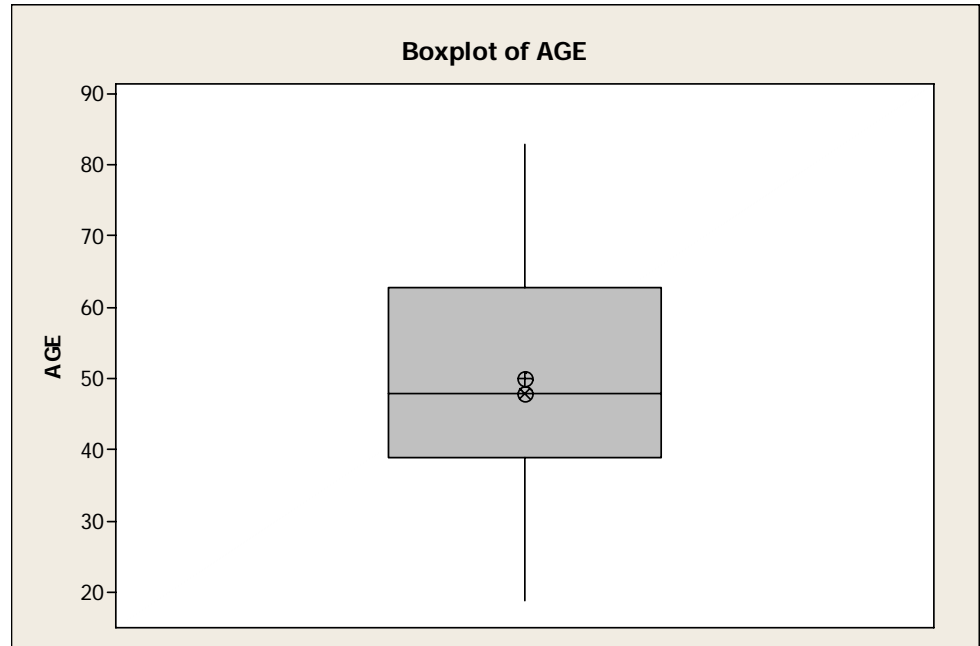


Medidas de Posición y Boxplot

Lección 11

Secc. 2.6

Prof. Pedro A. Torres
ESMA 3015



febrero de 2007

Objetivos

- Definir una medida de posición.
- Conocer y aprender a calcular las medidas de posición más usadas.
- Analizar un conjunto de datos usando las medidas de posición.
- Construir un boxplot e interpretarlo.

Medidas de Posición

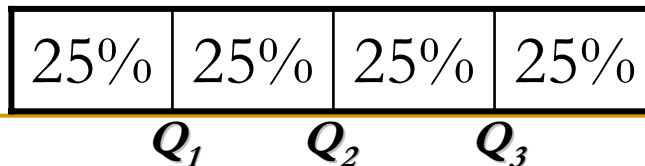
Son valores numéricos que describen la posición que tiene un dato o valor específico con respecto a todos los datos. Los más usados son los *cuartiles* y los *percentiles*.

Cuartiles

Son los valores que dividen los datos en cuatro partes iguales (cuartos).

- El primer cuartil (Q_1) es el valor tal que el 25% de los datos son menores a él.
- El segundo cuartil (Q_2) es el valor tal que el 50% de los datos son menores a él (la mediana).
- El tercer cuartil (Q_3) es el valor tal que el 75% de los datos son menores a él.

Los datos deben estar ordenados de menor a mayor



Medidas de Posición

Percentiles

Son los valores que dividen los datos en cien partes iguales. El *percentil* k (se escribe P_k) es el valor tal que el $k\%$ de los datos son menores a él.

Contrariamente, el $(100-k)\%$ de los datos son mayores a este valor.

Por ejemplo, si en la variable peso el percentil 13 es igual a 110 libras, esto indica que el 13% de las personas tienen un peso menor a 110 libras.

Notas:

- El cuartil 1, Q_1 , es el mismo P_{25} .
- El cuartil 2, Q_2 , es el mismo P_{50} . Ambos son iguales a la mediana.
- El cuartil 3, Q_3 , es el mismo P_{75} .

