MM104/ MM106/ BM110

Topic 2: Variability, Skewness and Percentiles
Common Measures of Variability

Ainsley Miller
ainsley.miller@strath.ac.uk

Spread (Variability)

We have already discussed location in the previous lecture. In today's lecture we are going to discuss **spread**.
The location (mean, median and mode) tell us what to expect from the data.

> The spread (variability) tells how much sample values depart from expectations and about the diversity of values within our sample.

This can be a good or bad thing

- Good: Job interview candidates as this maximises the number of possibilities for an employer.
- Bad: Manufacturing process where uniform production of goods is essential.

Measures of Variability

There are many ways to measure variability

- Range
- Upper and Lower Quartiles
- Interquartile Range
- Standard Deviation
- Variance

Range

The range is the most simple measure of variability and is simply the largest value in a data set minus the smallest value.

Example 1

Find the range of the following data: 4, 6, 9, 3 and 7.

The largest value is: 9

The smallest value is: 3

Range is $9 - 3$

$\underline{\underline{\text{The range is 6.}}}$

Although the range is simple and quick to calculate, it can be very misleading.

Quartiles

We will now introduce quartiles. Recall from the previous topic that the median is the middle value in your data, i.e. **it cuts the data in half**.

Therefore quartiles split the data into quarters, and there are three quartiles: $Q1, Q2$ and $Q3$.

- The first quartile is Q1 and it separates the first 25 % of the data from the rest.
- The second quartile is Q2 and is also the median, and it separates the first 50 % of the data from the rest.
- The third quartile is Q3 and it separates the first 75 % of the data from the rest.

First and Third Quartile

The first and third quartile are called the lower and upper quartiles respectively. Calculating the upper and lower quartile is very similar to calculating the median. Remember the data must be arranged in ascending numerical order.

$Q1$, the lower quartile, is located at the $\frac{n+1}{4}$ position in the data.

$Q3$, the upper quartile, is located at the $\frac{3(n+1)}{4}$ position in the data.

$Q1$ or $Q3$ = lower value + (remainder × (upper value - lower value))

What is the remainder ?

If we found that the lower quartile was located at position 3.5 then the remainder is what comes after the integer value i.e. 0.5

If we found that the upper quartile was located at position 10.75 then the remainder is what comes after the integer value i.e. 0.75

> The remainder is **always** a decimal, and $0 < \text{remainder} < 1$

Finding the upper and lower quartile

Example 2

Given the following data: 15, 1, 6, 2, 18, 8, 14, 6, 7 and 1. Find the upper and lower quartile.

Firstly put the data in ascending order:

$1, 1, 2, 6, 6, 7, 8, 14, 15, 18$

We'll find the lower quartile first

The position of the lower quartile is $\dfrac{n+1}{4}$, here $n = 10$.

$\Rightarrow$ Position: $\dfrac{10+1}{4} = \dfrac{11}{4} = 2.75$

The lower quartile is located 0.75 of the way between the 2nd and 3rd element.

Finding the upper and lower quartile

So what do we do ?

The second element is 1. The third element is 2.

$Q1$ = lower value + (remainder $\times$ (upper value - lower value))

The remainder is 0.75

$\Rightarrow Q1 = 1 + (0.75 \times (2 - 1))$

$\Rightarrow \underline{\underline{Q1 = 1.75}}$

Finding the upper and lower quartile

Now we'll find the upper quartile

The position of the upper quartile is $\dfrac{3(n+1)}{4}$, here $n = 10$.

$\Rightarrow$ Position: $\dfrac{3(10+1)}{4} = \dfrac{3 \times 11}{4} = 8.25$

The upper quartile is located 0.25 of the way between the 8th and 9th element.

The 8th element is 14. The 9th element is 15.

$Q3 =$ lower value + (remainder $\times$ (upper value - lower value))

The remainder is 0.25

$\Rightarrow Q3 = 14 + (0.25 \times (15 - 14))$

$\Rightarrow \underline{\underline{Q3 = 14.25}}$

Interquartile Range

The Interquartile Range (IQR) is $Q3 - Q1$.

Example 3

Given the following data: 15, 1, 6, 2, 18, 8, 14, 6, 7 and 1.
Find the interquartile range.

This is the same data as the previous example, therefore we know
$Q1 = 1.75$ and $Q3 = 14.25$
Therefore IQR= 14.25 -1.75
$\underline{\underline{\text{IQR} = 12.5}}$

Standard Deviation

The sample standard deviation is a measure of variability which uses all the data in its calculation. The sample standard deviation measures the variability about the mean value. The sample standard deviation is denoted by s.

$$s = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$$

where n is the sample size, $\bar{x}$ is the sample mean and $x_i, \ldots, x_n$ are the data points.

The sample variance, s^2 is the standard deviation squared.

Standard Deviation - Example

It is unlikely to calculate the standard deviation by hand, we usually use Minitab, but we will provide an example.

Example 4

Find the standard deviation and variance of the following data: 10, 7, 6, 8 and 5.

We firstly find the sample mean

$$\bar{x} = \frac{10 + 7 + 6 + 8 + 5}{5} \Rightarrow \bar{x} = 7.2$$

$$s = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

Standard Deviation - Example

Let's find $\sum(x_i - \bar{x})^2$ first

$$\sum(x_i - \bar{x})^2 = (10 - 7.2)^2 + (7 - 7.2)^2 + (6 - 7.2)^2 + (8 - 7.2)^2 + (5 - 7.2)^2$$

$$\Rightarrow \sum(x_i - \bar{x})^2 = 14.8$$

$$s = \sqrt{\frac{1}{5-1}14.8}$$

$$s = 1.9235$$

$$s^2 = 1.9235^2 = 3.6999$$

The sample standard deviation is 1.9235 and the sample variance is 3.6999.