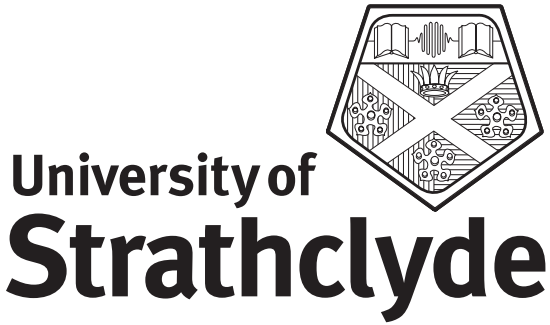


MM104/ MM106/ BM110

Topic 2: Variability, Skewness and Percentilies

Skewness

Ainsley Miller
ainsley.miller@strath.ac.uk

So many summary statistics

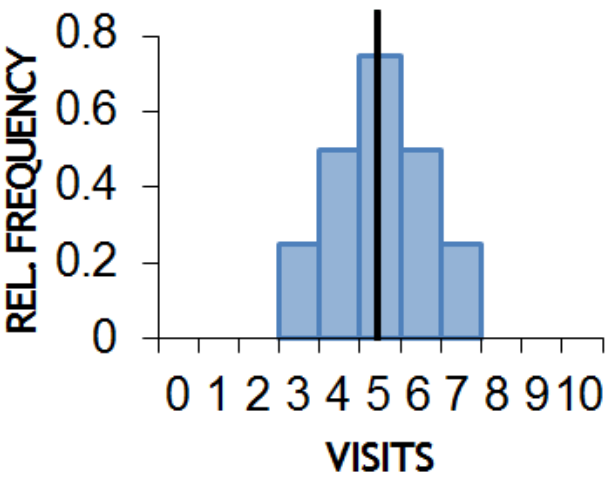
We have learned about a lot of summary statistics - mean, median, mode, range, upper and lower quartiles, IQR, standard deviation and variance.

In a statistical analysis is not appropriate to quote them all, you need to be able to decide when to quote which ones.

This is determined by the **skewness** of the data.

Skewness

Ideally you do not want your data to be skewed, you want it to be **symmetric**.
The following histogram shows how many times a sample of students visited the library during the exam diet.



If the data is symmetric we quote the mean and standard deviation.

If the data is not symmetric we quote the median and the interquartile range.

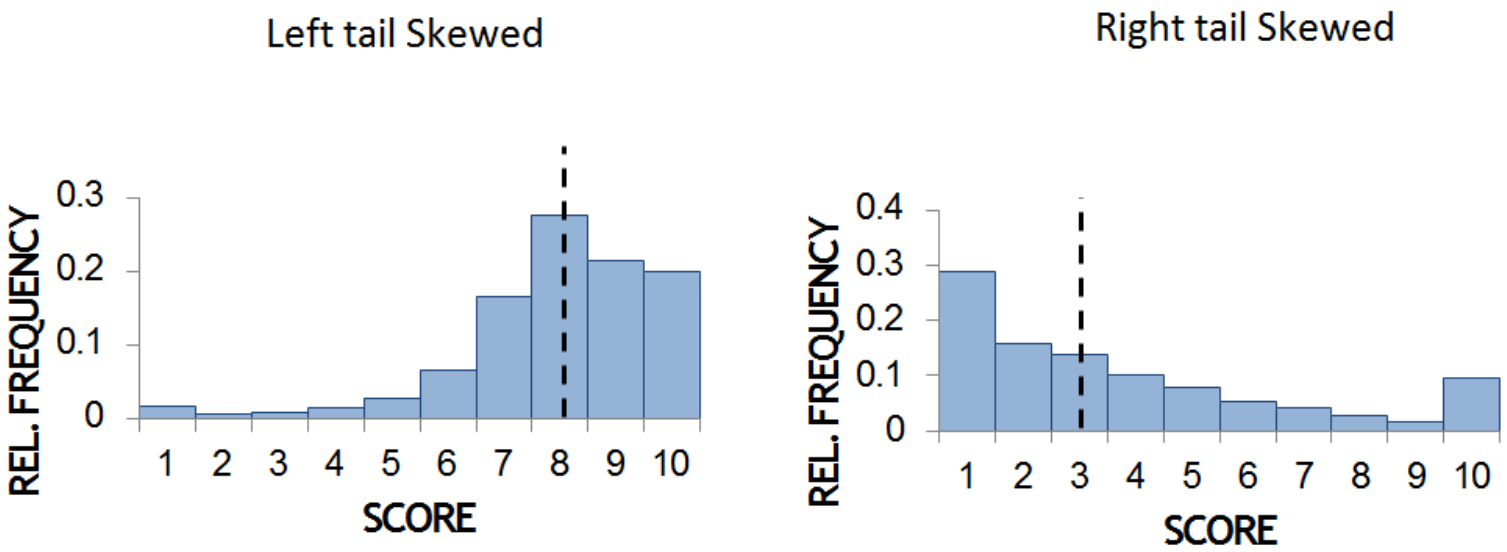
But looking at a plot by eye is very subjective so we need a measurement.

Skewness Statistic

The skewness statistic measures how skewed a data set is. The formula is too complicated to calculate by hand, so we use Minitab to calculate it. The steps are very similar to calculating the mean, median and mode. Navigate to Store Descriptive Statistics and tick the box for Skewness.

- If $-0.5 \leq$ Skewness Statistic ≤ 0.5 the data is roughly symmetric.
- If the Skewness Statistic > 0.5 the data is right tail skewed.
- If the Skewness Statistic < -0.5 the data is left tail skewed.

What do we mean by right and left skewed.



From the Figure above the plot on the left shows that the tails of the distribution are at the left side. Whereas, the plot on the right shows that the tails of the distribution are at the right side.

Relating this back to the mean and median.

Generally speaking, if the data is symmetric then the mean and median will be roughly the same.

Generally speaking, if the data is left tail skewed then the mean will be less than the median.

Generally speaking, if the data is right tail skewed then mean will be greater than the median.