

Subject: Leaving Certificate Maths

Teacher: Mr Murphy

Lesson 11: Statistics I

11.1 Learning Intentions

After this week's lesson you will be able to;

- ♦ Gather and organise data
- ♦ Describe the different types of data
- ♦ Explain the differences between populations and samples
- ♦ Describe a representative sample
- ♦ Use graphs to illustrate and interpret information

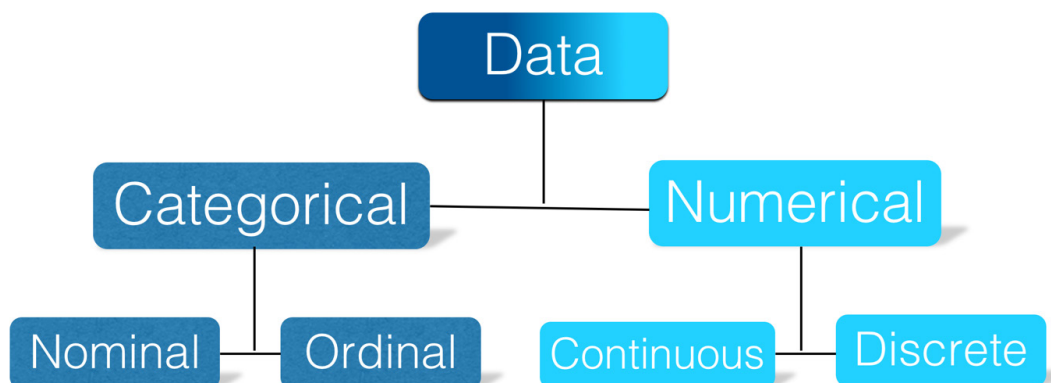
11.2 Specification

Students learn about	Students working at OL should be able to	In addition, students working at HL should be able to
1.5 Finding, collecting and organising data	<ul style="list-style-type: none">– select a sample (Simple Random Sample)– recognise the importance of representativeness so as to avoid biased samples– discuss different types of studies: sample surveys, observational studies and designed experiments– design a plan and collect data on the basis of above knowledge	<ul style="list-style-type: none">– recognise the importance of randomisation and the role of the control group in studies– recognise biases, limitations and ethical issues of each type of study– select a sample (stratified, cluster, quota – no formulae required, just definitions of these)– design a plan and collect data on the basis of above knowledge
1.6 Representing data graphically and numerically	Graphical <ul style="list-style-type: none">– describe the sample (both univariate and bivariate data) by selecting appropriate graphical or numerical methods– explore the distribution of data, including concepts of symmetry and skewness– compare data sets using appropriate displays including back-to-back stem and leaf plots– determine the relationship between variables using scatterplots– recognise that correlation is a value from -1 to +1 and that it measures the extent of the linear relationship between two variables– match correlation coefficient values to appropriate scatterplots– understand that correlation does not imply causality	Graphical <ul style="list-style-type: none">– analyse plots of the data to explain differences in measures of centre and spread– draw the line of best fit by eye– make predictions based on the line of best fit– calculate the correlation coefficient by calculator

11.3 Chief Examiner's Report

Section	Q	Mean Mark	Mean Mark (%)	Mark Ranking (Paper)	Main Topic
A	1	21.6	84	1	Probability
A	2	14.8	59	6	Inferential statistics

11.4 Types of Data



11.5 Sample vs. Population

The **population** is the entire group that is being studied, such as all Leaving Certificate students in 2016.

A **sample** is a small selection from the population.

There are many ways to choose a sample, but you must make the sample as representative of the population as possible. Make sure the sample is sufficiently large. Include all variances of the population, e.g. boys, girls, country students, city students, private students, public students etc.

Methods of Sampling:

Simple Random Sample:

This is where a sample of a certain size is selected from the population and every possible sample has the same chance of being selected.

Stratified Random Sample:

Divide the population into at least two different groups with common characteristic(s), then draw SOME subjects from each group (group is called **strata** or **stratum**).

