POSTERIS LVMEN MORITVRVS EDAT
1827
UNIVERSIDAD DEL CAUCA

Universidad del Cauca

icontec
ISO 9001
ISO 9001:2015 SC- CER 450832

CERTIFIED
IQNet
MANAGEMENT SYSTEM
IQNet: CO- SC-CER450832

Una Acreditación con Rostro Humano

Hacia una Universidad comprometida con la paz territorial

Estadística

Ingeniería de Sistemas
Otros indicadores estadísticos
Clase #7

Coeficiente de variación (*cv*): se obtiene dividiendo la desviación estándar entre media aritmética, expresándose el resultado en términos porcentuales:

$$cv = \frac{s}{\bar{x}} 100\%$$

- Se emplea cuando se desea comparar dos o más distribuciones (con unidades de medida diferentes), con el fin de determinar cuál de ellas tiene mayor o menor variabilidad relativa.
- No resulta un buen índice de comparación si las dimensiones de los datos son muy diferentes.
- Hay que tener cuidad cuando dos distribuciones con medias aritméticas diferentes e iguales desviaciones estándar, es decir, con la misma variabilidad absoluta, dan diferentes resultados en su coeficiente de variación.

Considere dos distribuciones con $\bar{x} = 24.5$ y $\bar{y} = 30$, y con desviaciones $s_x = s_y = 2$ (tienen el mismo grado de variación absoluta) ¿Pero que pasa con cv?

Ejemplos

1. Supongamos un grupo de profesionales, que trabajan en un sector de la actividad industrial en un Colombia, tienen un salario promedio de $968.000 y varianza de $144.000.000. Otro grupo de empleados profesionales que trabajan en un Venezuela, en la misma actividad industrial, tienen un salario promedio de Bls. 85.700 y una desviación estándar de Bls. 800. Se quiere determinar cuál grupo de salarios presenta una menor variabilidad.
2. Los salarios mensuales (en promedio) de una muestra de empleados de Unicauca fue de $892.000, con una desviación estándar de $400.000. ¿Es o no representativo ese promedio de salarios?

- Si $cv \leq 30\%$, entonces $\bar{x}$ se considera representativo, y la muestra es homogénea.
- Si $cv > 30\%$, entonces $\bar{x}$ se considera no representativo, y la muestra es no homogénea.

Asimetría

Considere una población de tamaño N (o muestra) que tiene media μ, desviación estándar σ, moda M_o, mediana $\tilde{x}$, primer cuartil Q_1 y tercer cuartil Q_3.

- Coeficiente de asimetría de Pearson:

$$A_P = \frac{\mu - M_o}{\sigma}.$$

El coeficiente de asimetría de Pearson solamente se puede calcular si es una distribución unimodal.

- Coeficiente de asimetría de Bowley:

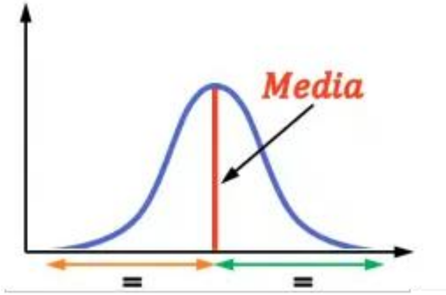
$$A_B = \frac{Q_3 + Q_1 - 2\tilde{x}}{Q_3 - Q_1}.$$

- Coeficiente de asimetría de Fisher:

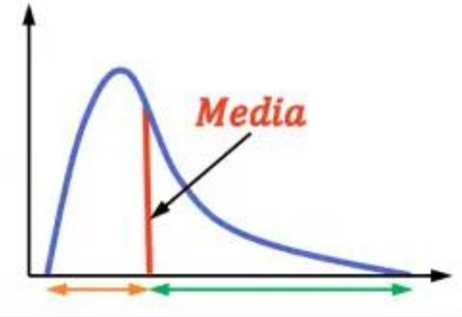
$$A_F = \frac{\sum_{i=1}^{N} (x_i - \mu)^3}{N\sigma^3}.$$

