


# Percentiles and Quantiles

The mean and standard deviation can be used to summarize a data set. Frequency tables and histograms give more details about the distribution of the data. Percentiles (including quantiles) are another tool for summarizing the data distribution. The median is an example of a percentile. It is the 50th percentile. Half the data is less than the median and half is greater.

**Example**. Here is a small data set:

6 2 17 5 5 3 20 19 10 12 19

To compute the median, we first need to sort the data.

2 3 5 5 6 10 12 17 19 19 20

Here are a few examples of percentiles:

- The minimum is the 0% since none of the data is smaller. The 0% is 2.
- The median is 10.
- The maximum is the 100% since none of the data is greater. The 100% is 20.
- The 25% is a value $v$ for which 25% of the data is at most $v$. It is the median of the first half of the data: 2, 3, 5, 5, 6, 10. So, the 25% is 5.
- The 75% is the median of the second half of the data: 10, 12, 17, 19, 19, 20. So, the 75% is any value between 17 and 19. Most statistical software will take the average and output 18.

Here is how to do this in `R`.

```
> x = c(6, 2, 17, 5, 5, 3, 20, 19, 10, 12, 19)
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   2.00    5.00   10.00   10.73   18.00   20.00
> fivenum(x)
[1]  2  5 10 18 20
> range(x)
[1]  2 20
```

1. Explain in words what the `summary`, `fivenum` and `range` commands in `R` do.

2. For the following data, find the 0%, 25%, 50%, 75% and 100% by hand then check your answers with `R`.

6 10 20 19 1 1 11 13 13 17 2

.

3. Repeat #2 using the following data.

98 83 65 56 58 13 97 4 100 63

.