Systematic Random Sample:

Select a random starting point and then select every n^{th} subject in the population.

Cluster Sampling:

This is based on dividing the population up into groups or clusters. A selection of these clusters is then chosen and all members of the cluster are included.

Quota Sampling:

This is where the participant must satisfy certain criteria in order to be considered for selection. Then a sample is chosen from those.

Convenience Sampling:

This is a non-probability-based method where subjects are chosen in the most convenient way possible.

11.6 Collecting Data

Face to Face Interview:

Gain a lot more than just verbal responses
Adds a human element to the research
Participant may be nervous
Can be time consuming and expensive

Online Questionnaire:

Cheap
Fast
Responses are organized
Anonymity of participants improves honesty

Postal Questionnaire:

Faster than face to face
Cheaper than face to face
Depending on participants returning surveys

Telephone Interview:

Select a sample from entire population
Can explain questions
Expensive compared to postal and online

Observation:

Cheap
Easy to carry out
Cannot explain questions

11.7 Designing a Questionnaire

When Designing a questionnaire:

Simple Language - Avoid Jargon

Brief - Short concise questions

Ease people in - Start with easy questions, closed in type.

Not too personal - Use age ranges rather than specific numbers, wages etc.

Avoid Leading Questions - Maths is the best subject isn't it?

Avoid dual-layered questions - Do you think Ronaldo is the best player in the world and Should move club?

11.8 Stem & Leaf Plots

Each number is separated into a stem and a leaf

54 4

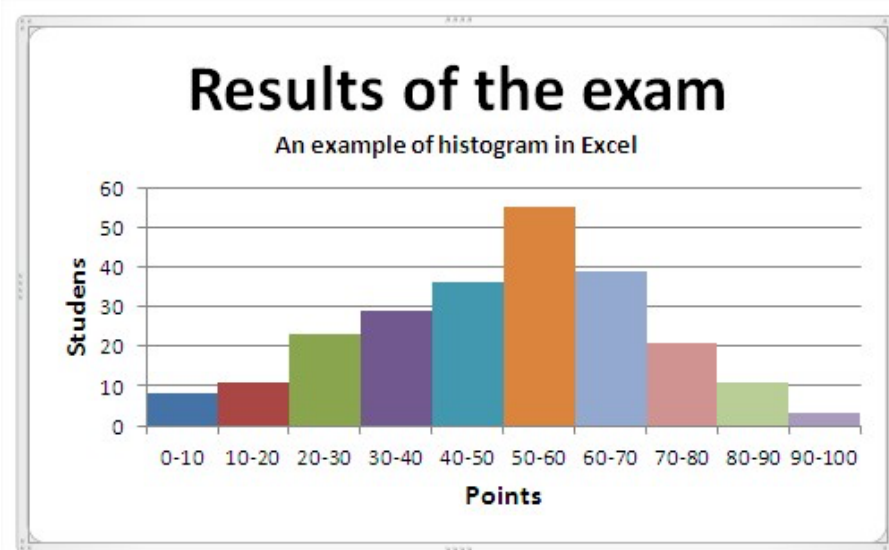
Ensure when drawing a stem and leaf plot that there is always a key at the bottom.
Draw in the stem and leaf diagram from the video below:

Back to Back Stem and Leaf Plots:

We use this version to compare data from two different groups
Draw in the stem and leaf diagram from the video below:

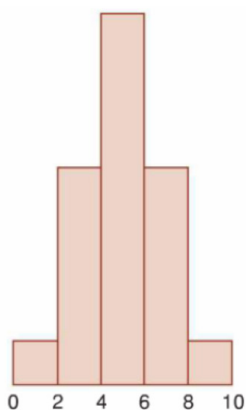
11.9 Histograms

Histograms are used for numerical continuous data.

