Análisis Cuantitativo I

Medidas de Dispersión

Carlos Cardona

Universidad del Rosario

10 de febrero de 2017

Quiz

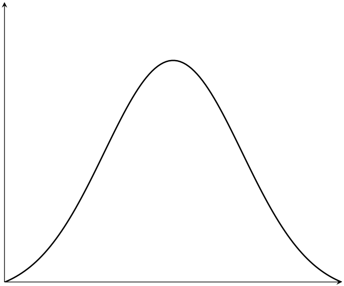
1. Hallar la media, mediana y moda de la siguiente tabla de distribución de frecuencias:

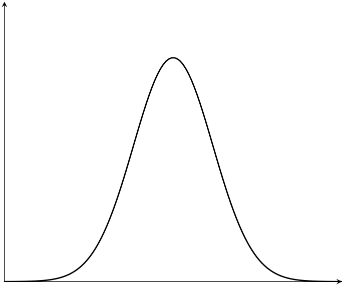
X	f
8	4
6	5
4	7
2	3

2. ¿Cuáles son las dos características de un diseño experimental?

Variabilidad

- Una distribución es descrita por su forma, su tendencia central y su dispersión.
- La dispersión mide las diferencias entre los valores en una distribución.
- Una buena medida de dispersión cumple dos objetivos:
  1. Describir la distribución. Específicamente, si los valores están agrupados entre sí o si están esparcidos a lo largo de una gran distancia.
  2. Medir cuan bien un valor individual representa toda la distribución.


Peso


Estatura

El Rango

El Rango

Es la distancia cubierta por los valores en una distribución, es decir, la distancia entre menor y el mayor valor.

- Se calcula de la siguiente manera:

$$Rango = X_{max} - X_{min} + 1$$

- En ocasiones es más útil reportar los valores mínimo y máximo que reportar el rango.
- Como el rango no tiene en cuenta todos los valores de la distribución, no es una medida precisa de la variabilidad de toda la distribución.

La Desviación Estándar

La Desviación Estándar

Describe la forma en que los valores de una variable se dispersan a lo largo de la distribución en relación a la media.

- La desviación estándar poblacional se denota σ_X, mientras que s_X es la notiación para la desviación estándar muestral.
- Se calcula siguiendo la fórmula:

$$s_X = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$$

Cálculo de la Desviación Estándar

1. Calcular la media.
2. Determinar el valor de desviación, o distancia de la media, para cada valor individual.
   - Para una distribución con una media de $\bar{X}$=50, si el valor X=53, el valor de desviación es:

$$X - \bar{X} = 53 - 50 = 3$$

   - Si el valor X=45, el valor de desviación será:

$$X - \bar{X} = 45 - 50 = -5$$

   - El valor de desviación nos dice la distancia del valor X a la media y la dirección de esa distancia.

3. Elevar al cuadrado los valores de desviación y sumar esos cuadrados.

$$Suma \quad de \quad cuadrados = \sum (X - \bar{X})^2$$

4. Dividir la suma de cuadrados entre $n - 1$.
   - Dado que la suma de cuadrados depende del número de observaciones, no es una medida estándar.
   - Por lo tanto, se promedia la suma de cuadrados por el número de valores menos 1.
5. El resultado de esto es llamado **varianza**. Cuyo símbolo es σ_X^2 o s_X^2.

$$s_X^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$$

1. Sacar raíz cuadrada a la varianza para obtener la desviación estándar.
   - Aunque la varianza es muy usada en algunos métodos inferenciales, el concepto de distancia al cuadrado no es una medida descriptiva intuitiva.
   - Por ejemplo, mencionar que entre Bogotá y Medellín hay 175561 kilómetros cuadrados no resulta informativo.
   - Por el contrario, si la sacamos raíz $\sqrt{175561} = 479$ kilómetros, el dato es mucho más significativo.

Un ejemplo

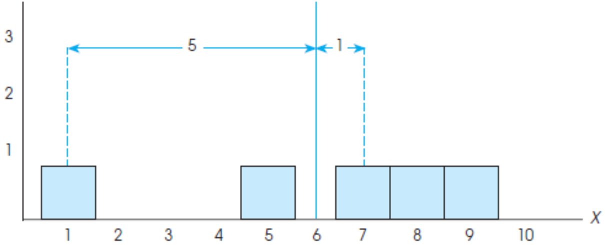
Valor de X	Desviación $X - \bar{X}$	Desviación^2 $(X - \bar{X})^2$
1	-5	25
9	3	9
5	-1	1
8	2	4
7	1	1
$\bar{X} = \frac{30}{5} = 6$	$\sum(X - \bar{X}) = 0$	$\sum(X - \bar{X})^2 = 40$

$$s_X^2 = \frac{\sum(X - \bar{X})^2}{n-1} = \frac{40}{4} = 10$$

$$s_X = \sqrt{10} = 3{,}16$$

Características de la Desviación Estándar

Figura 1



- Sumar la misma constante a cada valor, NO modifica la desviación estándar.
- Multiplicar cada valor por la misma constante, aumenta la desviación estándar en la misma proporción.

¿Por qué dividir sobre $n-1$?

- Imaginemos una población de N=6 valores: 0,0,3,3,9,9. La media $\mu = 4$ y la varianza $\sigma^2 = 14$.
- Ahora elegimos muestra de n=2 de esta población.

