ANEXO MATEMÁTICA #2

Estadística: Medidas de variabilidad

Luego de determinar las medidas de tendencia central en un conjunto de datos, lo que se estudia en ese conjunto de datos, son las medidas de dispersión. Algunas de estas son: el recorrido, recorrido intercuartílico, la varianza y la desviación estándar. Entre más disgregados o dispersos estén los datos, mayores serán los valores de estas medidas y, entre más cercanamente agrupados estén los datos, menores serán los valores obtenidos con estas medidas de dispersión.

1. Recorrido

Es la medida de dispersión de datos más fácil de obtener. Se llama recorrido (también se le conoce como amplitud o rango) a la diferencia que hay entre el valor máximo (V_{max})y el valor mínimo (V_{min}) obtenidos en un conjunto de datos.

Recorrido = $V_{max} – V_{min}$

Ejemplo 1

Obtenga la amplitud o recorrido en la siguiente muestra de datos.

4, 5, 4, 9, 5, 3, 4, 5, 7, 10, 3, 6.

Solución

En este caso el valor máximo de la lista de datos anterior, es 10 y el valor mínimo es 3, es decir: $V_{max} = 10$; $V_{min} = 3$

Por lo que el recorrido lo calculamos así:
Recorrido = $V_{max} – V_{min} = 10 – 3 = 7$

Respuesta

El recorrido de la muestra de datos es 7.

NOTA: *Lo que indica el recorrido es que los datos de la lista se encuentran a una distancia de 7 unidades en la recta numérica.*

2. Recorrido intercuartílico

Cuartil: Los cuartiles son tres valores que dividen, de igual manera porcentual, un conjunto ordenado de datos. Normalmente se usa la letra Q para denotar simbólicamente a los cuartiles: Q_1 (primer cuartil), Q_2 (segundo cuartil), Q_3 (tercer cuartil). En general, Q_1 coincide con la mediana de la primera mitad de los valores, Q_2 con la mediana del conjunto de datos y Q_3 con la mediana de la segunda mitad de los valores.

Recorrido intercuartílico: Es la diferencia entre el tercer cuartil (Q_3) y el primer cuartil (Q_1).

Recorrido intercuartílico: $R_i = Q_3 – Q_1$

Ejemplo 2

Calcule el recorrido, los cuartiles y recorrido intercuartílico, de la siguiente muestra de datos.

Longitud del lado de unas cajas de baldosas de cerámica defectuosas dadas en centímetros

32,9	32,8	33,1	33,0	33,2
33,1	33,1	33,0	32,8	32,8
33,2	33,1	32,7	32,9	33,3

Solución

a) Primero hay que ordenar los datos de menor a mayor:

$x_1 = 32,7$ $x_2 = 32,8$ $x_3 = 32,8$ $x_4 = 32,8$
$x_5 = 32,9$ $x_6 = 32,9$ $x_7 = 33,0$ $x_8 = 33,0$
$x_9 = 33,1$ $x_{10} = 33,1$ $x_{11} = 33,1$ $x_{12} = 33,1$
$x_{13} = 33,2$ $x_{14} = 33,2$ $x_{15} = 33,3$

b) Hallar el recorrido:

Recorrido = $V_{max} - V_{min}$

Recorrido = 33,3 – 32,7 = 0,6

> ***Recuerde***
>
> *x_i es el valor posicional de cada dato, es decir que x_1 es el primer dato, x_2 el segundo dato y así sucesivamente.*

c) Hallar los cuartiles:

Q_2: Primero se puede calcular la media, ya que Me = Q_2, entonces como son 15 datos, se tiene que: Me = x_8 = 33,0, es decir que Q_2 = 33,0

Q_1: Como $Q_2 = x_8$, la mitad de la primera mitad de los datos estaría entre x_4 y x_5 es decir que $Q_2 = \frac{x_4 + x_5}{2} = \frac{32,8 + 32,9}{2} = 32,85$, es decir que Q_1 = 32,85.

Q_3: Como $Q_2 = x_8$, la media de la segunda mitad de los datos estaría entre x_{11} y x_{12} es decir que Q_3 = 33,1 porque $Q_{11} = Q_{12}$.

d) Recorrido intercuartílico:
$R_i = Q_3 - Q_1$ = 33,1 – 32,85 = 0,25

Respuesta: El recorrido es 0,6 cm, los cuartiles son: Q_1 = 32,85 cm; Q_2 = 33,0 cm; Q_3 = 33,1 cm y el recorrido intercuartílico es 0,25 cm

3. Varianza y desviación estándar

Varianza: La varianza o variancia es una medida de dispersión de datos en la que se calcula el promedio del valor de cada dato menos la media aritmética del total de los datos. Normalmente, la varianza se simboliza mediante la letra "s^2" cuando se calcula para una muestra y con σ^2 cuando se calcula para la población del estudio. Al calcular la varianza se obtienen las unidades al cuadrado, lo cual no tiene mucho sentido. Por ejemplo, si el estudio es acerca de las estaturas dadas en centímetros, la varianza obtiene unidades en centímetros cuadrados. Esta medida es útil porque nos permite hallar fácilmente la desviación estándar.

