

PRESENTACIÓN DEL AED

- Inicio del AED: John Tukey (1977)
Nuevas técnicas gráficas y analísticas
Perspectiva exploratoria

- EDC frente al AED



- Inconvenientes de la estadística clásica

- a) Parte de hipótesis difíciles de verificar

- b) Supone errores repartidos alrededor de un valor central

- c) Uso exclusivo de modelos lineales (relación entre variables)

- AED: nuevo enfoque metodológico, cuyo objetivo es entender el análisis de los datos.

■ En distribuciones univariantes, el AED nos informa sobre:

a) Localización, desviación y forma de la distribución de los datos.

b) Número y localización de agujeros y puntas.

c) Presencia y número de valores alejados.

■ Componentes principales del AED

1) Representaciones gráficas

- *) Steam and leaf (Tronco y hojas)
 - *) Boxplots (Diagrama de cajas)
- Detectan el conjunto de índices descriptivos.

2) Análisis de datos

Diferencias entre datos observados y valores ajustados. (RESIDUOS)
Modelos de ajuste (no lineales)

3) Transformación de datos

Simplificación del análisis.
Uso de funciones matemáticas simples como raíz cuadrada y logaritmo.

4) Resistencia

Valores de datos extraños no influyen en los resultados de un análisis.

5) Robustos

Busca estadísticos poco sensibles a desviaciones.

■ Campos de aplicación

- Ciencias Sociales
- Ciencias Humanas
- Ciencias de la Salud

.*PSICOLOGÍA* (tendencias, patrones de conductas, formación de actividades...)

.*HISTORIA y LINGÜÍSTICA* (descubrir indicadores de cambio histórico o lingüístico)

.*ECONOMÍA, SOCIOLOGÍA y PEDAGOGÍA* (empleo de técnicas del AED antes de confirmar modelos)

.*MEDICINA* (revelaciones en investigación...)

.*EMPRESAS* (rendimiento de plantilla, control de calidad...)

■ Conclusión

El AED no es solo un complemento a las técnicas de la EDC, sino que es una alternativa en caso de incumplimiento de alguna condición de aplicación, pues no es tan restrictivo en sus supuestos.

MEJOR CALIDAD DE ANÁLISIS DE DATOS GLOBALMENTE.

ORGANIZACIÓN, REDUCCIÓN Y REPRESENTACIÓN DE DATOS

■ ÍNDICES DEL AED

El AED desarrolla unos nuevos índices descriptivos basados en la mediana y en los parámetros de forma y posición.

Índices Descriptivos

AED	Estadística Clásica
Localización	Medidas de posición y tendencia central
Dispersión	Medidas de dispersión
Forma	Medidas de forma (simetría y curtosis)

- Lo que llamamos muestra en EDC → lote (batch) en AED.

Vamos a considerar el siguiente ejemplo con dos lotes:

LOTE 1

10
10
11
12
12
13
14
15
15
15
16
18
19

LOTE 2

10
10
11
12
12
13
14
15
15
15
16
18
57

Ejercicio: Calcular los estadísticos clásicos: media, mediana, moda, varianza, simetría y curtosis.

		LOTE1	LOTE2
N	Válidos	13	13
	Perdidos	0	0
Media		13,85	16,77
Mediana		14,00	14,00
Moda		15	15
Varianza		8,141	151,859
Asimetría		,310	3,368
Error típ. de asimetría		,616	,616
Curtosis		-,688	11,774
Error típ. de curtosis		1,191	1,191

A primera vista, los únicos valores no afectados por valores extremos son la media y la moda, que serán las más utilizadas en el AED.

La media está basada en todos los datos del lote... A veces nos interesa saber cómo están centrados los datos y cuando encontramos valores extremos o **outliers** (en este caso, la media no es representativa, y utilizamos la mediana). Además, la media se usa bajo condiciones de normalidad, cosa que no es necesaria para la mediana.

Por tanto, si hay casos extremos uso la mediana y sino los hay uso la media. Además si hay normalidad uso la media y en otro caso la mediana.

■ Índices de Localización

Al pretender buscar estimadores robustos, usamos los centiles o percentiles, pues no se ven afectados por valores extremos.

Ejercicio: Calcular los percentiles 10, 25, 50 ,75 y 90

Estadísticos

		LOTE1	LOTE2
N	Válidos	13	13
	Perdidos	0	0
Percentiles	10	10,00	10,00
	25	11,50	11,50
	50	14,00	14,00
	75	15,50	15,50
	90	18,60	41,40

El único afectado a primera vista es el percentil 90.

Dentro de los índices de localización, vamos a ver los siguientes:

■ Promedio de cuartiles (\bar{Q})

Hace un promedio entre los 2 cuartiles centrales que recoge el 50% de los datos

$$\bar{Q} = \frac{C_{25} + C_{75}}{2} = \frac{Q_1 + Q_3}{2}$$

Q recoge el 50% de los datos, eliminando la influencia de los valores extremos.