Muestra	Valor 1	Valor 2	$\bar{X}$	Varianza Usando n	Varianza Usando $n-1$
1	0	0	0.00	0.00	0.00
2	0	3	1.50	2.25	4.50
3	0	9	4.50	20.25	40.50
4	3	0	1.50	2.25	4.50
5	3	3	3.00	0.00	0.00
6	3	9	6.00	9.00	18.00
7	9	0	4.50	20.25	40.50
8	9	3	6.00	9.00	18.00
9	9	9	9.00	0.00	0.00
		Totales	36.00	63.00	126.00

Valores Z

- Supongamos que en un examen cualquiera, la nota recibida es $X = 76$. ¿Es buena o mala la nota?

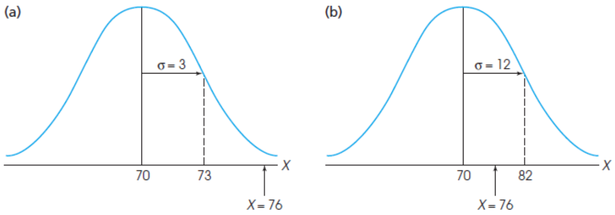
Valores Z

- Supongamos que en un examen cualquiera, la nota recibida es $X = 76$. ¿Es buena o mala la nota?
- Si la media es $\mu = 70$, sé que estoy 6 puntos por encima de la media.

Valores Z

- Supongamos que en un examen cualquiera, la nota recibida es $X = 76$. ¿Es buena o mala la nota?
- Si la media es $\mu = 70$, sé que estoy 6 puntos por encima de la media.
- Aún teniendo información de la media, no es posible saber dónde está localizado el valor de la nota.

Figura 2



- La ubicación relativa del valor dentro de una distribución depende tanto de la media como de la desviación estándar.
- Los valores Z tienen dos objetivos:
  1. Reportar la ubicación exacta de un valor dentro de una distribución.
  2. Permitir la comparación entre dos distribuciones estandarizadas.
- ¿Se puede comparar un puntaje de 64 en la sección de matemáticas de Saber Pro con un puntaje de 66 directamente, si se sabe que cada prueba fue realizada en períodos diferentes?
- Si la media de ambos exámenes fue 62, además $\sigma_1 = 1$ y $\sigma_2 = 4$ ¿A quién le fue mejor?

Valor Z

Un valor Z reporta la ubicación precisa de cada valor X dentro de la distribución. Su signo señala si el valor está por encima o por debajo de la media. Además, el valor numérico especifica la distancia de la media al contar el número de desviaciones estándar entre el valor y la media.

- La fórmula para calcular el valor Z es la siguiente:

$$Z_X = \frac{X - \bar{X}}{s_X}$$

- La unidad de medida del valor Z son el número de desviaciones estándar (SD).

- Volvamos al ejemplo de la prueba Saber Pro.
- Para el puntaje de 64, donde $\bar{X} = 62$ y $\sigma_1 = 1$:

$$Z_X = \frac{64 - 62}{1} = \frac{2}{1} = +2SD$$

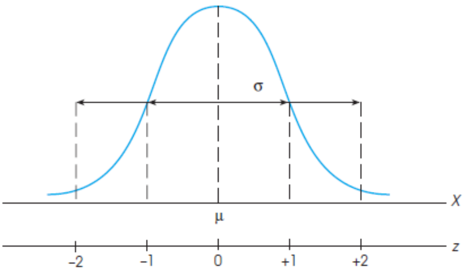
- Para el puntaje de 66, donde $\bar{X} = 62$ y $\sigma_1 = 4$:

$$Z_X = \frac{66 - 62}{4} = \frac{4}{4} = +1SD$$

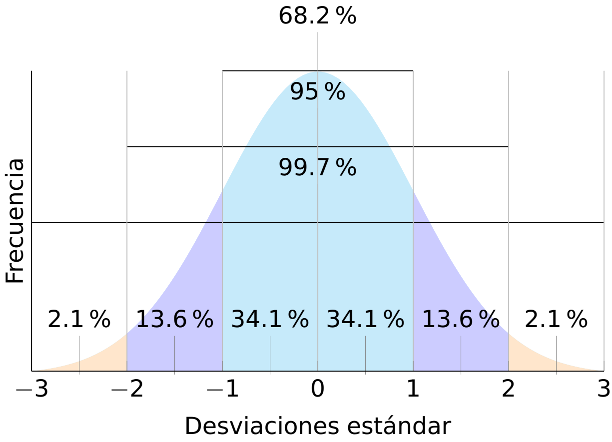
- La fórmula en muchas ocasiones no es necesaria. Si $\bar{X} = 10$ y $s_x = 2$, cuál es el valor de Z para un X=8?

- Al estandarizar una distribución:
  1. su forma se mantiene.
  2. La media se convierte en $\mu = 0$.
  3. La desviación estándar será $\sigma = 1$.

Figura 3



La Regla 68-95-99



La Utilidad de la Regla

- Imaginemos una muestra de 1000 mujeres de la universidad. Su peso medio es de $\bar{X} = 60kgs$ y su desviación estándar es $s_X = 5kgs$.
- 500 mujeres tendrán menos de 60 kgs.
- Cerca de 680 mujeres pesarán entre 55 y 65 kgs.
- Alrededor de 950 mujeres pesan entre 50 y 70 kgs.
- Aproximadamente 3 pesarán menos de 45 y más de 75 kgs.