Est. 1841 YORK ST JOHN UNIVERSITY
Student Life
Library and Learning Services

Interquartile range and box plots

Study Development Factsheet

Interquartile range is similar to range, but is for data sets with extreme outliers. For example, in a survey about number of study hours that students do in a week, an answer of 0 hours might be an outlier, and a value of 100 hours would definitely be an outlier. Calculating range from these values would give an inaccurate picture of the spread of the data. In this case, it would be more appropriate to calculate the interquartile range.

Method

There are 5 ‘quartile’ values we can calculate, and the method is very similar to calculating a median (in fact, one of the quartiles *is* the median).

1. Place all the values in order from smallest to largest.
2. The lowest value is called Q_0 (quartile 0). This is often not really needed, but can be found if useful.
3. The highest value is called Q_4 (4^{th} quartile). The range can be calculated by finding $Q_4 - Q_0$.
4. Find the median. For n of data points, find the value in position $\frac{n}{2} + 0.5$ (for example, for 5 data points, find the 3^{rd} data point. This is the median). If n is even, we find the values in positions $\frac{n}{2}$ and $\frac{n}{2} + 1$ and find the point between them by adding them together and dividing by 2 (for example, for 10 data points, add together the 5^{th} and 6^{th} values and divide that by 2).

   The median is called Q_2 (2^{nd} quartile).
5. We are now left with two sets of numbers: those above the median, and those below. We find the median of each of these sets (make sure not to put Q_2 in either set before finding the median).
6. The median of the numbers lower than Q_2 is called Q_1 (1^{st} quartile) and the median of the numbers higher than Q_2 is called Q_3 (3^{rd} quartile).
7. The interquartile range (IQR) is given by $Q_3 - Q_1$.

Library and Learning Services
Study Development
Email: studydevelopment@yorksj.ac.uk

Example

A bookshop records how many books each customer bought one day. They get the following results:

2, 1, 1, 3, 95, 1, 4, 5, 1, 0, 1, 0, 1, 0, 0, 0

Find the interquartile range and justify your reason for finding the IQR instead of the range.

Answer

1. We put the values in order from smallest to largest:
   0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 2, 3, 4, 5, 95
2. $Q_0 = 0$
3. $Q_4 = 95$
4. There are 16 values, so we find the values in position 8 and 9 and average them:
   $8^{th} = 1$, $9^{th} = 1$, average = 1.
   $Q_2 = 1$
5. We now have two sets of numbers: those lower than Q_2: 0, 0, 0, 0, 0, 1, 1, 1, and those higher than Q_2: 1, 1, 1, 2, 3, 4, 5, 95.
6. The median of those lower than Q_2 is found by averaging those in the 4^{th} and 5^{th} positions:
   $4^{th} = 0$, $5^{th} = 0$, average = 0.
   $Q_1 = 0$
7. The median of those higher than Q_2 is found by averaging those in the 12^{th} and 13^{th} positions: $12^{th} = 2$, $13^{th} = 3$, average = 2.5.
   $Q_3 = 2.5$
8. The IQR is $Q_3 - Q_1 = 2.5 - 0 = 2.5$

We find the IQR rather than the range since 95 books is obviously an outlier for the shop (perhaps a large order was placed for an event- but it seems strange next to all the other values).