1%	1%	1%	1%	1%	...	1%	1%
----	----	----	----	----	-----	----	----

———— **Min.** P_1 P_2 P_3 P_4 P_5 ... P_{98} P_{99} **Máx.** ————

Procedimiento para calcular un *percentil k*

$$P_k$$

Objetivo: hallar el valor del percentil identificando el número de dato que corresponde al percentil que se quiere buscar.

- Ordene los datos de menor a mayor.
- Calcule $nk/100$.
- Si el número anterior ($nk/100$) es un decimal, aproxime al siguiente número. El dato correspondiente a este número es el P_k .
- Si el número ($nk/100$) es un entero entonces el P_k es igual al promedio del dato correspondiente a este número y el siguiente dato.

Ejemplo

Calcule los cuartiles (Q1, Q2 y Q3) y los percentiles 15 y 65 de los siguientes datos de edades:

37, 54, 15, 25, 30, 68, 26, 23, 25, 26, 27, 24

1. Ordene los datos:

15, 23, 24, 25, 25, 26, 26, 27, 30, 37, 54, 68

2. Calcule $nk/100$:

$$n = 12$$

Percentil (k)	$nk/100$	EL percentil es igual a:	Valor del Percentil
15	1.8	Dato 2	$P_{15} = 23$
25 (Q1)	3	Promedio entre el dato 3 y 4	$P_{25} = 24.5$
50 (Q2, mediana)	6	Promedio entre el dato 6 y 7	$P_{50} = 26$
65	7.8	Dato 8	$P_{65} = 27$
75 (Q3)	9	Promedio entre el dato 9 y 10	$P_{75} = 33.5$

Interpretación: por ejemplo, el $P_{65} = 27$, indica que el 65% de las personas tienen edades menores a 27 años.

Otras medidas

Cuartil medio: Esta es otra medida de tendencia central. Es el promedio entre Q_1 y Q_3 :

$$\text{Cuartil medio} = \frac{Q_1 + Q_3}{2}$$

Ejemplo: siguiendo con los datos del ejemplo anterior:

$$\text{Cuartil medio} = \frac{24.5 + 33.5}{2} = 29$$

La interpretación de esta medida es similar a un promedio. Sin embargo, su uso es limitado en la práctica.

Rango Inter-cuartílico: Esta es otra medida de dispersión. Es la diferencia entre Q_3 y Q_1 :

$$\text{Rango IQ} = Q_3 - Q_1$$

Ejemplo: Rango IQ = $33.5 - 24.5 = 9$

Valores grandes indican mayor dispersión o variabilidad en los datos.

Resumen con cinco medidas

Algunas veces se analiza la distribución de los datos usando cinco medidas importantes:

1. El mínimo
2. Q_1 o el P_{25}
3. La mediana
4. Q_3 o el P_{75}
5. El máximo

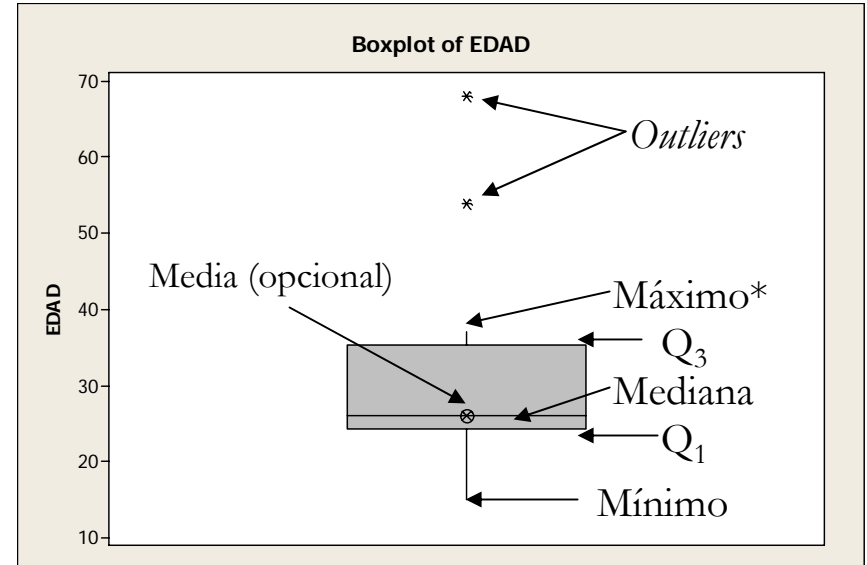
Ejemplo:

