de los valores cuadrados. Para una muestra, una expresión casi análoga se construye con la ayuda de su media.

Para una población

$$\sigma^2 = \Sigma(X - \mu)^2 / N$$

Para una muestra

$$s^2 = \Sigma(X - \overline{X})^2 \; / \; (n - 1)$$

EJEMPLO Calcule la varianza para una población de N = 5 valores: 2, 2, 4, 7 y 15.

SOLUCION La tabla que muestra la forma en que la varianza se calcula para datos poblacionales, procedimiento por demás tedioso cuando el número de observaciones es grande. Los programas modernos para computadora efectúan con suma rapidez este tipo de operación.

Tabla

VALORES OBSERVADOS X	MEDIA DE LA POBLACION	DESVIACIONES	CUADRADO DE LAS DESVIACIONES
2	6	-4	16
2	6	-4	16
4	6	-2	4
7	6	1	1
15	6	9	81
$\Sigma = 30$			$\Sigma = 118$
	$\sigma = 118/5 = 23.6$		

PROBLEMAS PRACTICOS

Por desgracia hay dos problemas prácticos relacionados con el uso de concepto de varianza. Primero la varianza tiende a ser un número grande en comparación con las observaciones cuya dispersión haya de describirse. Cuando las observaciones originales son iguales a unos pocos miles de millones, su varianza puede ser igual a muchos cientos de miles de millones. En segundo término, y más grave es que la varianza, siendo un número elevado al cuadrado no se expresa en las mismas unidades que los valores observados en sí.

Pero también hay buenas noticias: ambas dificultades conceptuales se pueden vencer de un solo golpe al trabajar con la raíz cuadrada de la varianza, concepto el cual vemos en seguida.

1. 2. 2 Desviación estándar.

VARIANZAY DESVIACIÓN ESTÁNDAR

La varianza se asemeja a la desviación media absoluta en que se basa en la diferencia entre cada valor del conjunto de datos y la media del grupo. Pero se distingue de ella en un muy importante aspecto: cada diferencia se eleva al cuadrado antes de sumarse. En el caso de una población, la varianza se representa con V(X) o, más habitualmente, con la letra griega minúscula o^2 ("sigma cuadrada"). La fórmula es

$$V(X) = \sigma^2 = \frac{\Sigma(X-\mu)^2}{N}$$

A diferencia de lo que ocurre con las demás estadísticas muestrales ya expuestas, la varianza de una muestra no equivale exactamente, en términos de cálculo, a la varianza de una población. El denominador de la fórmula de la varianza muestral es un tanto distinto. En esencia, en esta fórmula se incluye un factor de corrección, a fin de que la varianza muestral sea un estimador insesgado de la varianza de la población. La varianza muestral es representada por s^2; su fórmula es

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1}$$

En general, es difícil interpretar el significado del valor de una varianza, porque las unidades en las que se le expresa son valores elevados al cuadrado. Debido en parte a esta razón, es más frecuente el uso de la raíz cuadrada de la varianza, representada por la letra griega a (o por s en el caso de una muestra) y llamada desviación estándar. Las fórmulas son:

Desviación estándar de la población: $$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$$

Desviación estándar de la muestra: $$s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}}$$

La desviación estándar es particularmente útil en conjunción con la así llamada distribución normal.

$$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}} = \sqrt{\frac{86}{8}} = \sqrt{10.75} = 3.3$$

Hoja de trabajo para el cálculo de la desviación estándar de la población de los datos de ventas (u= 10.5)

X	$X - \mu$	$(X - \mu)^2$
5	−5.5	30.25
8	−2.5	6.25
8	−2.5	6.25
11	0.5	0.25
11	0.5	0.25
11	0.5	0.25
14	3.5	12.25
16	5.5	30.25
		Total 86.00

CÁLCULOS SIMPLIFICADOS DE LA VARIANZA Y LA DESVIACIÓN ESTÁNDAR

Las fórmulas se llaman fórmulas de desviaciones, porque en cada caso deben determinarse las desviaciones específicas de los valores individuales respecto de la media. Sin embargo, se han derivado ya otras fórmulas, matemáticamente equivalentes pero que no requieren de la determinación de cada desviación. Dado que por lo general estas fórmulas son más fáciles de utilizar en la realización de cálculos, se llaman fórmulas de cálculo.

Las fórmulas de cálculo son:

Varianza de la población: $$\sigma^2 = \frac{\sum X^2 - N\mu^2}{N}$$

Desviación estándar de la población: $$\sigma = \sqrt{\frac{\sum X^2 - N\mu^2}{N}}$$

Varianza de la muestra: $$s^2 = \frac{\sum X^2 - n\bar{X}^2}{n-1}$$

Desviación estándar de la muestra: $$s = \sqrt{\frac{\sum X^2 - n\bar{X}^2}{n-1}}$$

Tabla Hoja de trabajo para el cálculo de la desviación estándar de la población de los datos de ventas

X	X^2
5	25
8	64
8	64
11	121
11	121
11	121
14	196
16	256
	Total 968

Desviación típica o estándar

Es la medida de dispersión más utilizada en las investigaciones por ser la más estable de todas, ya que para su cálculo se utilizan todos los desvíos con respecto a la media aritmética de las observaciones, y además, se toman en cuenta los signos de esos desvíos. Se le designa con la letra castellana S cuando se trabaja con una muestra y con la letra griega minúscula s (Sigma) cuando se trabaja con una población. Es importante destacar que cuando se hace referencia a la población él número de datos se expresa con N y cuando se refiere a la muestra él número de datos se expresa con n. La desviación típica se define como:

Interpretación de la desviación estándar
La desviación típica como medida absoluta de dispersión, es la que mejor nos proporciona la variación de los datos con respecto a la media aritmética, su valor se encuentra en relación directa con la dispersión de los datos, a mayor dispersión de ellos, mayor desviación típica, y a menor dispersión, menor desviación típica.