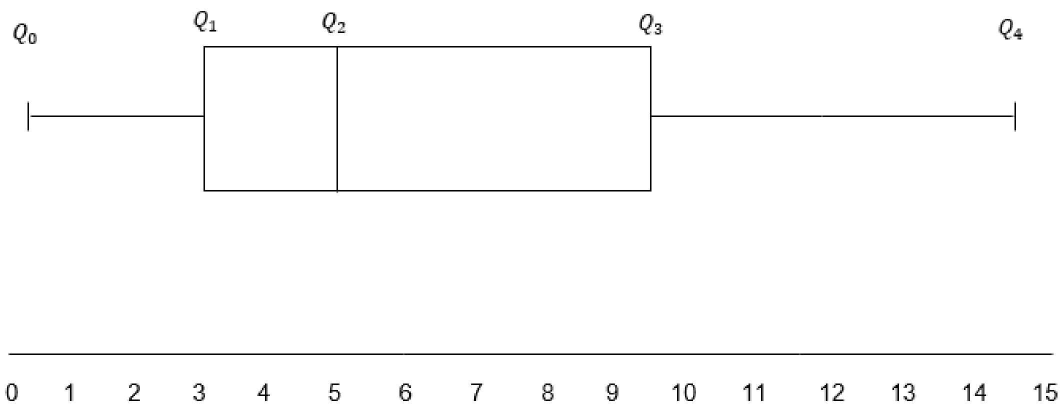
Interquartile range and box plots

Study Development Factsheet

Drawing a box plot

We can use the values for Q_0, Q_1, Q_2, Q_3 and Q_4 to draw a box plot (also called a box and whisker plot). These are useful for quickly seeing if data is skewed.

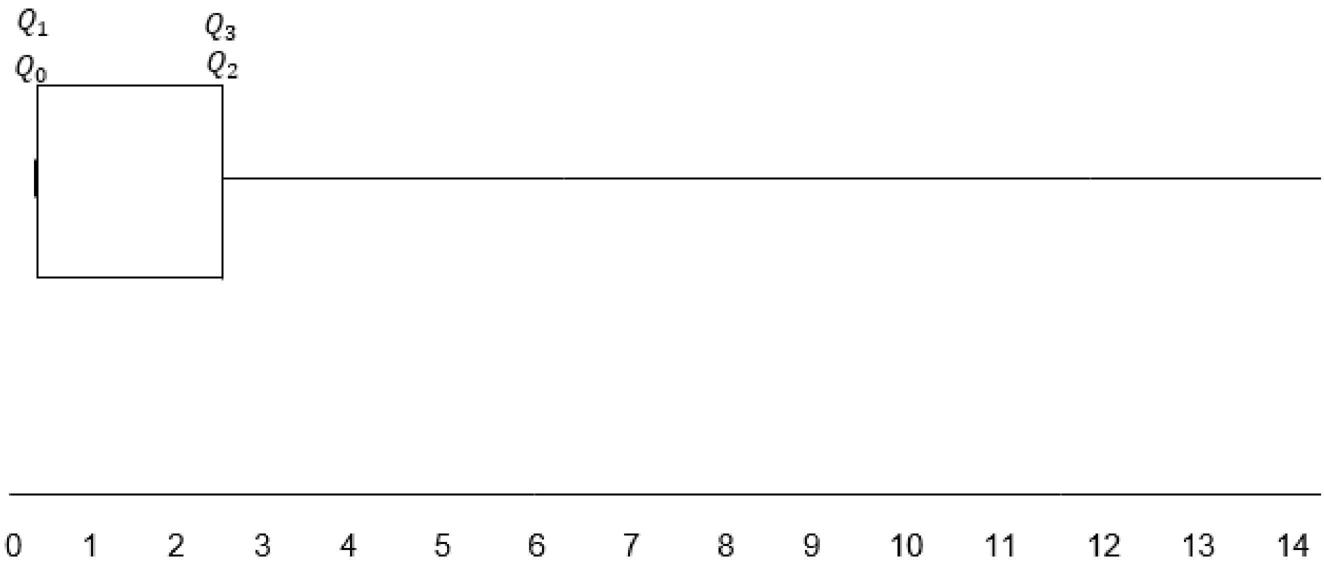
A box plot is drawn as follows:



The lines that show Q_0 to Q_1 and Q_3 to Q_4 are called the 'whiskers' of the box and whisker plot, and the box between Q_1 and Q_3 is the 'box'.

We can tell the skew by looking at Q_2, if it is closer to Q_1 than Q_3, this is called 'left skew' or 'positive skew'. If Q_2 is closer to Q_3 than Q_1, this is called 'right skew' or 'negative skew'.

Sometimes these can look quite strange. For example, the box plot for the earlier bookshop example looks like this:



With Q_4 so far away that it isn't actually shown on this diagram. This is still a box and whisker plot, and shows very strong negative skew.

Support: Study Development offers workshops, short courses, 1 to 1 and small group tutorials.

- Join a tutorial or workshop on the Study Development tutorial and workshop webpage or search 'YSJ study development tutorials.'
- Access our Study Success resources on the Study Success webpage or search 'YSJ study success.'