Chapter 2 - Measures of Location and Spread

Personal notes:

Prelude

Variables (x) in *algebra* vs *statistics*

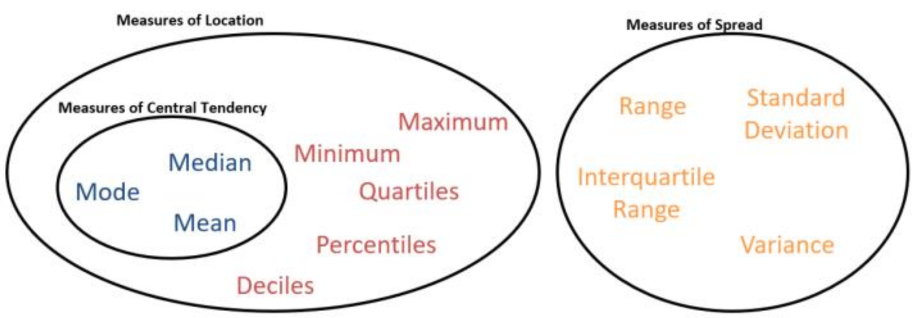
Similarities

- Variables in stats **represent the value of some quantity**, e.g. shoe size, height, colour
- Variables can be discrete or continuous
- **Multiplication reasoning**, e.g. if x represents the weight of a sample, then $2x$ represents twice sample's weight. In stats this is known as "**coding**" (see Chp 2.5)

Differences

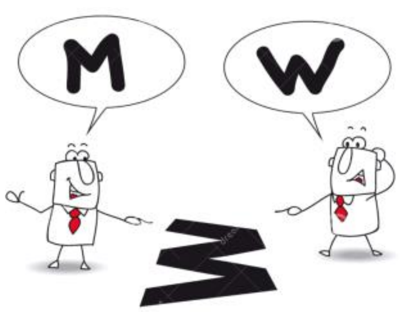
- Unlike algebra, a variable in stats represents the value of **multiple objects**, e.g. the weights of **all** people in a room, whereas in algebra, the value of x represents value of **an unit**, e.g. price of orange.
- Because of this, we can do operations on it as if it was a collection of values
  - E.g. If x represents people's weights, then $\sum x$ gives the **sum** of everyone's weights.
  - Counter e.g. (algebra), if $x = £4$, then $\sum x$ makes no sense.
- $\bar{x}$ refers to the mean of x. Notice that x is a **collection of values**, whereas $\bar{x}$ is a single value.
- We could **attach an associated probability** to each value inside a variable, e.g. weight < 60kg = 0.9, weight > 60kg = 0.1. This is known as **random variable** (see Chp 6)

Measures of...



- **Measures of location** are single values which describe a **position** of any particular data in a data set.
- **Measures of central tendency** are statistical indicators relative to **the centre of the data**, i.e. a notion of "average".
- **Measures of spread** measures how spread out is the data set.

BF MATHS



2.1 - Measures of central tendency

Notes

- **Mean, median** and **mode** are measures of central tendency.
- You should already know how to work out the mean, median and mode of a set of *grouped* as well as *ungrouped* frequency tables.
- Mean of **discrete data** $(\bar{x})$ =
- (Estimated) Mean of **ungrouped/grouped frequency tables** =

Example (discrete data)
Find the mean height of the following students:
Height of students, x cm
154 170 165 189 167 180 170

Example (Ungrouped frequency table)

Shoe Size (x)	Frequency (f)
3	6
4	2
5	3
6	4
7	1

The world's *Largest feet ever* belonged to the *Tallest man ever*, Robert Wadlow (USA, 1918–40), who wore US size 37AA (UK size 36) shoes, which equates to 47 cm (1 ft 6.5 in).

2.1 - Measures of central tendency

Example (Grouped frequency table)

Estimate the mean weight of the brown bears

Weight (x grams) of brown bears	Frequency (f)
$280 \leq x < 380$	4
$380 \leq x < 450$	14
$450 \leq x < 550$	30
$550 \leq x < 700$	40
$700 \leq x < 800$	2

Practice (Getting used to graphical calculator)

Find the mean value of each table.

1.

Num children (c)	Frequency (f)
0	2
1	6
2	1
3	1

2.

IQ of L6Ms2 (q)	Frequency (f)
$80 < q \leq 90$	7
$90 \leq q < 100$	5
$100 \leq q < 120$	3
$120 \leq q < 200$	1

3.