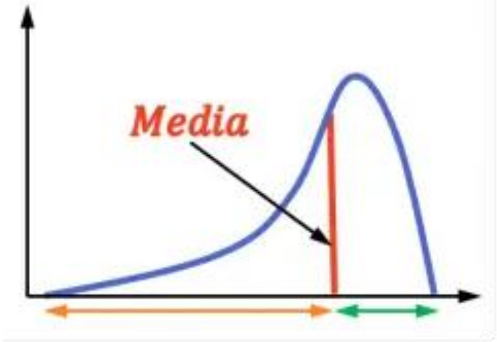
Interpretación del coeficiente de asimetría:

1. Si $A = 0$, la distribución es simétrica



2. Si $A > 0$, la distribución es asimétrica con sesgo positivo

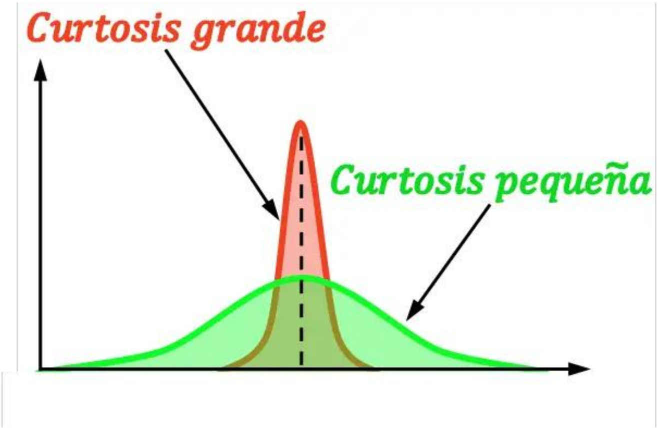


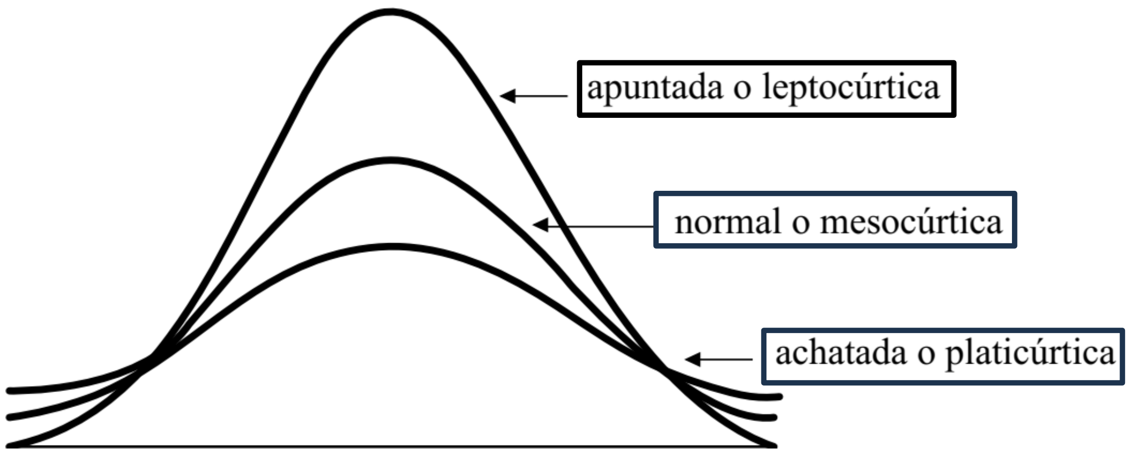


3. Si $A < 0$, la distribución es asimétrica con sesgo negativo.

Curtosis

También llamada apuntamiento, es una medida estadística que indica el grado de concentración de una distribución alrededor de su media, dicho de otra manera, es el grado de agudeza en la cima de la curva que representa la distribución.





