The Median and the Interquartile range (IQR)

Today, the statistical analysis of a laboratory proficiency testing (PT) program for checking the performance of test laboratories through inter-comparison of test results on standardized samples tends to go for:

(1) Non-parametric or distribution-free methods which handle data that may not be normally distributed.

(2) Robust methods which are based on the belief that the underlying population distribution may indeed be approximately normal, but because of the presence of outliers, this distribution may get distorted.

In essence, robust techniques operate by down-weighting the importance of outliers in the dataset, so they are appropriate in the cases of heavy-tailed distributions, and their acceptance and use have increase dramatically in recent years.

In here, we do not use the arithmetic mean or average as the measure of central tendency of a set of results because we have to assume symmetrical normal distribution.   Instead, we look at the **median** of the dataset.

To calculate the median of n observations, we arrange them in ascending order: in the unlikely event that n is very large, this sorting process can be performed very quickly by common software programs such as Excel® spreadsheets, SPSS or Minitab®.

The median is the value of the $\frac{1}{2}(n+1)$th observation if n is odd, and the average of the $\frac{1}{2}n$th and the $\frac{1}{2}(n+1)$th observations if n is even.

For example, if we have n = 5 data such as 1, 2, **2**, 3, 5, the median is 2, whilst if we have n = 6 data such as 1, 2, 2, 3, 5, 6, the median is (2+3)/2 or 2.5.

We see that the median of a set of experimental results is determined with little or no calculation, unaffected by outlying values, and is a more realistic measure of central tendency than the arithmetic mean.

In non-parametric statistics, the usual measure of dispersion (replacing the standard deviation) is the **interquartile range** (IQR). As we have seen, the median divides the sample of measurements into 2 equal halves (50% each);

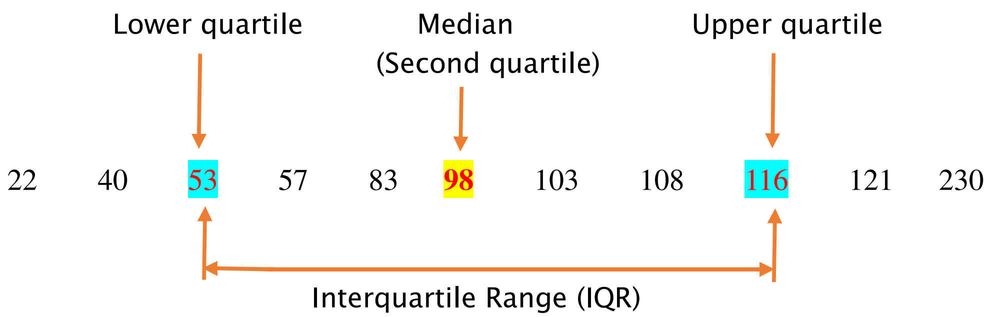
if each of these halves is further divided into two to become 25% each, the points of division are called the **upper** and **lower quartiles**. The range between the upper and lower quartiles is called **interquartile range** (IQR).

The following illustration demonstrates clearly the concepts of median and interquartile range:

Given a set of 11 data as:

22	57	40	83	103	53	98	230	121	108	116

Upon re-arranging the data in ascending order, we have:



The advantage of the IQR, which shows the spread of the observations, is that it is not affected by extreme values at either of the data distribution.