

Section 2.5, Measures of Position

1 The Five-Number Summary

The three **quartiles** Q_1 , Q_2 , and Q_3 divide the data into four “equal” parts. About 25% of the data falls at or below Q_1 , 50% of the data falls at or below Q_2 , and 75% falls at or below Q_3 . To find the quartiles:

1. First, order the data.
2. Q_2 is the median of the data. If the number of data points is even, keep the original data set. If the number of data points is odd, you will ignore the data entry used for the median from now on.
3. Q_1 is the median of the first half of the data.
4. Q_3 is the median of the second half of the data.

The **interquartile range (IQR)** is the range of the middle half of the data, so:

$$\text{IQR} = Q_3 - Q_1$$

Example

Find the quartiles and IQR of the following data:

1. 100 102 102 102 102 106 108

The median is 102, so $Q_2 = 102$. Since there were 7 data points (an odd number), we will ignore only the 102 used for the median. Eliminating this, we consider the first 3 data points, so $Q_1 = 102$. Considering the last 3 data points, we get that $Q_3 = 106$. Then, $\text{IQR} = 106 - 102 = 4$.

2. 7 10 11 12 13 14 16 17

The median is 12.5, so $Q_2 = 12.5$. Since there were 8 data points (an even number), we will not ignore any points when calculating the other quartiles. We consider the first 4 data points, so $Q_1 = 10.5$. Considering the last 4 data points, we get that $Q_3 = 15$. Also, $\text{IQR} = 15 - 10.5 = 4.5$

The **five-number summary** consists of the minimum, Q_1 , the median Q_2 , Q_3 , and the maximum of a data set.

A **box-and-whisker plot (boxplot)** is a graphical representation of the five-number summary. A horizontal line spans from the minimum entry to the maximum entry, with a box spanning from Q_1 to Q_3 . The median is represented by a vertical line the height of the box. In general, it looks like:

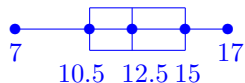


The lines that extend past the box are called “whiskers.” The rectangular part is referred to as a “box.”

Example

Find the five-number summary and sketch the box-and-whisker plot for: 7 10 11 12 13 14 16 17

We already calculated $Q_1 = 10.5$, $Q_2 = 12.5$, and $Q_3 = 15$. From the data, we see that the minimum is 7 and the maximum is 17, so the box-and-whisker plot will be:



Box-and-whisker plots can be drawn over a number line, or be “stacked” vertically to compare data from multiple data sets.

2 Standardized Data

The **standard score** or **z -score** is the number of standard deviations that a given value x falls from the population mean μ :

$$z = \frac{x - \mu}{\sigma},$$

where σ is the population standard deviation.

One benefit of a z -score is that it allows us to compare data from different data sets. For example, it’s used to determine what we consider “tall” for different genders- a male who is six feet tall isn’t normally considered “very tall,” but a female who is the same height normally is “very tall” since her height is further above the mean height for females.

Example

For healthy females, the red blood cell count (RBC, in millions per cubic millimeter of whole blood) has mean 4.8 and standard deviation 0.3. Is a RBC of 4.6 considered unusual?

First, we calculate the z -score: $z = \frac{4.6-4.8}{0.3} = -0.6667$. This measurement is within one standard deviation of the mean, so it is not unusual (by the Empirical rule).