Time t	Frequency (f)
$9.5 < t \leq 10$	32
$10 \leq t < 12$	27
$12 \leq t < 15$	47
$15 \leq t < 16$	11

2.2 - Other measures of location

<u>Starter (GCSE Combined mean)</u>

1) The mean maths score of 20 pupils in class A is 62.
The mean maths score of 30 pupils in class B is 75.

a) What is the overall mean of all the pupils' marks.
b) The teacher realises they mismarked one student's paper; he should have received 100 instead of 95. Explain the effect on the mean and median.

2) Archie the Archer competes in a competition with 50 rounds. He scored an average of 35 points in the first 10 rounds and an average of 25 in the remaining rounds. What was his average score per round?

<u>Median</u> (Discrete Data)

Fill the blanks of the following table:

Items		Position of median	Median
1,4,7,9,10	5		
4,9,10,15	4		
2,4,5,7,8,9,11	7		
1,2,3,5,6,9,9,10,11,12	10		

Rule to find position of median of n items:

<u>Median</u> (Grouped Data)

IQ (q)	Frequency (f)
$80 \leq q < 90$	7
$90 \leq q < 100$	5
$100 \leq q < 120$	3
$120 \leq q < 200$	2

Find the position of the median.

2.2 - Other measures of location

Notes (Percentile)

- 1^{st} percentile (P_1) = the value that sits in the
- 50^{th} percentile (P_{50} or Q_2) = the value that sits in the
- 25^{th} percentile (P_{25} or Q_1) = the value that sits in the
- 75^{th} percentile (P_{75} or Q_3) = the value that sits in the
- If the data is presented in a grouped frequency table, we could use **linear interpolation** to find out any percentile.

Example

Given a data set contains 100 numbers (from 1 to 100).

a) Find the 50^{th} percentile (median)
b) Find the 10^{th} percentile
c) Find the interquartile range

Example

Given a data set now contains 25 numbers (from 101 to 125).

a) Find the 50^{th} percentile
b) Find the 12^{th} percentile
c) Find the IQR

2.2 - Other measures of location

Example (Linear Interpolation)

The length of time (to the nearest minute) spent on the internet each evening by a group of students is shown in the table.

Length of time spent on internet (minutes)	30–31	32–33	34–36	37–39
Frequency	2	25	30	13

a Find an estimate for the upper quartile. **b** Find an estimate for the 10th percentile.

BF MATHS

2.2 - Other measures of location

Practice

Estimate the median and interquartile range of the following dataset, which gives the mass of 100 eggs:

Mass, m (g)	Frequency
$40 \leq m < 45$	4
$45 \leq m < 50$	15
$50 \leq m < 55$	15
$55 \leq m < 60$	22
$60 \leq m < 65$	17
$65 \leq m < 70$	16
$70 \leq m < 75$	11
$75 \leq m < 80$	0

Exam Practice

Edexcel S1 Jan 2007 Q4

Summarised below are the distances, to the nearest mile, travelled to work by a random sample of 120 commuters.

Distance (to the nearest mile)	Number of commuters
0 – 9	10
10 – 19	19
20 – 29	43
30 – 39	25
40 – 49	8
50 – 59	6
60 – 69	5
70 – 79	3
80 – 89	1

For this distribution,

~~(*a*) describe its shape. **(1)**~~

(*b*) use linear interpolation to estimate its median. **(2)**

2.3 - Measures of spread

Notes

- Measures of spread include the use of range, interquartile range (IQR), variance and standard deviation.
- IQR can be calculated using *linear interpolation*.
- Variance and standard deviation -> see 2.4 notes

Example

The table shows the masses, in tonnes, of 120 African bush elephants.

Mass, *m* (t)	$4.0 \leqslant m < 4.5$	$4.5 \leqslant m < 5.0$	$5.0 \leqslant m < 5.5$	$5.5 \leqslant m < 6.0$	$6.0 \leqslant m < 6.5$
Frequency	13	23	31	34	19

Find estimates for:

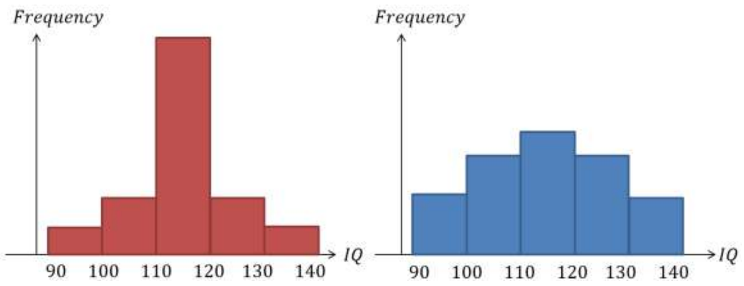
a the range **b** the interquartile range **c** the 10th to 90th interpercentile range.