Recuerde que el promedio en una muestra es:

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, es decir, la suma de todos los datos entre el número de datos.

a) Si se calcula la varianza para una muestra de datos se utiliza la siguiente fórmula:

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

en la que x_i es cada dato de la muestra y $\bar{x}$ es el promedio de los datos.

> ***Recuerde***
>
> ***n** es el número de datos de la muestra.*
>
> ***N** es el número de datos de la población.*
>
> ***i** siempre es igual a **1** porque es el primer dato.*

b) Si se calcula la varianza para una población, se usa la siguiente fórmula:

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

en la que **N** es el total de datos de la población,

μ es el promedio de la población: $\mu = \frac{\sum_{i=1}^{N} x_i}{N}$

(note que es la misma fórmula para hallar el promedio de la muestra)

*Es importante notar que en el caso de muestras relativamente pequeñas se utiliza como divisor **n – 1** ya que se hace una mejor estimación poblacional, mientras que si la muestra es muy grande no importa si se usa **N** o **n – 1** como divisor.*

4. Desviación estándar

La desviación estándar es una medida de dispersión de datos muy usada en estadística. En general, indica el promedio con el que se alejan los datos de la media aritmética ($\bar{x}$) de la muestra o de la población en estudio.

Su cálculo es relativamente sencillo si se obtiene primero la varianza, al realizar la raíz cuadrada de esta.

$s = \sqrt{s^2}$	$\sigma = \sqrt{\sigma^2}$
Desviación estándar de la muestra	Desviación estándar de la población

Las unidades obtenidas por la desviación estándar son las mismas que las unidades del estudio, es decir, que si un estudio trata de estaturas en centímetros de un grupo de personas, obtenemos la desviación estándar en centímetros (a diferencia de la varianza que se obtendrían en centímetros cuadrados).

Ejemplo 3

En una escuela se realizó un estudio acerca de las estaturas de las estudiantes de segundo grado que provienen de un determinado barrio aledaño. Para ello se tomó la muestra de la estatura 16 alumnas dada en centímetros.

Los valores en centímetros obtenidos luego de las mediciones son los siguientes:

126	127	121	122
130	123	125	121
123	122	123	124
120	125	122	119

Calcule la varianza y la desviación estándar de la muestra.

Solución

a) Cálculo de la varianza

Promedio: $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

$$\frac{126+127+121+122+130+123+125+121+123+122+123+124+120+125+122+119}{16} = \bar{x} = \frac{1973}{16} \approx 123{,}31$$

La estatura promedio de las alumnas es aproximadamente 123,31 cm.

Para hallar la varianza:

- **Paso 1:** se calcula los $x_i - \bar{x}$ (cada dato menos el promedio).
- **Paso 2:** se eleva al cuadrado los resultados anteriores $(x_i - \bar{x})^2$.
- **Paso 3:** se suman los resultados de $(x_i - \bar{x})^2$.
- **Paso 4:** el total se divide por n – 1.

Observe la siguiente tabla:

Dato x_i	PASO 1 $x_i - \bar{x}$	PASO 2 $(x_i - \bar{x})^2$
126	126 – 123,31 = 2,6875	$2{,}6875^2 \approx 5{,}23$
127	127 – 123,31 = 3,6875	$3{,}6875^2 \approx 13{,}60$
121	121 – 123,31 = –2,3125	$(-2{,}3125)^2 \approx 5{,}35$
122	122 – 123,31 = –1,3125	$(-1{,}3125)^2 \approx 1{,}72$
130	130 – 123,31 = 6,6875	$6{,}6875^2 \approx 44{,}72$
123	123 – 123,31 = –0,3125	$(-0{,}3125)^2 \approx 0{,}98$
125	125 – 123,31 = 1,6875	$6{,}6875^2 \approx 44{,}72$
121	121 – 123,31 = –2,3125	$(-2{,}3125)^2 \approx 5{,}35$
123	123 – 123,31 = –0,3125	$(-0{,}3125)^2 \approx 0{,}98$
122	122 – 123,31 = –1,3125	$(-1{,}3125)^2 \approx 1{,}72$
123	123 – 123,31 = –0,3125	$(-0{,}3125)^2 \approx 0{,}98$
124	124 – 123,31 = 0,6875	$0{,}6875^2 \approx 0{,}47$
120	120 – 123,31 = –3,3125	$(-3{,}3125)^2 \approx 10{,}97$
125	125 – 123,31 = 1,6875	$1{,}6875^2 \approx 2{,}85$
122	122 – 123,31 = –1,3125	$(-1{,}3125)^2 \approx 1{,}72$
119	119 – 123,31 = –4,3125	$(-4{,}3125)^2 \approx 18{,}60$
	Suma de todos los $(x_i - \bar{x})^2$	≈ 159,96

