



LICEO MARTA DONOSO ESPEJO

Diagrama de Caja y Bigotes

Los diagramas de Caja-Bigotes (boxplots o box and whiskers) son una presentación visual que describe varias características importantes, al mismo tiempo, tales como la dispersión y simetría.

Para su realización se representan los tres cuartiles y los valores mínimo y máximo de los datos, sobre un rectángulo, alineado horizontal o verticalmente.

Una gráfica de este tipo consiste en una caja rectangular, donde los lados más largos muestran el recorrido intercuartílico. Este rectángulo está dividido por un segmento vertical que indica donde se posiciona la mediana y por lo tanto su relación con los cuartiles primero y tercero (recordemos que el segundo cuartil coincide con la mediana).

Esta caja se ubica a escala sobre un segmento que tiene como extremos los valores mínimo y máximo de la variable. Las líneas que sobresalen de la caja se llaman bigotes. Estos bigotes tienen un límite de prolongación, de modo que cualquier dato o caso que no se encuentre dentro de este rango es marcado e identificado individualmente

EJEMPLO DISTRIBUCIÓN DE EDADES

Utilizamos la ya usada distribución de frecuencias (en tallos y hojas), que representan la edad de un colectivo de 20 personas.

36 25 37 24 39 20 36 45 31 31

39 24 29 23 41 40 33 24 34 40

ORDENAR LOS DATOS

Para calcular los parámetros estadístico, lo primero es ordenar la distribución

20 23 24 24 24 25 29 31 31 33 34 36 36 37 39 39 40 40 41 45

CALCULO DE CUARTILES

Q1, el cuartil Primero es el valor mayor que el 25% de los valores de la distribución. Como $N = 20$ resulta que $N/4 = 5$; el primer cuartil es la media aritmética de dicho valor y el siguiente:

$$Q1 = (24 + 25) / 2 = 24,5$$

Q2, el Segundo Cuartil es, evidentemente, la mediana de la distribución, es el valor de la variable que ocupa el lugar central en un conjunto de datos ordenados. Como $N/2 = 10$; la mediana es la media aritmética de dicho valor y el siguiente:

$$me = Q2 = (33 + 34) / 2 = 33,5$$

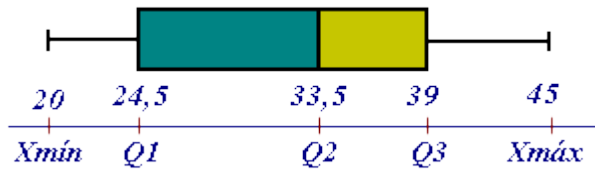
Q3, el Tercer Cuartil, es el valor que sobrepasa al 75% de los valores de la distribución. En nuestro caso, como $3N / 4 = 15$, resulta



LICEO MARTA DONOSO ESPEJO

$$Q2 = (39 + 39) / 2 = 39$$

DIBUJAR LA CAJA Y LOS BIGOTES



El *bigote* de la izquierda representa al colectivo de edades (X_{\min} , Q_1)

La primera parte de la caja a (Q_1 , Q_2),

La segunda parte de la caja a (Q_2 , Q_3)

El *bigote* de la derecha viene dado por (Q_3 , X_{\max}).

INFORMACIÓN DEL DIAGRAMA

Podemos obtener abundante información de una distribución a partir de estas representaciones.

Veamos alguna:

La parte izquierda de la caja es mayor que la de la derecha; ello quiere decir que las edades comprendidas entre el 25% y el 50% de la población está más dispersa que entre el 50% y el 75%.

El *bigote* de la izquierda (X_{\min} , Q_1) es más corto que el de la derecha; por ello el 25% de los más jóvenes están más concentrados que el 25% de los mayores.

El *rango intercuartílico* = $Q_3 - Q_1 = 14,5$; es decir, el 50% de la población está comprendido en 14,5 años.

Comparar distribuciones

La mayor utilidad de los diagramas caja-bigotes es para comparar dos o más conjuntos de datos.

Comparación distribución de edades

Comparación entrenamientos de un corredor

Comparación clasificación liga

COMPARACIÓN DISTRIBUCIÓN DE EDADES

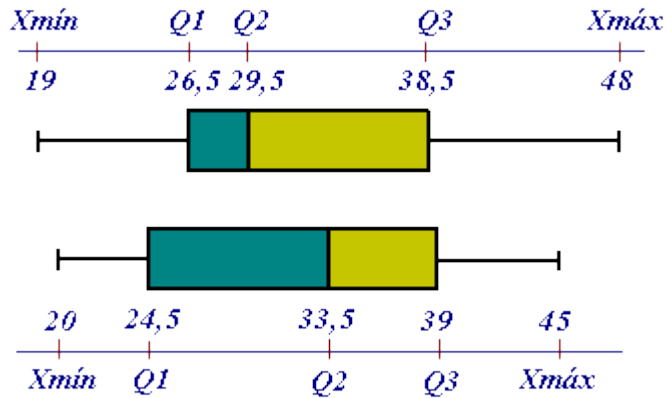
Análogamente a lo realizado con los diagramas de tallo y hojas, comparamos, mediante estos diagramas, esta distribución con la del otro ejemplo de distribución de edades.