Leptocúrtica: la distribución es muy apuntada, es decir, los datos están muy concentrados alrededor de la media.

Mesocúrtica: la curtosis de la distribución es equivalente a la curtosis de la distribución normal. No se considera ni apuntada ni achatada.

Platicúrtica: la distribución es muy achatada, es decir, la concentración en torno a la media es baja.

Coeficiente de curtosis:

$$A_p = \frac{\sum_{i=1}^{N} (x_i - \mu)^4}{N\sigma^4}$$

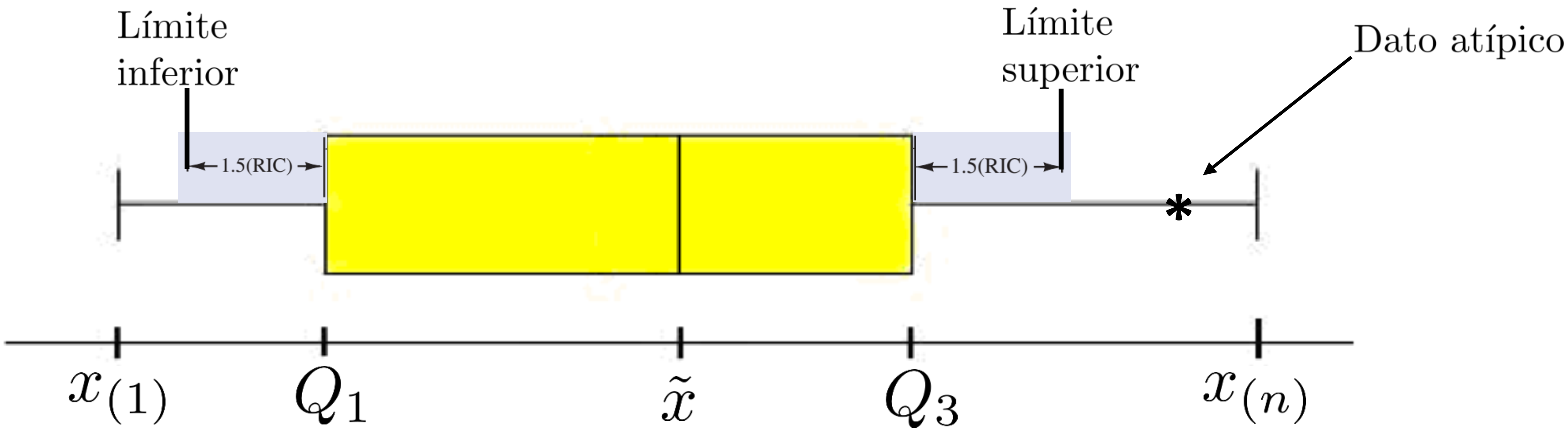
Interpretación del coeficiente de curtosis:

1. Si $A_p = 3$, la distribución es mesocúrtica (normal)
2. Si $A_p > 3$, la distribución es leptocúrtica (apuntada)
3. Si $A_p < 3$, la distribución es platicúrtica (achatada).

La curtosis y la asimetría son dos conceptos que se suelen estudiar juntos, ya que ambos sirven para describir la forma de una distribución, sin la necesidad de representación gráfica.

Diagrama de caja

Dispositivo gráfico que se usa para expresar en forma resumida, algunas medidas estadísticas de posición:



- La caja contiene el 50% de los datos.
- Rango intercuartílico: RIC$= Q_3 - Q_1$ ➡ Límites: $Q_1 - 1.5RIC$, y $Q_3 + 1.5RIC$
- Datos atípicos: Si $x_{(i)} < Q_1 - 1.5RIC$, o $x_{(i)} > Q_3 + 1.5RIC$, entonces $x_{(i)}$ se denomina **dato atípico**.