Desviación estándar de la población.

La desviación estándar de la población es simplemente la raíz cuadrada de la varianza de la población. Como la varianza es el promedio de las distancias al cuadrado que van desde las observaciones a la media, la desviación estándar es la raíz cuadrada del promedio de las distancias al cuadrado que van desde las observaciones a la media. La desviación estándar está en las mismas unidades que las que se usaron para medir los datos.

La raíz cuadrada de un número positivo puede ser tanto positiva como negativa. Cuando tomamos la raíz cuadrada de la varianza para calcular la desviación estándar, los estadísticos solamente consideran la raíz cuadrada positiva.

Para calcular la varianza o la desviación estándar, construimos una tabla utilizando todos los elementos de la población.

Usos de la desviación estándar.

La desviación estándar nos permite determinar, con un buen grado de precisión, dónde están localizados los valores de una distribución de frecuencias con relación a la media. El teorema de Chebyshev dice que no importa qué forma tenga la distribución, al menos 75% de los valores caen dentro de $\pm$ 2 desviaciones estándar a partir de la media de la distribución, y al menos 89% de los valores caen dentro de $\pm$ 3 desviaciones estándar a partir de la media.

Con más precisión:

- Aproximadamente 68% de los valores de la población cae dentro de $\pm$ 1 desviación estándar a partir de la media.
- Aproximadamente 95% de los valores estará dentro de $\pm$ 2 desviaciones estándar a partir de la media.
- Aproximadamente 99% de los valores estará en el intervalo que va desde tres desviaciones estándar por debajo de la media hasta tres desviaciones estándar por arriba de la media.

<u>Resultado estándar:</u>

La desviación estándar es también útil para describir qué tan lejos las observaciones individuales de una distribución de frecuencias se apartan de la media de la distribución. Una medida que se conoce como *resultado estándar* nos da el número de desviaciones estándar que una observación en particular ocupa por debajo o por encima de la media:

1.2.3. Rango.

Rango.- Dato mayor menos dato menor.
Rango:

Es la primera medida que vamos a estudiar, se define como la diferencia existente entre el valor mayor y el menor de la distribución. Lo notaremos como **R.** Realmente no es una medida muy significativa en la mayoría de los casos, pero indudablemente es muy fácil de calcular.

Hemos estudiado varias medidas de centralización, por lo que podemos hablar de desviación con respecto a cualquiera de ellas, sin embargo, la más utilizada es con respecto a la media.

RANGO

El rango, o R, es la diferencia entre los valores más alto y más bajo incluidos en un conjunto de datos. Así, cuando My representa al mayor valor del grupo y Mn al menor, el rango de datos no agrupados es

R = My - Mn

EJEMPLO Durante un mes de verano, los ocho vendedores de una empresa de equipos de calefacción y aire acondicionado vendieron los siguientes números de unidades centrales de aire acondicionado: 8,11, 5, 14, 8,11, 16, 11. El rango del número de unidades vendidas es

R =My - Mn = 16 - 5 = 11.0 unidades

Nota: Para efectos de comparación, generalmente reportamos las medidas de variabilidad con un decimal adicional al nivel original de medición.

RANGOS MODIFICADOS

Un rango modificado es un rango que se construye eliminando algunos de los valores extremos de cada una de las porciones finales de la distribución. El 50% central es el rango entre los valores en el 25o. punto percentil y el 75o. punto percentil de la distribución. De este modo, también es el rango entre el primer y tercer cuartiles de la distribución. Por este motivo, al rango del 50% central suele llamársele rango intercuartil (RIC). Así,

RIC = Q3 – Q1,

Otros rangos modificados de uso común son el 80% central, el 90% central y el 95% central.

EJEMPL0 Los datos de ventas de unidades centrales de aire acondicionado presentados en el ejemplo anterior son, en orden ascendente: 5, 8, 8, 11, 11, 11, 14, 16. En consecuencia, el número de observaciones es N= 8 en estos datos de la población. Para calcular el rango intercuartil, primero debemos determinar los valores en Q3 (el 75o. punto percentil) y Q1, (el 25o. punto percentil) y después restar Q1, de Q3:

$$Q_3 = X_{[(75N/100)+(1/2)]} = X_{[6+(1/2)]} = X_{6.5} = 12.5$$

$$Q_1 = X_{[(25N/100)+(1/2)]} = X_{[2+(1/2)]} = X_{2.5} = 8.0$$

$$RIC = Q_3 - Q_1 = 12.5 - 8.0 = 4.5 \text{ unidades}$$