15	24.5	26	33.5	68
Mínimo	Q1	Mediana	Q3	Máximo

Boxplot

Es un diagrama con una caja y *bigotes* que representa las cinco medidas anteriores. Se construye de la siguiente manera:

- El ancho de la caja esta definido por el rango inter-cuartílico (Q1 y Q3). A mayor amplitud de la caja, mayor variabilidad en los datos.
- Del centro de la caja salen dos segmentos, uno hasta el mínimo y el otro hasta el máximo. Estos segmentos representan los datos que están por fuera del rango inter-cuartílico.
- Dentro de la caja se dibuja una línea horizontal la cual indica la mediana de los datos.
- Algunas veces, los *outliers* son representados por símbolos especiales (*, +, etc).

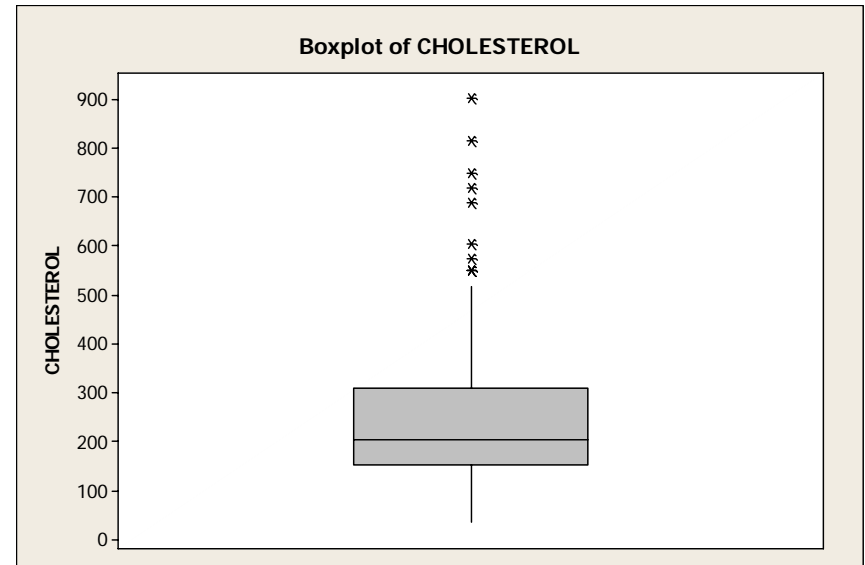
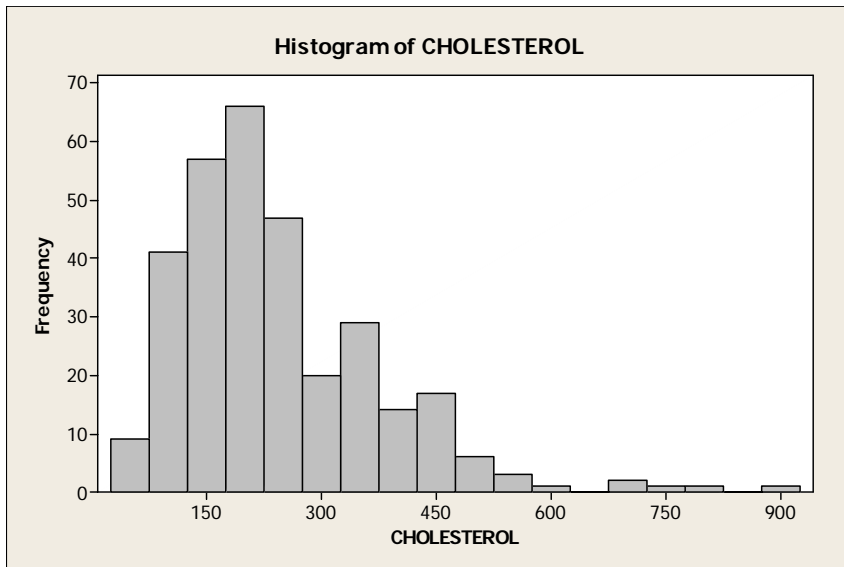


*Cuando existen *outliers*, los valores extremos (mínimo y máximo) se seleccionan sin tener en cuenta los *outliers*.