PASO 3

El resultado de sumar $(x_i - \bar{x})^2$ se divide entre $n-1$:

$$s^2 = \frac{159{,}96}{\underbrace{n-1}_{\text{PASO 4}}} = \frac{159{,}96}{16-1} = \frac{159{,}96}{15} \approx 10{,}664 \text{ cm}^2$$

b) Cálculo de la desviación estándar:

$$s = \sqrt{s^2} = \sqrt{10{,}664} \approx 3{,}27 \text{ cm}$$

Respuesta: La varianza es aproximadamente igual a 10,664 cm² y la desviación estándar es aproximadamente igual a 3,27 cm.

Representación gráfica

Diagrama de cajas

Un diagrama de caja y brazos es una representación gráfica para un conjunto de datos que permite determinar de manera visual entre qué par de números se encuentra el 50% de los datos, con respecto a la posición de la mediana (Me) y se representa mediante un rectángulo. Los brazos son dos segmentos de recta en los que se representa el otro 50% de los datos (25% en cada brazo). Se puede hacer de manera horizontal o vertical.

Ejemplo 4

Las calificaciones de 19 estudiantes en un examen son:

45, 50, 52, 55, 70, 70, 70, 73, 75,75, 78, 80, 80, 80, 83, 85, 90, 95, 98.

Represente los datos anteriores con un diagrama de cajas.

Solución

Ubique cada dato con su posición (x_i):

45	50	52	55	70	70	70	73	75	75
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}

78	80	80	80	83	85	90	95	98
x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}

Para la construcción del diagrama de cajas y brazos se requieren 5 números: el valor mínimo, el valor máximo, el extremo inferior de la caja, el extremo superior de la caja y la mediana (Me).

Primero: se ubica la mediana: $Me = x_{10} = 75$

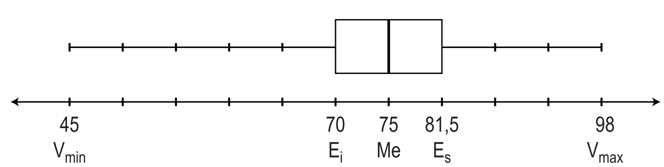
La posición de la mediana queda en el dato número 10.

Segundo: El valor de los extremos de la caja se calculan contando el número de unidades a partir del valor mínimo o máximo que indique el siguiente procedimiento: se usa la parte entera de la posición de la mediana más una unidad (Me + 1) y el resultado dividido entre 2: $\frac{Me+1}{2}$

a) Para el extremo inferior (E_i): $\frac{10+1}{2} = 5{,}5$ es decir que a partir del valor mínimo debemos contar de manera ascendente 5,5 unidades. En otras palabras, el extremo inferior de la caja queda entre x_5 y x_6 que coincide con 70.

b) Para el extremo superior (E_s): $\frac{10+1}{2} = 5{,}5$ es decir que a partir del valor máximo debemos contar de manera descendente 5,5 unidades, o sea, que el extremo superior de la caja queda entre x_{14} y x_{15} que sería: $\frac{80+83}{2} = 81{,}5$

Respuesta



Medidas relativas

En ocasiones sucede que se tienen datos de dos grupos diferentes y se requiere compararlos respecto de su variabilidad, como lo podría ser la comparación de las masas de dos especies de animales distintas. Por ejemplo, que se realice un estudio de la variabilidad de las masas de las vacas y las de los perros. Con el coeficiente de variación, se podría dar cuenta de cuál especie varía más respecto de su masa.

