#vivaFCA

Proceso de selección 2023

LICENCIATURA EN ACTUARÍA

SESIÓN 3

Medidas numéricas

MEDIDAS NUMÉRICAS

INTRODUCCIÓN

- Si las medidas se calculan con datos de una muestra se llaman **Estadísticos muéstrales.**
- Si estas medidas las calcula con los datos de una población se llaman **parámetros poblacionales.**

MEDIDAS DE LOCALIZACIÓN

MEDIA

- La medida de localización más importante es la media, o valor promedio, de una variable.
- La media proporciona una medida de localización central de los datos.
- Si los datos son datos de una muestra, la media se denota (x); si los datos son datos de una población, la media se denota con la letra griega **μ.**

MEDIA

MEDIA MUESTRAL

$$\bar{x} = \frac{\Sigma x_i}{n}$$

EJEMPLO

Egresado	Sueldo mensual inicial ($)	Egresado	Sueldo mensual inicial ($)
1	3450	7	3490
2	3550	8	3730
3	3650	9	3540
4	3480	10	3925
5	3355	11	3520
6	3310	12	3480

MEDIANA

- La mediana es otra medida de localización central. Es el valor de en medio en los datos ordenados de menor a mayor (en forma ascendente).
- Cuando tiene un número impar de observaciones, la mediana es el valor de en medio.
- Cuando la cantidad de observaciones es par la mediana es definida como el promedio de las dos observaciones de en medio.

MEDIANA

- Aunque la media es la medida de localización central más empleada, en algunas situaciones se prefiere la mediana. A la media la influyen datos en extremo pequeños o considerablemente grandes.
- Si hay algunos sueldos demasiado altos, la mediana proporciona una medida de tendencia central mejor que la media.

MODA

- La moda es el valor que se presenta con mayor frecuencia.
- Hay situaciones en que la frecuencia mayor se presenta con dos o más valores distintos. Cuando esto ocurre hay más de una moda. Si los datos contienen más de una moda se dice que los datos son bimodales.
- Si contienen más de dos modas, son multimodales.

PERCENTILES

- Un percentil aporta información acerca de la dispersión de los datos en el intervalo que va del menor al mayor valor de los datos.
- Cerca de p por ciento de las observaciones tienen valores menores que el percentil p y aproximadamente $(100 - p)$ por ciento de las observaciones tienen valores mayores que el percentil p .

PERCENTIL

El percentil p es un valor tal que por lo menos p por ciento de las observaciones son menores o iguales que este valor y por lo menos $(100 - p)$ por ciento de las observaciones son mayores o iguales que este valor.

CÁLCULO DEL PERCENTIL p

Paso 1. Ordenar los datos de menor a mayor (colocar los datos en orden ascendente).

Paso 2. Calcular el índice i

$$i = \left(\frac{p}{100}\right)n$$

donde p es el percentil deseado y n es el número de observaciones.

Paso 3. (a) Si i *no es un número entero, debe redondearlo*. El primer entero mayor que i denota la posición del percentil p.

(b) Si i *es un número entero*, el percentil p es el promedio de los valores en las posiciones i e $i + 1$.

EJEMPLO DE PERCENTIL

3310, 3355, 3450, 3480, 3480, 3490, 3520, 3540, 3550, 3650, 3730, 3925.