En Minitab, se hacen usando el menú **Graph**→**Boxplot**. Esta opción permite personalizar bastante el diagrama.

Interpretaciones del Boxplot

Una buena forma de ver como se interpreta el boxplot es comparándolo con un histograma



Histograma sesgado a la derecha.

Concentración alrededor de 200.

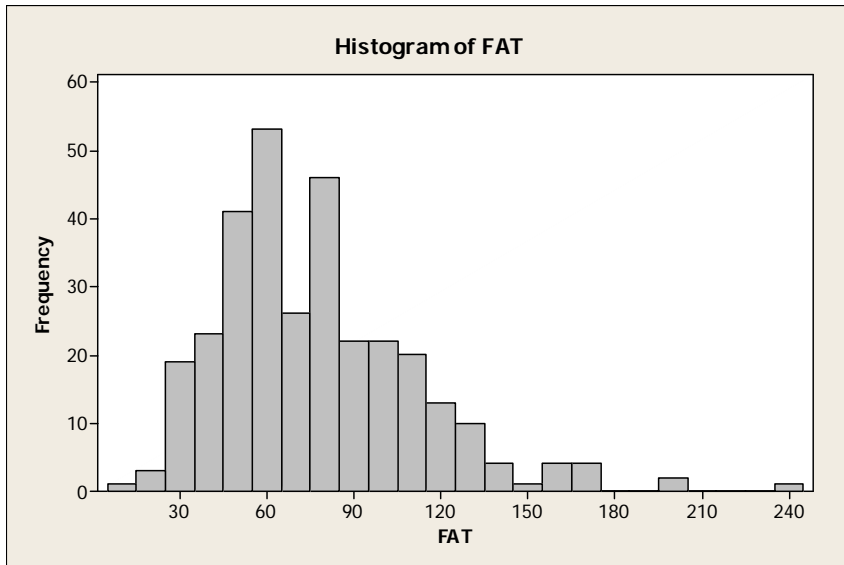
Presencia de *outliers*.

Los asteriscos indican que hay *outliers*.

El bigote largo hacia arriba y la línea que divide la caja más hacia abajo (mediana alrededor de 200) indica sesgo a la derecha.

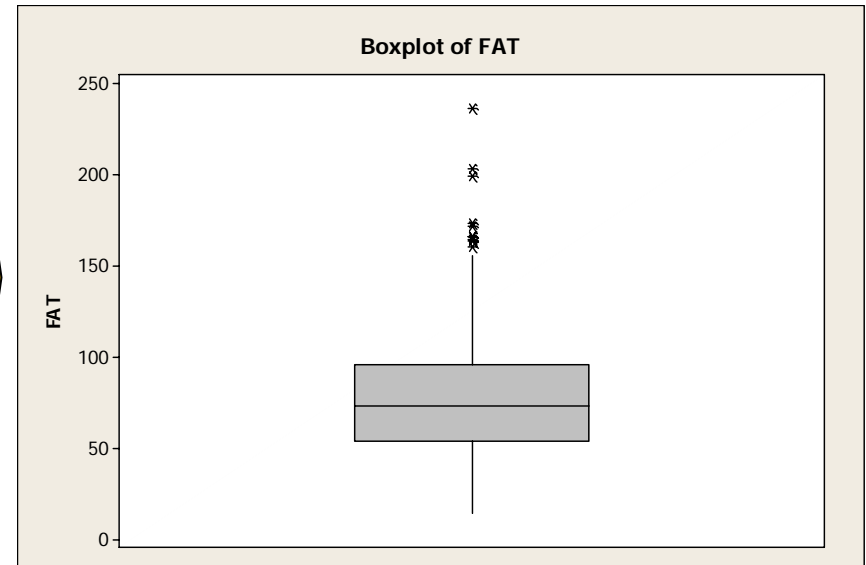
Interpretaciones del Boxplot

Otra comparación para una distribución mas cercana a la simetría.



