

- ☐ Consiste en una caja rectangular, donde los lados más largos muestran el Intervalo Intercuartílico (II). Este rectángulo está dividido por una línea vertical que coincide con la mediana (o segundo cuartil).
- ☐ La caja comprende el 50% de los valores centrales, extendiéndose entre el primer y el tercer cuartil (C1 y C3), es decir, coincide con II.
- ☐ Puede representarse tanto vertical como horizontalmente, pues sólo existe un eje, donde se colocan los datos o medidas.
- ☐ En el centro de la caja, mediante una aspa, se representa el valor de la media.

- ☐ Las líneas que sobresalen de la caja se llaman bigotes. Estos bigotes tienen un límite de prolongación ($1,5 \cdot II$), y marcan “*hasta donde mis datos son normales*”, de modo que cualquier dato o caso que no se encuentre dentro de este rango es marcado e identificado individualmente (*Dato anómalo*).

El *proceso para su construcción* es el siguiente:

1. Dibujar un rectángulo cuyos lados verticales pasen por el *primero y tercer cuartil*.
2. Dibujar otra vertical a la altura de la *mediana*.
3. Dibujar un aspa en mitad del rectángulo, a la altura de la *media*.
4. De cualquiera de las aristas del rectángulo se extiende una línea, o *bigote, que va hacia los valores extremos* (valor mínimo y valor máximo). Estas son observaciones que se encuentran entre cero y 1.5 veces el rango intercuartílico a partir de las aristas del rectángulo.
5. Las observaciones que están entre 1.5 y 3 veces el rango intercuartílico a partir de las aristas del rectángulo reciben el nombre de *valores atípicos* y se representan mediante cuadrados separados.

Ejemplo 1:

Los siguientes datos recogen los volúmenes de facturación de una muestra de 10 empresas de un determinado sector:

3,2 2,6 2,1 2,4 6,2 2,8 2,2 4,7 2,0 3,8

- Dibujar un diagrama Box-Whisker de los datos.
- A la vista del diagrama ¿qué signos cabe esperar que tengan el coeficiente de asimetría y el coeficiente de curtosis de los datos?

Pasos para su resolución:

1. Ordenar los datos de menor a mayor.

2 2,1 2,2 2,4 2,6 2,8 3,2 3,8 4,7 6,2

2. Obtener la mediana (C_2) y cuartiles (C_1 y C_3) y la media.

$C_1 = 2,2$; $C_2 = 2,7$; $C_3 = 3,8$; $\bar{X} = 3,2$

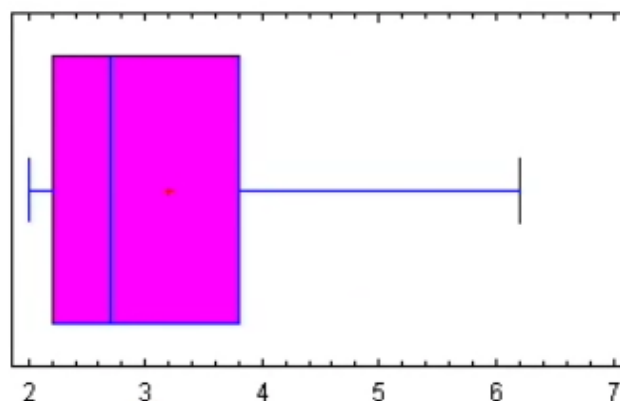
3. Obtener el Intervalo Intercuartílico (II).

$II = C_3 - C_1 = 3,8 - 2,2 = 1,6$

4. Obtener los valores de los límites superior (LS) e inferior (LI), de hasta donde podría prolongar mis bigotes si mis datos fuesen normales:

$LS = 3,8 + (1,5 * II) = 6,2$

$LI = 2,2 - (1,5 * II) = -0,2$



Ejemplo 2:

Los siguientes datos recogen el número de vehículos vendidos en 11 concesionarios distintos de Valencia en el último trimestre :

25 110 42 10 8 180 70 14 56 17 30

Dibujar el diagrama Box-Whisker de los datos.

Pasos para su resolución: Procedemos como en el caso anterior, ordenado las datos de menor a mayor, calculando cuartiles y media, estimando el II y los límites LS y LI.

25 **110** 42 10 **8** 180 70 14 56 17 30

$$C_1 = 14; C_2 = 30; C_3 = 70; \bar{X} = 56,2$$

$$II = C_3 - C_1 = 70 - 14 = 56$$

$$LS = 70 + (1,5 * II) = 154$$

$$LI = 30 - (1,5 * II) = -54$$

Box-and-Whisker Plot

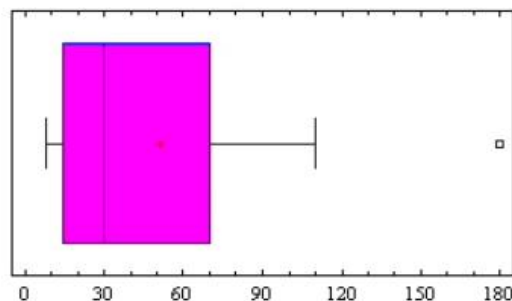


Imagen 3. Diagrama Box-Whisker

Fuente: Elaboración Propia

Ejemplo 3:

Una empresa fabricante de componentes mecánicos tiene dos líneas de fabricación (L1 y L2). Se muestra el diagrama Box-Whisker codificado según la línea en que se fabricó cada pieza. Teniendo cuenta que las especificaciones de fabricación dicen que la longitud de las piezas debe estar en el intervalo 100 ± 1 cms, indicar cuál sería la línea más adecuada para fabricar las piezas, justificando la respuesta.

Los diagramas de caja son muy útiles al hacer comparaciones gráficas entre conjuntos de datos.

Para ello debemos fijarnos en tres cosas básicas:

"Normalidad, Amplitud de caja y presencia de datos anómalos".

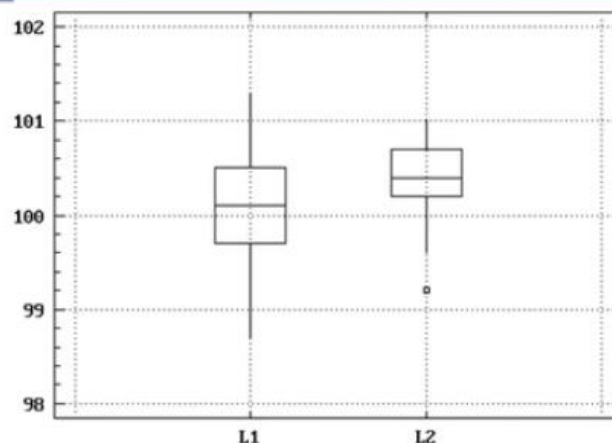


Imagen 4. Diagrama Box-Whisker

Fuente: Romero, R y Zúnica, L.R. (1993)

La línea L1 fabrica mayor proporción de piezas con dimensión cercana al nominal, pero produce unas pocas piezas fuera de especificación. La línea L2 produce las piezas con menor dispersión, todas dentro de la especificación, pero con una dimensión sistemáticamente mayor que el nominal. Convendría usar la línea L2 pero intentar centrar el proceso de fabricación alrededor de la media deseada.