Elaborar percentil 85
Elaborar percentil 50

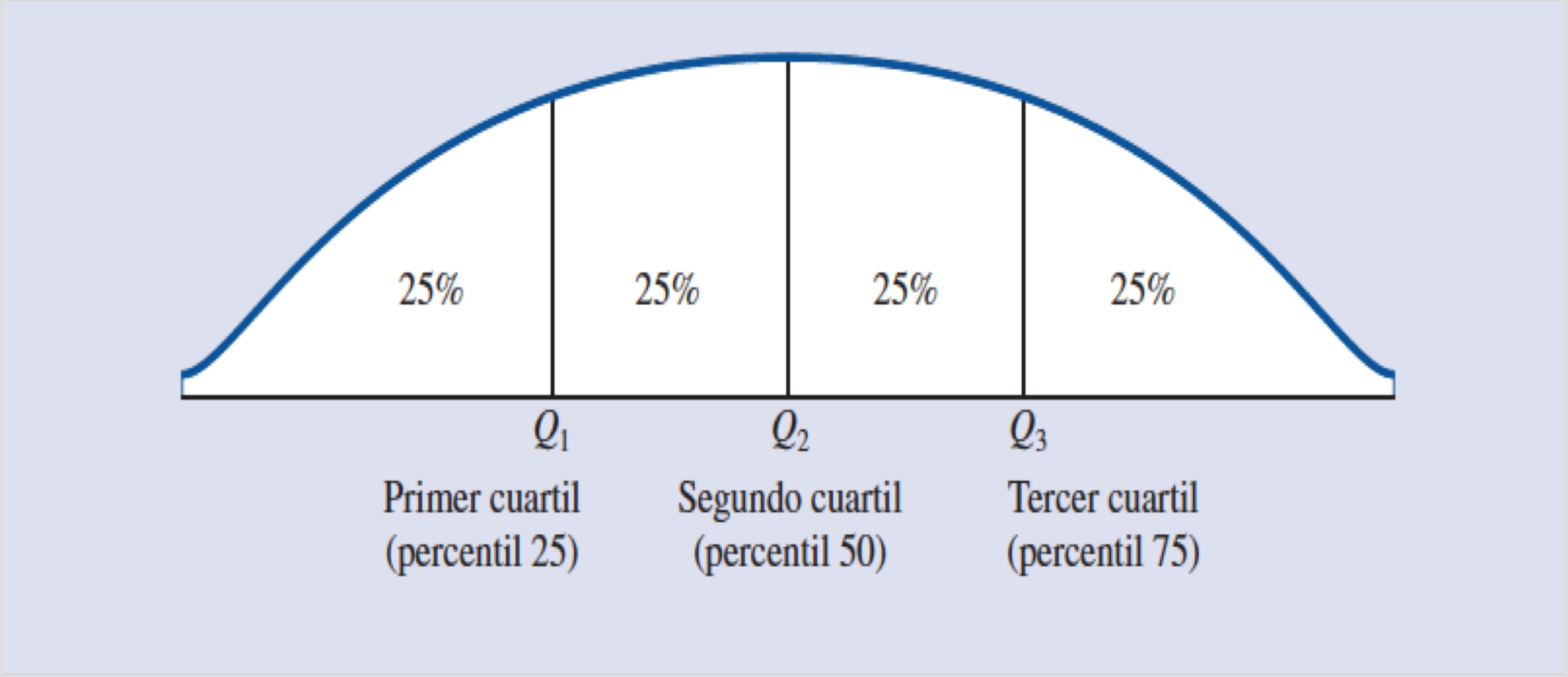
CUARTILES

- Con frecuencia es conveniente dividir los datos en cuatro partes; así, cada parte contiene una cuarta parte o 25% de las observaciones.

Q1 primer cuartil, o percentil 25

Q2 segundo cuartil, o percentil 50

Q3 tercer cuartil, o percentil 75


25%
25%
25%
25%
Q_1
Q_2
Q_3
Primer cuartil
(percentil 25)
Segundo cuartil
(percentil 50)
Tercer cuartil
(percentil 75)

MEDIDAS DE VARIABILIDAD

RANGO

RANGO

$$\text{Rango} = \text{Valor mayor} - \text{Valor menor}$$

Egresado	Sueldo mensual inicial ($)	Egresado	Sueldo mensual inicial ($)
1	3450	7	3490
2	3550	8	3730
3	3650	9	3540
4	3480	10	3925
5	3355	11	3520
6	3310	12	3480

RANGO INTERCUARTÍLICO

- Esta medida de variabilidad es la diferencia entre el tercer cuartil **Q**3 y el primer cuartil **Q**1 . En otras palabras, el rango intercuartílico es el rango en que se encuentra el 50% central de los datos.

RANGO INTERCUARTÍLICO

$$\mathrm{IQR} = Q_3 - Q_1$$

VARIANZA

- La varianza es una medida de variabilidad que utiliza todos los datos. La varianza está basada en la diferencia entre el valor de cada observación ($\mathbf{x_i}$) y la media.
- Si los datos son de una población, el promedio de estas desviaciones elevadas al cuadrado es la varianza poblacional. La varianza poblacional se denota con la letra griega $\boldsymbol{\sigma}^2$. En una población en la que hay N observaciones y la media poblacional es $\boldsymbol{\mu}$.

VARIANZA POBLACIONAL

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

- La varianza muestral , que se denota por s^2 , se define como sigue.