Histograma más concentrado alrededor de un valor (70-80), pero con cierto sesgo a la derecha.

Presencia de *outliers*.

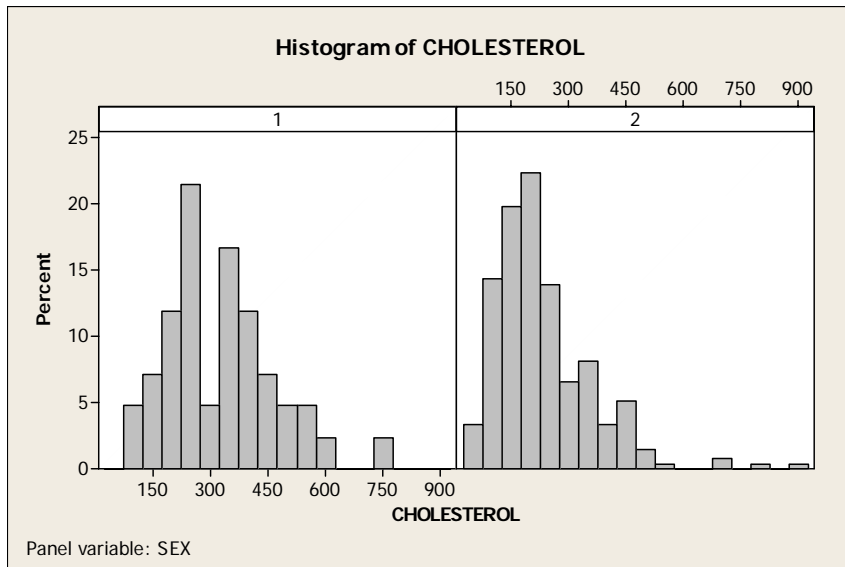


Los asteriscos indican que hay *outliers*.

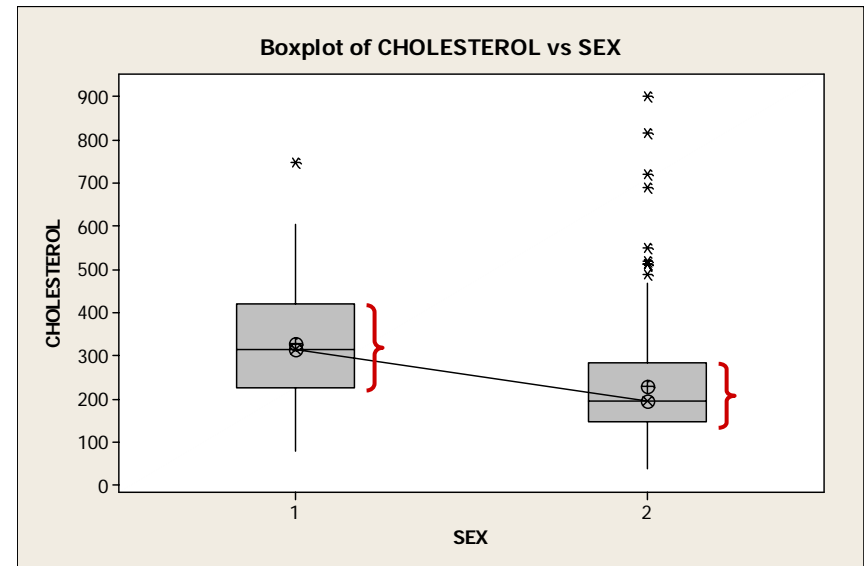
Los bigotes casi simétricos indican una distribución cercana a la simetría. Hay un ligero sesgo a la derecha.

Interpretaciones del Boxplot

El boxplot es una herramienta valiosa para comparar la distribución de dos o más grupos.



Mediante un histograma, muchas veces es difícil decidir que grupo es más disperso o cual es superior.



