3.5 Measures of Spread

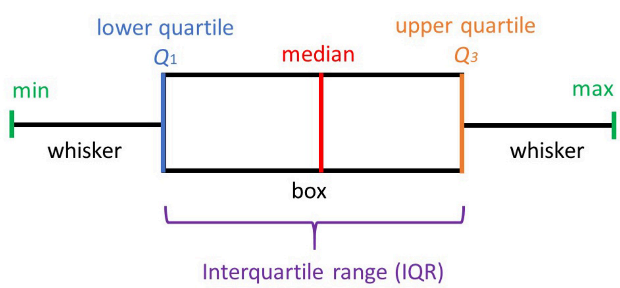
A measure of spread is how we discuss how data is clustered.

A mean of 75 tells us where the data is centered but a standard deviation of 10 tells us how tightly the data is grouped.

Interquartile Range - a measure of spread about the median. IQR = $Q_3 - Q_1$

A quartile is breaking the data into 4 groups of equal number-like taking the median on either side of the median. Q_1 is called the 'lower quartile' and is the median of the bottom half of the data. Q_3 is called the 'upper quartile' and is the median of the top half of the data.

A Box and Whisker Plot displays all this information in one graphic.



Standard Deviation - a measure of spread about the mean.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \mu)^2}{n}}$$

the f_i can be omitted when the data isn't grouped.

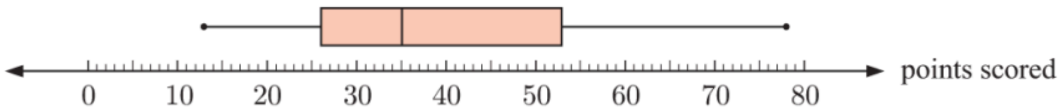
For now: the standard deviation gives a number value to the clustering of the data about the mean. Higher value, wider spread; lower value, tighter spread.

Variance is the square of the standard deviation, takes σ^2 as its variable and uses the formula above, without the square root, as its formula.

For the data set 5 5 7 3 8 2 3 4 6 5 7 6 4, find:

a the median **b** Q_1 and Q_3 **c** the interquartile range.

1 The box plot below summarises the points scored by a basketball team.



State the:

a median
b maximum value
c minimum value
d upper quartile
e lower quartile
f range
g interquartile range.

- **upper boundary = upper quartile + 1.5 × IQR**
  Any data larger than the upper boundary is an outlier.
- **lower boundary = lower quartile − 1.5 × IQR**
  Any data smaller than the lower boundary is an outlier.

The table alongside summarises the examination scores for 80 randomly selected students. Estimate the standard deviation for the data.

Mark	*Frequency*
0 - 9	1
10 - 19	1
20 - 29	2
30 - 39	4
40 - 49	11

Mark	*Frequency*
50 - 59	16
60 - 69	24
70 - 79	13
80 - 89	6
90 - 99	2

p.331#3,5,6
p.336#2,3
p348#7,10,13