VARIANZA MUESTRAL

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

EJEMPLO

- Numero de estudiantes en un grupo:
  - 46
  - 54
  - 42
  - 46
  - 32

Egresado	Sueldo mensual inicial ($)	Egresado	Sueldo mensual inicial ($)
1	3450	7	3490
2	3550	8	3730
3	3650	9	3540
4	3480	10	3925
5	3355	11	3520
6	3310	12	3480

Sueldo mensual (x_i)	Media muestral ($\bar{x}$)	Desviación respecto de la media ($x_i - \bar{x}$)	Cuadrado de la desviación respecto de la media $(x_i - \bar{x})^2$
3450	3540	−90	8 100
3550	3540	10	100
3650	3540	110	12 100
3480	3540	−60	3 600
3355	3540	−185	34 225
3310	3540	−230	52 900
3490	3540	−50	2 500
3730	3540	190	36 100
3540	3540	0	0
3925	3540	385	148 225
3520	3540	−20	400
3480	3540	−60	3 600
		0	301 850
		$\Sigma(x_i - \bar{x})$	$\Sigma(x_i - \bar{x})^2$

Empleando la ecuación (3.5),

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n - 1} = \frac{301\,850}{11} = 27\,440.91$$

DESVIACIÓN ESTÁNDAR

- La desviación estándar se define como la raíz cuadrada positiva de la varianza. Continuando con la notación adoptada para la varianza muestral y para la varianza poblacional, se emplea s para denotar la desviación estándar muestral y σ para denotar la desviación estándar poblacional.

DESVIACIÓN ESTÁNDAR

$$\text{Desviación estándar muestral} = s = \sqrt{s^2}$$

$$\text{Desviación estándar poblacional} = \sigma = \sqrt{\sigma^2}$$

COEFICIENTE DE VARIACIÓN

- En algunas ocasiones se requiere un estadístico descriptivo que indique cuán grande es la desviación estándar en relación con la media. Esta medida es el coeficiente de variación y se representa como porcentaje.

- En general, el coeficiente de variación es un estadístico útil para comparar la variabilidad de variables que tienen desviaciones estándar distintas y medias distintas.

COEFICIENTE DE VARIACIÓN

$$\left(\frac{\text{Desviación estándar}}{\text{Media}} \times 100\right)\%$$

- Una empresa quiere sacar un nuevo producto al mercado, para este fin el administrador de la empresa le pide a tres de sus ingenieros que cada uno realice un proceso piloto. Una vez terminados los procesos piloto el administrador tomo una muestra de 10 productos para cada proceso y midió los tiempos que tardaban dentro del proceso. Los datos son los siguientes:

	PROCESO A	PROCESO B	PROCESO C
	156	160	236
	166	165	238
	148	184	257
	160	192	242
	139	197	282
	151	172	253
	158	189	270
	167	179	256
	142	200	267
	219	193	259
MEDIA	160.6	183.1	256
DES. EST	22.5398413	13.7149553	14.6515073

- ¿Cuál de los procesos debería de escoger el administrador de la empresa?
- Si el administrador de la empresa cree que los tiempos de producción debe de estar dentro de S+1. ¿Cuál es el porcentaje de productos que quedarían por fuera de esta medida?

MEDIDAS DE LOCALIZACIÓN RELATIVAS Y DETECCIÓN DE VALORES ATÍPICOS

INTRODUCCIÓN

- Además de las medidas de localización, variabilidad y forma, interesa conocer también la ubicación relativa de los valores de un conjunto de datos. Las medidas de localización relativa ayudan a determinar qué tan lejos de la media se encuentra un determinado valor.

PUNTO Z

- A partir de la media y la desviación estándar, se puede determinar la localización relativa de cualquier observación. Suponga que tiene una muestra de n observaciones, en que los valores se denotan $\mathbf{x}_1$, $\mathbf{x}_2$, . . . , $\mathbf{x}_\mathbf{n}$. Suponga además que ya determinó la media muestral, que es y la desviación estándar muestral, que es **s** . Para cada valor $\mathbf{x}_\mathbf{i}$ existe otro valor llamado punto **z** .

PUNTO z

$$z_i = \frac{x_i - \bar{x}}{s}$$

donde

z_i = punto z para x_i

$\bar{x}$ = media muestral

s = desviación estándar muestral

PUNTO Z

- Al punto z también se le suele llamar valor estandarizado . El punto z_i puede ser interpretado como el número de desviaciones estándar a las que x_i se encuentra de la media . Por ejemplo si $z_1 =$ 1.2, esto indica que x_1 es 1.2 desviaciones estándar mayor que la media muestral. De manera similar, z_2= 0.5 indica que x_2 es 0.5 o 1/2 desviación estándar menor que la media muestral.

- Puntos z mayores a cero corresponden a observaciones cuyo valor es mayor a la media, y puntos z menores que cero corresponden a observaciones cuyo valor es menor a la media. Si el punto z es cero, el valor de la observación correspondiente es igual a la media.

