MM104/ MM106/ BM110

Topic 1: Data types, Central Location and Missing Values

**Central Location**

Ainsley Miller
ainsley.miller@strath.ac.uk

# Central Location

A measure of central location is a summary statistic that represents the centre point or typical value of a dataset.

**The location tells us what to expect from our data.**

In statistics, the three most common measures of central location are the

- Mode
- Mean
- Median

# Central Location

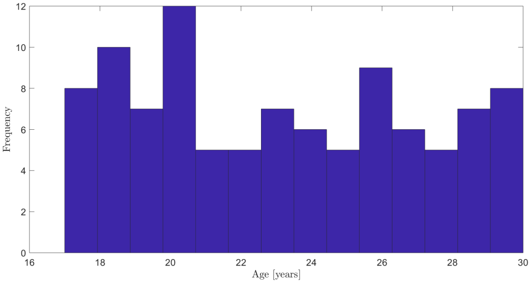
The mean, median and mode are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others. In the following slides, we will look at the mean, mode and median, and learn how to calculate them and under what conditions they are most appropriate to be used.

# Mode

The mode is arguably the simplest measure of central location to calculation.

| |
|---|
| The mode is the most common value occurring in the data set. |

# Mode Example




This histogram shows a sample of 100 customers who were surveyed in Topshop Oxford St. One of the survey questions was "How old are you ?".

From the histogram, the modal age of customers in Topshop Oxford St is 20 years.

# Mean

The mean is the most well known measure of central location.
The mean is equal to the sum of all the values in the data set divided by the number of values in the data set. So, if we have $n$ values in a data set and they have values $x_1$, $x_2$, …, $x_n$ the sample mean,denoted by $\bar{x}$ (pronounced $x$ bar) is:

$$\boxed{\bar{x} = \frac{x_1 + x_2 + \ldots x_n}{n}}$$

# Sample Mean

You may have noticed that the previous formula refers to the **sample mean**. So, why have we called it a sample mean? This is because, in statistics, **samples** and **populations** have very different meanings and these differences are very important, even if, in the case of the mean, they are calculated in the same way.

$\bar{x}$ is the sample mean.
$\mu$ is the population mean.

$\mu$ is the Greek letter mu.

# Samples and Population

The main difference between a population and sample has to do with how observations are assigned to the data set.

A population includes all of the elements from a set of data.

A sample consists one or more observations drawn from the population.

# Calculating the Mean - Example

The weights of 10 dogs in a veterinary practise were recorded in kg.
The weights were: 55, 18, 58, 68, 69, 33, 44, 55, 46, 51.
Find the mean weight of a dog in the veterinary practise.

We are calculating the sample mean as this data set is not representative of all dogs in the practise

$$\bar{x} = \frac{55 + 18 + 58 + 68 + 69 + 33 + 44 + 55 + 46 + 51}{10} = \frac{497}{10}$$

$$\Rightarrow \underline{\underline{\bar{x} = 49.7 \text{ kg}}}$$

# Median

The median is the middle score for a set of data that has been arranged in order of magnitude.

How to calculate the median

- Arrange the data in ascending numerical order
- The median is the $\frac{n+1}{2}$ element of the data set, where $n$ is the number of observations
- The median is equal to lower value $+\frac{\text{(upper value - lower value)}}{2}$

If the sample size is odd the median will be an exact value from the data set.
If the sample size is even the median will lie half way between two values.

# Calculating the Median - Example

We will use the same example which we used for the mean.

The weights of 10 dogs in a veterinary practise were recorded in kg.
The weights were: 55, 18, 58, 68, 69, 33, 44, 55, 46, 51.
Find the median weight of a dog in the veterinary practise.

Firstly order the data in ascending numerical order

$18, 33, 44, 46, 51, 55, 55, 58, 68, 69$

Now find the position of the median, $n$=10

Position of median: $\frac{n+1}{2} = \frac{10+1}{2} = 5.5$

The median will lie half way between the 5th and 6th element.

# Calculating the Median - Example

The 5th element is 51

The 6th element is 55

The median is: $51 + \dfrac{55-51}{2} = 51 + \dfrac{4}{2} = 51 + 2$

The median weight is 53 kg.