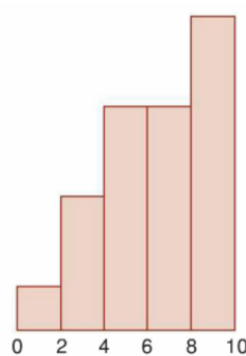


The overall shape of a histogram gives us a snapshot of the data. We can have three situations for a histogram:

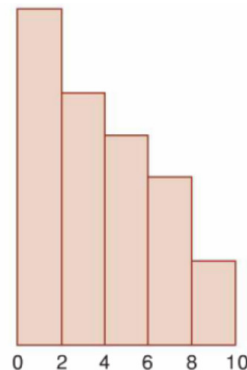
1) Symmetric



2) Negative (Skewed Left)



3) Positive (Skewed Right)

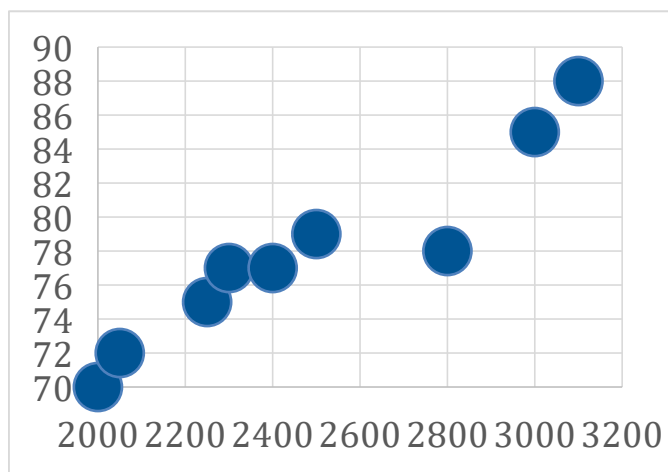


11.10 Scatter Plots

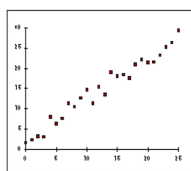
We use scatter plots to compare to sets of data. This data is referred to as **Paired** or **Bivariate Data**

When the points are plotted on a graph the closer, they are to resembling a straight line, the stronger the **correlation**.

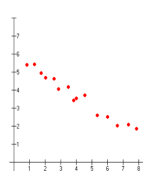
Body Mass vs. Daily Caloric Intake



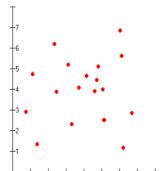
With a correlation there are 5 different types we can have:



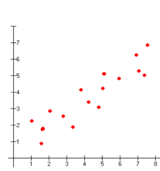
Strong Positive



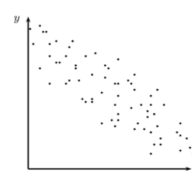
Strong Negative



No Correlation



Weak Positive

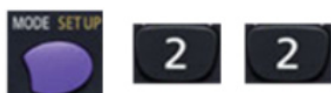


Weak Negative

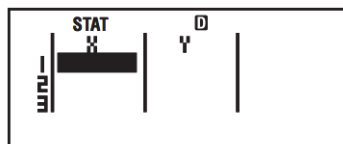
However, although these give us a general relationship, in order to accurately compare different correlations we need to establish a value for the correlation coefficient. We call this value r .

Calculating the r value (Casio fx-83GT):

1. Set up to deal with paired data



2. Input the data into the table



3. Once data entry is finished, press



Screen will read 0, this is ok

4. To finish the calculation press the following



5. Your value should range from -1 up to +1

The correlation coefficient for the below data is _____, indicating a _____ correlation.

Calories	Body Mass
2000	70
2500	79
2250	75
2300	77
3000	85
2800	78
3100	88
2400	77
2050	72

11.11 Recap of the Learning Intentions

After this week's lesson you will be able to;

- ♦ Gather and organise data
- ♦ Describe the different types of data
- ♦ Explain the differences between populations and samples
- ♦ Describe a representative sample
- ♦ Use graphs to illustrate and interpret information

11.12 Homework Task

Question:

75 mrks

An economics student wants to find out whether the length of time people spend in education affects the income they earn. The student carries out a small study. Twelve adults are asked to state their annual income and the number of years spent in full-time education. The data are given in the table below, and a partially completed scatter plot is given.