Número de estudiantes en un grupo (x_i)	Desviación respecto de la media ($x_i - \bar{x}$)	Puntos z $\left(\dfrac{x_i - \bar{x}}{s}\right)$
46	2	$2/8 = 0.25$
54	10	$10/8 = 1.25$
42	−2	$-2/8 = -0.25$
46	2	$2/8 = 0.25$
32	−12	$-12/8 = -1.50$

TEOREMA DE CHEBYSHEV

- El teorema de Chebyshev permite decir qué proporción de los valores que se tienen en los datos debe estar dentro de un determinado número de desviaciones estándar de la media.

TEOREMA DE CHEBYSHEV

Por lo menos $(1 - 1/z^2)$ de los valores que se tienen en los datos deben encontrarse dentro de z desviaciones estándar de la media, donde z es cualquier valor mayor que 1.

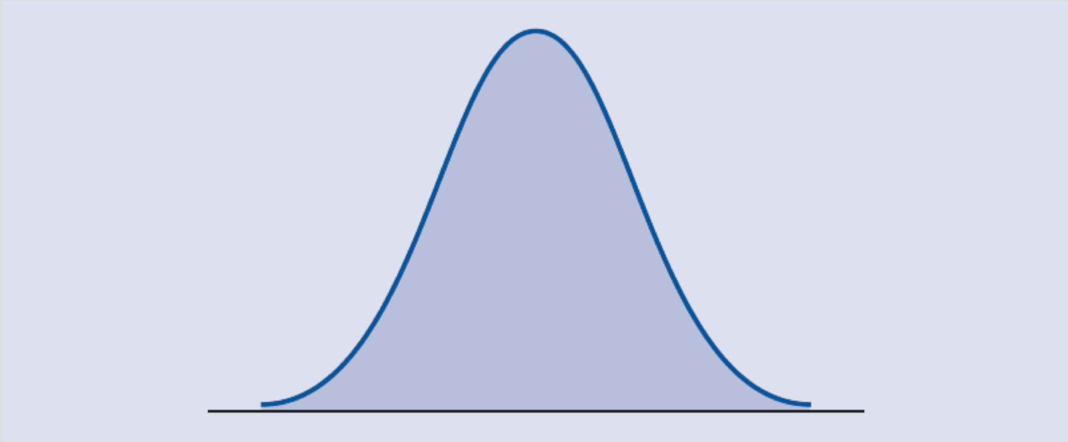
- De acuerdo con este teorema para z 2, 3 y 4 desviaciones estándar se tiene:
  - Por lo menos 0.75, o 75%, de los valores de los datos deben estar dentro de z 2 desviaciones estándar de la media.
  - Al menos 0.89, o 89%, de los valores deben estar dentro de z 3 desviaciones estándar de la media.
  - Por lo menos 0.94, o 94%, de los valores deben estar dentro de z 4 desviaciones estándar de la media.

EJEMPLO

- Para dar un ejemplo del uso del teorema de Chebyshev, suponga que en las calificaciones obtenidas por 100 estudiantes en un examen de estadística para la administración, la media es 70 y la desviación estándar es 5. ¿Cuántos estudiantes obtuvieron puntuaciones entre 60 y 80?, ¿y cuántos tuvieron puntuaciones entre 58 y 82?

REGLA EMPÍRICA

- En muchas aplicaciones prácticas los datos muestran una distribución simétrica con forma de montaña o de campana.
- Cuando se cree que los datos tienen aproximadamente esta distribución, se puede emplear la regla empírica para determinar el porcentaje de los valores de los datos que deben encontrarse dentro de un determinado número de desviaciones estándar de la media.



REGLA EMPÍRICA

Cuando los datos tienen una distribución en forma de campana:

- Cerca de 68% de los valores de los datos se encontrarán a no más de una desviación estándar desde la media.
- Aproximadamente 95% de los valores de los datos se encontrarán a no más de dos desviaciones estándar desde la media.
- Casi todos los valores de los datos estarán a no más de tres desviaciones estándar de la media.

DETECCIÓN DE OBSERVACIONES ATÍPICAS

- Para identificar las observaciones atípicas se emplean los valores estandarizados (puntos **z**). Recuerde que la regla empírica permite concluir que en los datos con una distribución en forma de campana, casi todos los valores se encuentran a no más de tres desviaciones estándar de la media. Por tanto, si usa los puntos **z** para identificar las observaciones atípicas, es recomendable considerar cualquier dato cuyo punto **z** sea menor que 3 o mayor que 3 como una observación atípica. Debe examinar la exactitud de tales valores y si en realidad pertenecen al conjunto de datos.