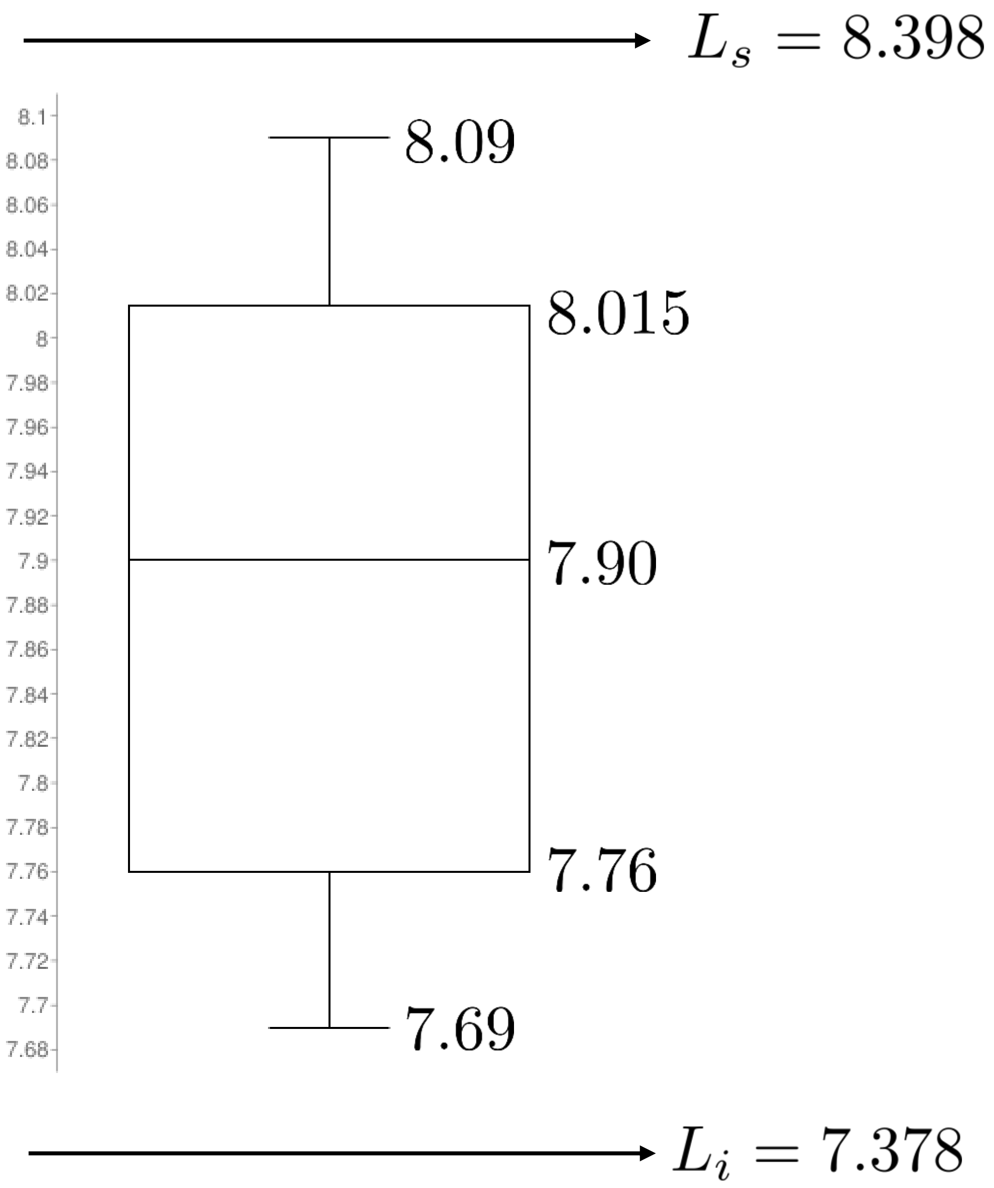
- Si los datos son simétricos, la mediana estará en el centro de la caja.