El coeficiente de variación se puede calcular con las siguientes fórmulas:

$CV = \frac{s}{\bar{x}} \cdot 100$ CV para muestras	$CV = \frac{\sigma}{\mu} \cdot 100$ Desviación estándar de la población

Ejemplo 5

Calcule el coeficiente de variación de una muestra de 10 vacas y 10 perros que hay en una determinada finca, de acuerdo con la siguiente tabla:

Masas (en kilogramos) de animales de una granja

Vacas	600	630	650	660	630	625	630	670	595	640
Perros	4	3,5	6	2,2	8	12	5	10,5	3,5	8,5

Solución

Cálculo del coeficiente de variación de las vacas:

Promedio: $\bar{x} = 633$ kg desviación estándar: s = 23,71 kg

$$CV = \frac{s}{\bar{x}} \cdot 100 = \frac{23,71}{633} \cdot 100 = 3,75\%$$

Cálculo del coeficiente de variación de los perros:

Promedio: $\bar{x} = 3,75$ desviación estándar: s = 3,29 kg

$$CV = \frac{s}{\bar{x}} \cdot 100 = \frac{3,29}{3,75} \cdot 100 = 87,73\%$$

Respuesta: Las vacas presentan una variación de 3,75% mientras que los perros 87,73%. Esto podría interpretarse que posiblemente los perros, sean de razas diferentes o también que mezclaron perros cachorros con adultos.

Estandarización

La estandarización es un procedimiento por el cual se obtiene la posición de un valor particular x en términos de desviaciones estándar que lo separan del promedio de los datos. El valor z se calcula mediante la fórmula:

$z = \frac{x - \bar{x}}{s}$ Valor z para muestras	$z = \frac{x - \mu}{\sigma}$ Valor z para poblaciones

Recuerde

x: valor del dato

$\bar{x}$: promedio en la muestra

μ: promedio en la población

s: desviación estándar de la muestra

σ: desviación estándar de la población

Ejemplo 6

Determine los valores z para $x_1 = 92$ y $x_2 = 68$, con respecto a las calificaciones de un examen de un grupo de estudiantes, en el que el promedio fue 73,89 y la desviación estándar 14,74.

Solución

Para x_1: $z = \frac{x_1 - \bar{x}}{s} = \frac{92 - 73,89}{14,74} \approx 1,23$

Para x_2: $z = \frac{x_2 - \bar{x}}{s} = \frac{68 - 73,89}{14,74} \approx -0,40$

Respuesta: Los datos obtenidos anteriormente significan que el valor 92 está a una y un cuarto desviaciones estándar por arriba de la media, mientras que para el valor 68, indica que está a menos de la mitad de una desviación estándar por debajo de la media.

ACTIVIDAD

1. Observe la siguiente lista de datos

4, 4, 5, 5, 5, 6, 6, 7, 8, 8, 9, 10, 10, 11, 11, 11, 11, 14, 16, 20, 28

De acuerdo con la lista de datos anterior, calcule:

a) El recorrido
b) Los cuartiles y el recorrido intercuartílico
c) La varianza y la desviación estándar
d) Represente los datos de la lista en un diagrama de cajas

2. Observe las siguiente tabla que muestra las temperaturas máximas registradas en Nicoya y en San José durante 11 días del mes de marzo.

Día	1	2	3	4	5	6	7	8	9	10	11
Nicoya	31°	35°	38°	39°	35°	33°	34°	37°	40°	34°	33°
San José	24°	26°	27°	28°	22°	20°	22°	24°	26°	22°	20°

a) Construya un diagrama de cajas para cada conjunto de datos (Nicoya y San José).

b) Calcule la desviación estándar y la media aritmética de cada conjunto de datos.

c) Calcule el coeficiente de variación de las temperaturas de cada ciudad.

3. Lea la siguiente situación:

Un estudiante obtuvo un 73 en el examen de la materia 1 y un 85 en el examen de la materia 2. El desea saber en cuál de las dos materias obtuvo un mejor rendimiento. En la materia 1 el promedio fue 59 y la desviación estándar 8,5. En la materia 2 la media aritmética fue 83 y la desviación estándar 5,5.

De acuerdo con el texto anterior, calcule el **valor z** para cada examen de cada materia y determine en cuál de las dos materias obtuvo mejor rendimiento.

Bibliografía de consulta

Gómez Barrantes, Miguel. *Elementos de Estadística Descriptiva*. EUNED, 20.ª reimpresión de la 3.ª edición. San José, Costa Rica, 2010.

http://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S0253-29482003000200003, visto el 10/04/2016, a las 14:47. (Revista Costarricense de Ciencias Médicas, Aileen Fernández Ramírez, José Moncada Jiménez, vol.24 n.3-4 San José Jul. 2003).

Johnson, Robert. *Estadística Elemental*. EDIMUSA, 1.ª Edición al español, México D.F., México, 1995.

Ministerio de Educación Pública de Costa Rica, *Programas de Estudio de Matemáticas*, San José, Costa Rica, 2012.

Quintana, Carlos. *Elementos de Inferencia Estadística*. EUCR, 1.a edición, San José, Costa Rica, 1989.

Elaborado por: Prof. Federico Scriba Pasos
Profesor de Matemática