En este caso, el boxplot permite visualizar dos aspectos relevantes:

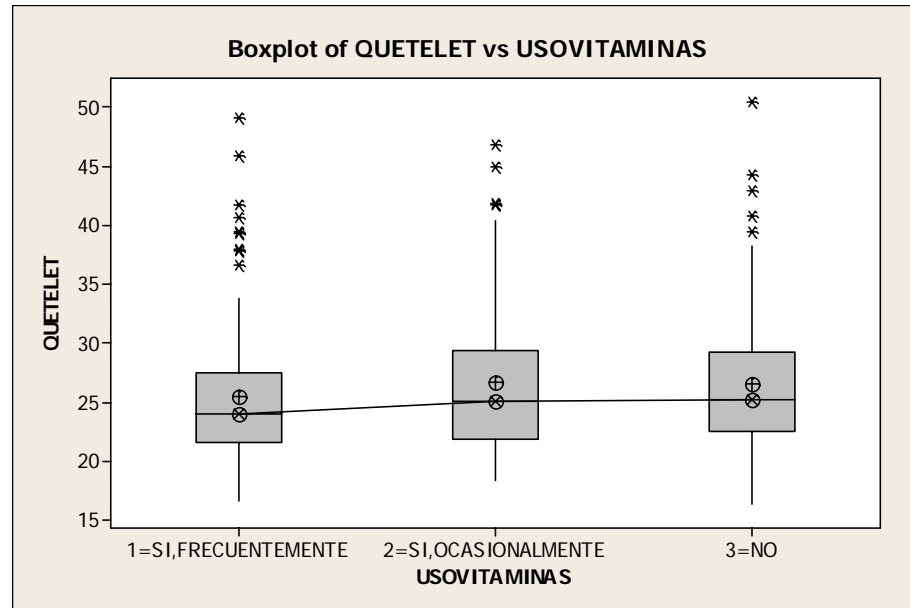
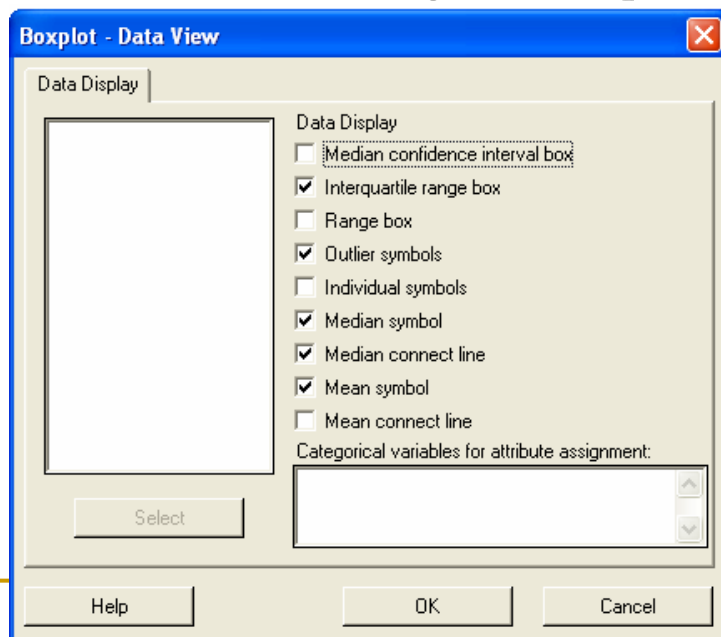
- Los datos de colesterol en los hombres (Sex=1) presentan mayor variabilidad que en las mujeres (caja más ancha).
- Los niveles de colesterol tienden a ser mayores en los hombres (caja y mediana más altas).

Ejercicio usando Minitab

Con la base de datos Retinol_Plasma, abrir Minitab y crear el siguiente boxplot.

Use la opción **Graph**→**Boxplot**→**Within Groups**.

Para personalizar la información que muestra el boxplot use la opción **Data View**. Seleccione las siguientes opciones:



¿Qué puede concluir acerca del Índice de Masa Corporal (Quetelet) entre los diferentes usos de vitaminas?

Lo que vimos:

- Las diferentes medidas de posición y cómo calcularlas.
- Interpretación de las medidas de posición.
- Uso del boxplot para analizar distribuciones de variables individualmente y comparando grupos.
- Como generar los boxplots usando Minitab.

Bibliografía

- **Just the Essentials of Elementary Statistics**, Robert, Johnson, Novena Edición, International Thomson (ISBN: 0-534-99945-X).

Recomendaciones:

- Leer las sección 2.6 del libro.