- Si los datos están sesgados, la mediana estará más cerca de la parte superior o inferior de la caja.

- Los bigotes se usan generalmente para indicar la variabilidad fuera de los cuartiles inferior y superior.

- Un rango intercuartil pequeño, es decir, una caja estrecha, indica baja dispersión del 50% de los datos centrales.

Ejemplo.



El departamento de control de calidad de cierta compañía de alimentos verifica el peso de un frasco de mantequilla de maní de ocho onzas. Los pesos de la muestra de nueve frascos fabricados la hora pasada son los siguientes:

7.69	7.72	7.8	7.86	7.90	7.94	7.97	8.06	8.09

Recuerde que: $x_{(1)} = 7.69$, $Q_1 = 7.76$, $\tilde{x} = 7.90$, $Q_3 = 8.015$, $x_{(9)} = 8.09$

$$L_i = 7.76 - 1.5(8.015 - 7.76) = 7.378$$

$$L_s = 8.015 + 1.5(8.015 - 7.76) = 8.398$$

Ejemplo.

Suponga que la bolsa de trabajo de una universidad del Cauca envía cuestionarios a los recién egresados de la carrera Ing. Civil solicitándoles información sobre sus sueldos (en dólares) mensuales iniciales. Los resultados se presentan en la siguiente tabla.

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

Elabore un diagrama de caja para estos datos, y responda las siguientes preguntas:

1. ¿La distribución de los datos es simétrica o sesgada positiva/negativa? (Verifique su respuesta mediante el coeficiente de simetría de Bowley)
2. ¿Hay datos atípicos en la muestra?

¡Gracias por su atención!

Universidad del Cauca

www.unicauca.edu.co

Ejercicios.

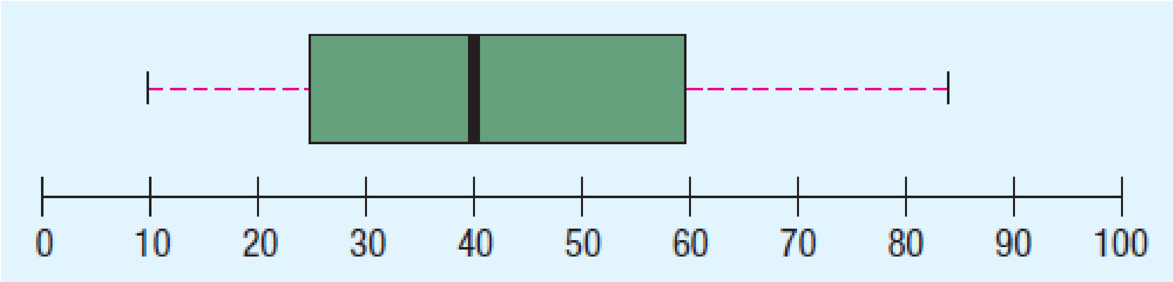
1. Para la media y la varianza de un conjunto de datos se han hallado, respectivamente, los valores 4 y 25. ¿Qué opinión le merece la media de esos datos?

2. Se analizaron en el segundo semestre del año pasado los gastos semanales de 30 empresas y se obtuvo en promedio de \$174 millones y una desviación estándar de \$9 millones. Se determinó luego que los contadores de cada una de las empresas habían prescindido de \$3 millones en los gastos por un error de apreciación. Corrigiendo las medidas enunciadas, obtener el coefciente de variación de los gastos de estas empresas.

3. Sumando 4 a cada uno de los números de la muestra: 2, 6, 5, 9, 1, se obtiene la muestra 6, 10, 9, 13 y 5. Comprobar que ambas muestras tienen la misma varianza, distintas medias y diferentes coefcientes de variación.

