

Using box plots to describe and compare distributions

Sunday, 10 February 2019 4:36 pm

★ By the end of the lesson I would hope that you have an understanding and be able to apply to questions the following concepts:

- Know how to read what a box-plot is trying to tell us about its:
 - Shape
 - Centre
 - IQR
 - Any outliers which might exist
- Write statements and short reports about distributions described by box-plots
- Compare data using box-plots

RECAP:

This series of lessons is looking at how we can describe data.

We have, to this point looked at:

- Dot plots and stem plots
- The mean, range and interquartile range
- The five number summary and box plot
- Knowing how to relate a box plot to shape

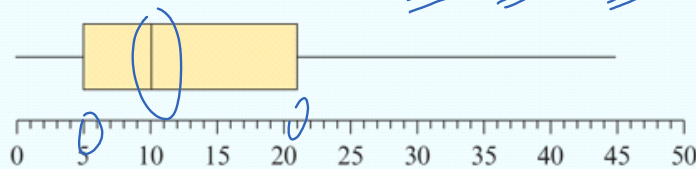
It's time to write some reports (or statements) about box plots.

The best way to do this is to use examples!

Example 1: Using a box plot to describe a distribution with outliers

Example has been extracted from the Cambridge Further Mathematics Units 3 and 4 Textbook

Describe the distribution represented by the box plot in terms of shape centre and spread. Give appropriate values.



We can describe the data in terms of:

- Shape
- Outliers
- Centre (Median value)
- Spread of data (Range and IQR)

$$\begin{aligned} \text{UQ} - \text{LQ} &= 16 \\ 21 - 5 &= 16 \end{aligned}$$

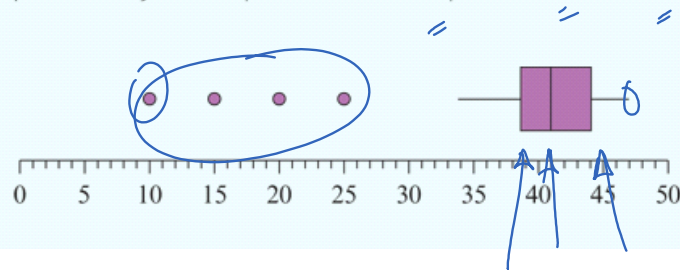
One possible description might be:

The distribution is positively skewed with no outliers. The distribution is centred at 10, the median value. The spread of the distribution, as measured by the IQR, is 16 and, as measured by the range, 45.

Example 2: Using a box plot to describe a distribution with outliers

Example has been extracted from the Cambridge Further Mathematics Units 3 and 4 Textbook

Describe the distributions represented by the box plot in terms of shape and outliers, centre and spread. Give appropriate values.



We can describe the data in terms of:

- Shape
- Outliers
- Centre (Median value)
- Spread of data (Range and IQR)
- The number and values of outliers

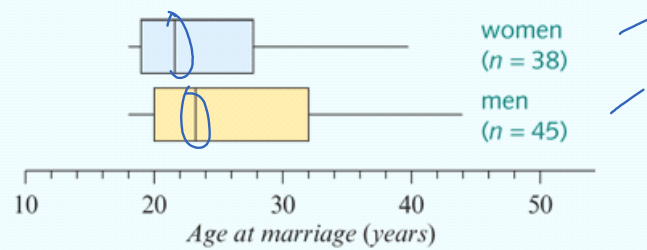
One possible description might be:

The distribution is symmetric but with outliers. The distribution is centred at 41, the median value. The spread of the distribution, as measured by the IQR, is 5.5 and, as measured by the range, 37. There are four outliers: 10, 15, 20 and 25.

Example 3: Using a box plot to compare distributions

Example has been extracted from the Cambridge Further Mathematics Units 3 and 4 Textbook

The parallel box plots show the distribution of ages of 45 men and 38 women when first married.



- Compare the two distributions in terms of shape (including outliers, if any), centre and spread. Give appropriate values at a level of accuracy that can be read from the plot.
- Comment on how the age of the men when first married compares to that of women.

The distributions of age at marriage are positively skewed for both men and women. There are no outliers. The median age at marriage is higher for men ($M=23$ years) than women ($M=21$ years). The IQR is also greater for men ($IQR=12$ years) than women ($IQR=9$ years). The range of age at marriage is also greater for men ($R=26$ years) than women ($R=22$ years).

For this group of men and women, the men, on average, married later and the age at which they first married was more variable.