35 38 32 28 30 29 27 19 48 40

39 24 24 34 26 41 29 48 28 22



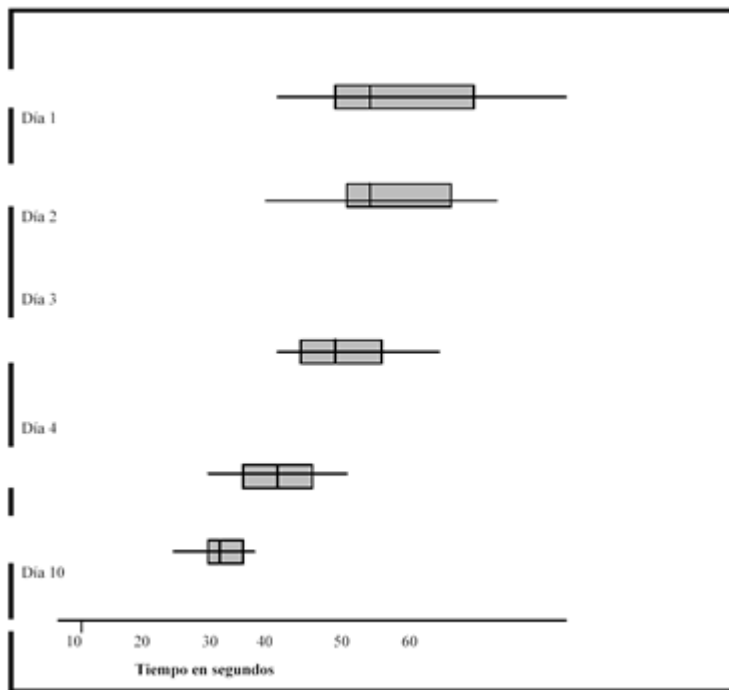
LICEO MARTA DONOSO ESPEJO



A partir de dicha comparación puede obtenerse bastante información de ambas distribuciones.

COMPARACIÓN ENTRENAMIENTOS DE UN CORREDOR

Un corredor entrena para una determinada carrera y se toman los tiempos que necesita para recorrer los 100m, durante 10 días consecutivos (cada día se toman varios tiempos y se calculan mediana, cuartiles, valores mínimo y máximo)



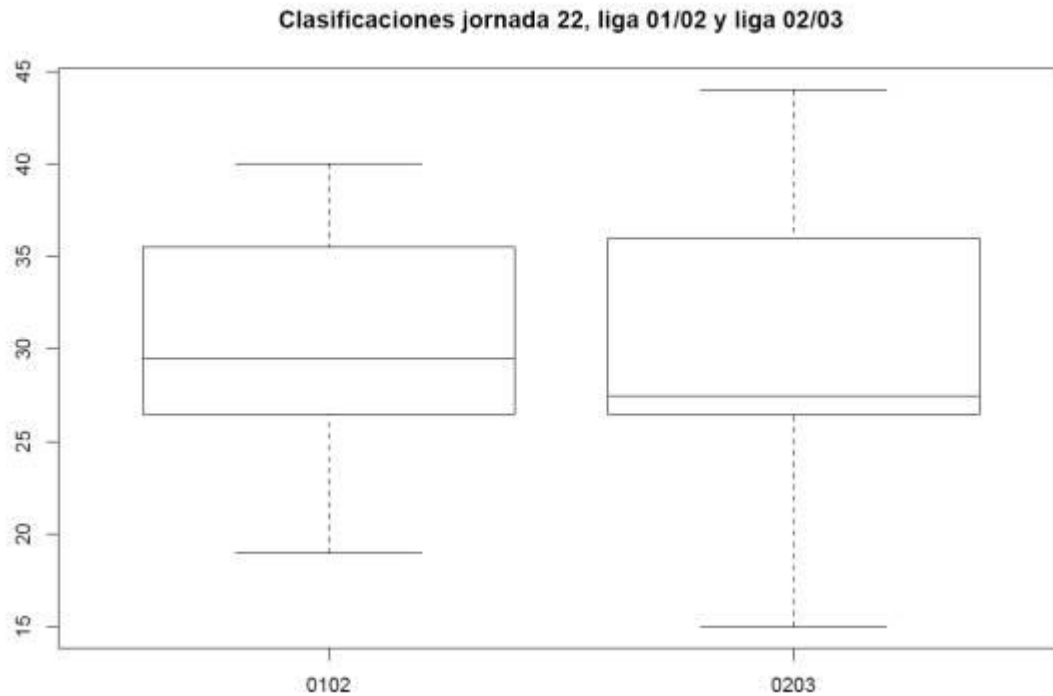
Observamos que el desplazamiento de las gráficas de caja hacia la izquierda indica que el entrenamiento ha dado resultado, ya que se tardan menos segundos en recorrer la misma distancia, siendo la diferencia entre el máximo y el mínimo menor, como así también la diferencia intercuartílica.



LICEO MARTA DONOSO ESPEJO

COMPARACIÓN CLASIFICACIÓN LIGA

Las puntuaciones de los equipos de la liga de la temporada 01/02 y 02/03 en primera división se pueden comparar con un diagrama caja y bigotes, como aparece aquí,



Comentarios: No hay datos muy atípicos, es decir que no hay equipo que se haya destacado por arriba o por abajo del resto de los equipos. Hay más diferencia de puntos entre el primer y el último clasificado para la liga 02/03 que en la liga anterior. Los equipos del tercer cuarto de la clasificación están más apelotonados en la liga 02/03.