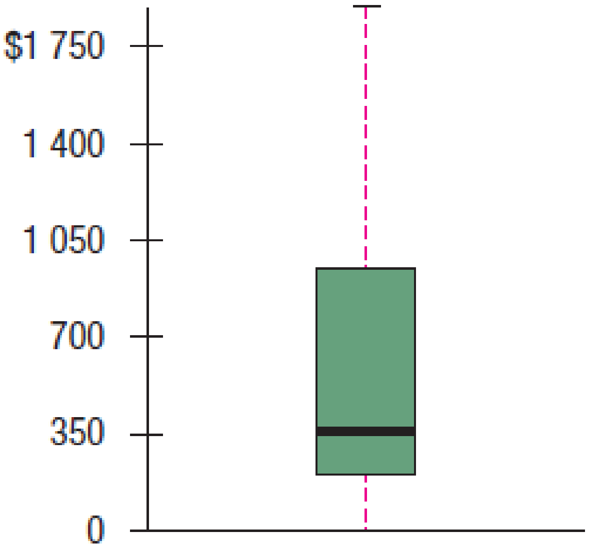
4. Multiplicando por 4 cada uno de los valores de la muestra: 3, 2, 0, 5, se obtiene la muestra: 12, 8, 0, 20, comprobar que ambas muestras tienen el mismo coefciente de variación.

5. El siguiente diagrama de caja muestra los activos en millones de dólares de cooperativas de crédito en Colombia.



¿Cuáles son los valores mínimo y máximo, los cuartiles primero y tercero, y la mediana? ¿Estaría usted de acuerdo en que la distribución es simétrica? ¿Hay datos atípicos? (Calcule los limites) ¿Cual es el rango intercuartílico?

6. El diagrama de caja muestra la suma que se gastaron en libros y suministros durante un año los estudiantes de cuarto semestre de universidades públicas.



(a) Calcule la mediana de la suma que se gastó.

(b) Calcule el primero y el tercer cuartiles de la cantidad que se gastó.

(c) Calcule el rango intercuartil de la cantidad que se gastó.

(d) ¿Más allá de qué punto un valor se considera dato atípico?

(e) Identifique cualesquiera datos atípicos y calcule su valor.

(f) ¿Es la distribución simétrica, o tiene sesgo positivo o negativo?

7. En un estudio sobre el rendimiento en millas por galón de gasolina de automóviles modelo 2011, la media fue de 27.5 y la mediana de 26.8. El valor más pequeño fue de 12.70 millas por galón y el más grande de 50.20. El primer y tercer cuartil fueron 17.95 y 35.45 millas por galón, respectivamente. Elabore un diagrama de caja y haga algún comentario sobre la distribución. ¿Es una distribución simétrica? ¿Hay datos atípicos?

8. En un estudio sobre el rendimiento en millas por galón de gasolina de automóviles modelo 2011, la media fue de 27.5 y la mediana de 26.8. El valor más pequeño fue de 12.70 millas por galón y el más grande de 50.20. El primer y tercer intercuartiles fueron 17.95 y 35.45 millas por galón, respectivamente. Elabore un diagrama de caja y haga algún comentario sobre la distribución.

9. Elabore un diagrama de caja para los datos del ejercicio 9 (farmaceuticas) de la clase 5, compare las cajas y los rangos intercuartilicos y tome una decisión.
10. Una empresa fabrica bombillas eléctricas de dos clases, A y B. Con base en muestras de la producción se sabe que las distribuciones de la duración en horas de esas bombillas son tales, que tienen las siguientes medias y varianzas.

TIPO	MEDIA	VARIANZA
A	800 horas	7.800
B	650 horas	5.400

    (a) Comparar ambas distribuciones en cuanto a su variabilidad absoluta y relativa.
    (b) Si se extrajo una bombilla de cada tipo y su duración fue de 700 y 630 horas, respectivamente, ¿cuál tipo de bombilla tiene menor posición relativa?
    (c) Determinar el coefciente de variación para el total de las bombillas examinadas.