2.4 - Variance and Standard Deviation

Starter

Pair and discuss:

Here are the distribution of IQs in two classes of the same size. What's the same, and what's different? (Try to use statistical terminologies such as mean, median, range, "spread")



Notes (Variance)

- Variance measures the spread of the data using the mean as the centre, looking at the spread generally.
- Variance, by definition, is the **average squared distance from the mean**.
- The higher the variance, the wider the spread of numbers in the data.
- Var (σ^2) =
- More practical version:

Notes (Standard deviation)

- Standard deviation is calculated as the square root of variance by figuring out the variation between each data point relative to the mean.
- The higher the standard deviation, the further the data is to the mean.
- Standard deviation (σ) =

BF MATHS

2.4 - Variance and Standard Deviation

<u>Example</u>

The marks gained in a test by seven randomly selected students are:

3 4 6 2 8 8 5

Find the variance and standard deviation of the marks of the seven students.

<u>Example / Practice</u>

Find the variance and standard deviation of the following sets of data:

a) 2, 4, 6
b) 2, 3, 3, 5, 7
c) 1, 2, 3, 4, 5

Grouped frequency table

•

<u>Example</u>

Shamsa records the time spent out of school during the lunch hour to the nearest minute, x, of the female students in her year. The results are shown in the table.

Time spent out of school (min)	35	36	37	38
Frequency	3	17	29	34

Calculate the standard deviation of the time spent out of school.

BF MATHS

2.4 - Variance and Standard Deviation

<u>Example</u>

Andy recorded the length, in minutes, of each telephone call he made for a month. The data is summarised in the table below.

Length of telephone call (l min)	$0 < l \leqslant 5$	$5 < l \leqslant 10$	$10 < l \leqslant 15$	$15 < l \leqslant 20$	$20 < l \leqslant 60$	$60 < l \leqslant 70$
Frequency	4	15	5	2	0	1

Calculate an estimate of the standard deviation of the length of telephone calls.

<u>Example/ Practice (Ex. 2E Q8)</u>

The daily mean windspeed, x (kn) for Leeming is recorded in June 2015. The summary data is:

$\Sigma x = 243 \qquad \Sigma x^2 = 2317$

a Use your calculator to work out the mean and the standard deviation of the daily mean windspeed in June 2015. **(2 marks)**

The highest recorded windspeed was 17 kn and the lowest recorded windspeed was 4 kn.

b Estimate the number of days in which the windspeed was greater than one standard deviation above the mean. **(2 marks)**

c State one assumption you have made in producing this estimate. **(1 mark)**

2.5 - Coding

Starter

Here's a scenario….

Imagine you have to measure the weight of 10 chipmunks.

Now, all of them are stuffed with 5 walnuts (assume exactly the same).

Would the mean weight change?

Would the variance change?

If so, how?

Notes

- Coding is a method to mask a set of data into a new set of data which is easier to work with.

Coding Formula	Effect on $\bar{x}$	Effect on σ

Example

A scientist measures the temperature, x °C, at five different points in a nuclear reactor. Her results are given below:

332 °C 355 °C 306 °C 317 °C 340 °C

a Use the coding $y = \frac{x - 300}{10}$ to code this data.

b Calculate the mean and standard deviation of the coded data.

c Use your answer to part **b** to calculate the mean and standard deviation of the original data.

BF MATHS

2.5 - Coding

Example (Ex. 2E Q8)

A teacher standardises the test marks of his class by adding 12 to each one and then reducing the mark by 20%.

If the standardised marks are represented by t and the original marks by m:

a write down a formula for the coding the teacher has used. **(1 mark)**

The following summary statistics are calculated for the standardised marks:

$n = 28 \qquad \bar{t} = 52.8 \qquad S_{tt} = 7.3$

b Calculate the mean and standard deviation of the original marks gained. **(3 marks)**

Practice

Old mean $\bar{x}$	Old σ_x	Coding	New mean $\bar{y}$	New σ_y
36	4	$y = x - 30$		
		$y = 4x$	72	16
35	4	$y = 2x + 15$		
		$y = \frac{x}{5}$	20	$\frac{3}{2}$
11	27	$y = \frac{x+20}{3}$		
		$y = \frac{x-105}{6}$	40	5

BF MATHS