

## About the test

You have a one hour test in the **second session** next week

It is open book and you can use a blank spreadsheet or a calculator to do the work

Students with surnames A - L take the test **in the first hour**

Students with surnames M-Z take the test **in the second hour**

You will write your answers out, but **MUST SHOW WORKING OUT** for full marks

It will be on

**Nominal**

Identifying data types - and explaining why

**Ordinal**

**Ratio**

**Interval**

**Calculating Mean, Median, Mode**

Being able to say what these mean in terms of a data set

It will be on

Finding mean, median and mode from the table

Creating and using a frequency distribution table **Relative FD**

**Cumulative FD**

**Calculating Variance**

Being able to say what variance means in terms of a data set

## Measures of Variability

Sometimes the variance will be calculated from a frequency distribution table

### Example from last time ....

Example: Suppose two machines produce mobile phone covers which are on average 10cm long. 11 covers are selected from each machine.

Machine A: 6, 8, 8, 10, 10, 10, 10, 10, 12, 12, 14

Machine B: 6, 6, 6, 8, 8, 10, 12, 12, 14, 14, 14

Which machine is better?

### Example from last time ....

Machine A: 6, 8, 8, 10, 10, 10, 10, 10, 12, 12, 14

Representing this in a frequency distribution table ..

Machine A: 6, 8, 8, 10, 10, 10, 10, 10, 12, 12, 14

item	frequency
6	1
8	2
10	5
12	2
14	1

Recap: what would the mean, median and mode be?

### Represent in Excel

Add these columns

relative frequency  
percentage frequency  
cumulative frequency

item	frequency
6	1
8	2
10	5
12	2
14	1

Hint: cumulative frequency - add them up cumulatively down the column - can you use a formula?

item	frequency	relative frequency	percentage frequency	cumulative absolute frequency	cumulative relative frequency
	<i>f</i>				
6	1	0.09	9%	1	0.09
8	2	0.18	18%	3	0.27
10	5	0.45	45%	8	0.73
12	2	0.18	18%	10	0.91
14	1	0.09	9%	11	1.00
	11	1	100%		

**Represent in Excel**

**Add these columns**

relative frequency  
percentage frequency  
cumulative frequency

item	frequency
6	1
8	2
10	5
12	2
14	1

**Represent in Excel**

**Now to calculate the variance**

item	frequency
6	1
8	2
10	5
12	2
14	1

The variance of a data set measures how much the data is spread around the mean. To work it out the variance from a frequency distribution

Calculate mean

Find differences to mean

Square differences

Multiply by the frequencies

**The NEW bit**

Add them up and divide by the no of items

item	frequency			
x	f	x-mean	(x-mean) <sup>2</sup>	(x-mean) <sup>2</sup> *f
6	1	-4	16	16
8	2	-2	4	8
10	5	0	0	0
12	2	2	4	8
14	1	4	16	16
	11			48
So mean is		110 / 11	Variance is	48/11 = 4.36

The variance of a data set measures how much the data is spread around the mean. To work it out the variance from a frequency distribution

Calculate mean

Find differences to mean

Square differences

Multiply by the frequencies

Add them up and divide by the no of items

**The NEW bit**

## Now try with Machine B

Machine A: 6, 8, 8, 10, 10, 10, 10, 10, 12, 12, 14      4.36  
 Machine B: 6, 6, 6, 8, 8, 10, 12, 12, 14, 14, 14      10.18

Set up as frequency distribution table  
 Calculate variance (should get 10.18 as above)

- Calculate mean
- Find differences to mean
- Square differences
- Multiply by the frequencies
- Add them up and divide by the no of items

## Exercises

Which numerical measure may be used with any data?

1 mark

If two data sets have the same mean does this imply that they contain the same data?

Fully explain your answer and provide an illustrative example.

3 marks

Table 1 presents a summary of the ages of students on a particular undergraduate degree at the University.

Student Age	Number of Students
18	40
19	28
20	17
21	5
22	10
23	6
24	4

Generate the following statistics for the student age data in Table 1.

- i. the mean      1 mark
- ii. the median      1 mark
- iii. the mode      1 mark
- iv. the variance      3 marks

If five students aged 18, 18, 19, 19 and 22 respectively leave the degree, how would this affect the mean and median values? Your answer does not need to include the new values of these statistics.

2 marks