Para el lote 1 y para el lote 2

$$\bar{Q} = 13,5$$

$$\bar{Q} = 13,5$$

■ Trimedia (TRI)

$$TRI = \frac{M_d + \bar{Q}}{2} = \frac{M_d + \frac{C_{25} + C_{75}}{2}}{2} = \frac{C_{25} + 2M_d + C_{75}}{4}$$

TRI elimina el 25% de observaciones en cada extremo (eliminando así casos extremos)

Para el lote 1 TRI = 13,75, y para el lote 2, TRI = 13,75

■ Centrimea o media intercuartílica (MID)

$$MID = \frac{x_{c_{25}+1} + \dots + x_{c_{75}-1}}{n_i}$$

Elimina valores repetidos y valores extremos que no nos interesan (promedia datos entre el cuartil 25 y el 75, sin incluirlos).

En el cálculo no deben incluirse los valores repetidos, y debe procurarse que el número de éstos a un lado y otro de la Md sean los mismo, es decir, ni debe ser un número impar (se puede incluir uno de los valores repetidos en el “lado” que presente menos).

En estos índices destaca el uso del 50% central de los datos, y en especial de la Md. Si los valores se hallan agrupados, el valor de ésta será muy semejante al de la media aritmética clásica, sin embargo en caso de valores muy alejados la Md reflejará mejor el valor promedio del grupo.

Pasos para calcular la centrimedia

- 1º. Calcular los cuartiles 25 y 75, y colocar la muestra entre estos, sin coger sus valores.
- 2º. Eliminamos un dato de cada extremos (si están repetidos).
- 3º Si hubiera algún valor extremos que se repite también lo eliminamos

(Se procura que el número de datos sea par, y que no haya valores repetidos.)

■ Índices de Dispersión

■ Amplitud intercuartílica (IQR)

$$\text{IQR} = c75 - c25$$

Es más resistente que la desviación típica y más fácil de calcular, y nos indica el 50 % de los casos que hay.

Se usa para comparar lotes y hacer transformaciones de variables.

Lote – 1

IQR = 4

Lote – 2

IQR = 4

■ Mediana de las desviaciones absolutas (MAD)

$$\text{MAD} = \text{Md} (|x_i - \text{Md}|)$$

Pasos:

- 1º) Cálculo Md
- 2º) Calcular $|x_i - \text{Md}|$
- 3º) Reordenar los datos
- 4º) Calcular la “nueva” Md \rightarrow MAD

Ejercicio: Calcular la MAD de los lotes 1 y 2

■ Coeficiente de variación intercuartílico

Sirve para comparar distribuciones, y no está afectado por valores extremos:

$$CV_c = \frac{\frac{IQR}{2}}{Q} = \frac{c_{75} - c_{25}}{c_{75} + c_{25}}$$

Lote 1

$$CV_c = 0.15$$

Lote 2

$$CV_c = 0.15$$

El que sea más alto, indica que hay más dispersión.

■ Índices de FORMA

■ Índice de simetría de Yule (H_1)

(Mide la simetría en el centro de la distribución)

$$H_1 = \frac{c_{25} + c_{75} - 2M_d}{2M_d}$$

Interpretación:

$H_1 = 0 \rightarrow$ Simetría

$H_1 > 0 \rightarrow$ Asimétrica positiva (sesgada derecha)

$H_1 < 0 \rightarrow$ Asimétrica negativa (sesgada izqda.)

Lote – 1

Lote – 2

$H_1 = -0.03$

$H_1 = -0.03$

■ Índice de simetría de Kelly (H_2)

Mide la simetría en las colas

$$H_2 = M_d - \frac{c_{10} + c_{90}}{2}$$

Misma interpretación que el índice de Yule.

■ Juntando ambos índices obtenemos el siguiente:

$$H_3 = \frac{c_{10} + c_{90} - 2M_d}{2M_d} = \frac{-H_2}{M_d}$$

Misma interpretación que los otros dos.

■ Coeficiente de CURTOSIS

$$K_1 = \frac{c_{90} - c_{10}}{1,9(c_{75} - c_{25})}$$

Indica el apuntamiento de la distribución.
Compara la dispersión entre el 90% de los casos centrales y la existente en el 50%.

Interpretación:

$K_1 > 1 \rightarrow$ leptocúrtica (concentrada en el centro)

$K_1 = 1 \rightarrow$ mesocúrtica (normal)

$K_1 < 1 \rightarrow$ platicúrtica.

Lote – 1

$K_1 = 1,13$

Lote – 2

$K_1 = 4,13$

Ejemplo

- Supongamos que conocemos las calificaciones en una asignatura por un grupo de chicos y chicas:

2.1	5.0	6.1	8.7	2.2	4.0	5.0	7.5
2.2	5.0	6.5	8.7	2.3	4.0	5.2	7.5
2.7	5.0	7.2	8.8	2.5	4.2	5.3	8.2
3.2	5.0	7.3	8.8	2.9	4.3	5.5	9.0
3.5	5.0	7.5	9.2	3.0	4.4	5.6	10.0
3.7	5.1	7.5	9.2	3.5	4.5	6.0	10.0
4.0	5.2	7.7	9.5	3.5	5.0	6.0	10.0
4.0	5.3	8.2	10.0	3.5	5.0	6.0	
5.0	5.4	8.4	10.0	3.7	5.0	6.2	

1º) Calcular los estadísticos tradicionales y los del AED.

	NIÑOS	NIÑAS
N		
Media		
Moda		
Varianza		
Simetría		
Curtosis		