NOTTINGHAM ACADEMY

Year 10

GCSE Statistics

Summer Work

SCAN ME

Name .....

Types of Data

The data we collect can be split into two categories:

Qualitative Data which is _____

and **Quantitative** which is _____

We can then further split quantitative data into more specific categories:

Discrete Data which is _____

some examples of this could be:

- _____
- _____
- _____

and **Continuous** Data which is _____

some examples of this could be:

- _____
- _____
- _____
-

SCAN ME

QUICK QUESITON:

Mark is preparing for his French Speaking Exam. He writes out what he wants to say in English first.

'My name is **MARK** and I am **15** years old. I have **BLOND** hair and **GREEN** eyes and I am **170**cm tall. My favourite sport is **FOOTBALL**. Last season I scored **7** goals for my team. Most days I get the **BUS** to school but I try to walk **1** day a week'

Look at the **BOLD** words and numbers. How many of these are examples of:

- ☐ Qualitative data
- ☐ Discrete data
- ☐ Continuous data

The way in which we collect our data can also be categorized into two categories:

Primary Data which is _____

an example of this could be:

- _____

Primary data is good because

1. _____
2. _____

and **Secondary** Data which is _____

an example of this could be:

- _____

Secondary data is good because

1. _____
2. _____

Finally, the way in which data is measured can be categorised, usually into one of three groups:

Categorical data is when _____

_____

Ranked data is when _____

_____

Bivariate data is when _____

_____

QUICK QUESITON:

For each of the following sets of data, choose whether they can be classed as **categorical, bivariate** or **ranked**

a) The year group a student is in

b) The age and heights of the students

c) The league positions of the football teams

Sampling

Everything or everybody that could possibly be involved in an investigation is known as the _____

Data containing everything about every member of a population is known as a _____

A set of data which contains information about part of a population is called a _____

Sample Units are _____

_____

A **Sample Frame** is _____

_____

SCAN ME

Sampling Methods

A **RANDOM SAMPLE** is _____

_____

Everyone and everything from the population has an _____ of being chosen

You must use the following format when writing how to take a **RANDOM SAMPLE**

1) Number the pupils from the sampling frame

2) Choose random numbers from a random number generator

3) Ignore any repeats and numbers > n

4) _____

A **STRATIFIED SAMPLE** is _____

_____

To calculate the number of subjects in each strata we use the formula: $\frac{\square}{\square} \times \square$

A **SYSTEMATIC SAMPLE** is _____

_____

A **CONVENIENCE SAMPLE** is _____

_____

A **QUOTA SAMPLE** is _____

_____

A **CLUSTER SAMPLE** is _____

_____

QUICK QUESTION

For each of the following scenarios, choose which method of sampling is being described:

R-Random **ST**-Stratified **S**-systematic **C**-Convenience **Q**-Quota **C**-Cluster

- I stop the first 10 people I see on the corridor ☐
- I number all the students and randomly select 10 numbers and hence the corresponding students ☐
- I stop the first 5 boys and 5 girls I see on the corridor ☐
- I number all the forms in school and randomly select all the pupils from one form ☐
- I choose every 10th person from the school register ☐

Interviews

Interviews come in three forms:

Cheapest ↓

	Least likely to have interviewer bias ↓
Most Expensive	Most likely to have interviewer bias

Questionnaires

A **Pilot Survey** is a _____ of some of the possible questions you may want to ask. You ask a smaller sample than you intend to give the questionnaire to

We use them to:

1. _____
2. _____
3. _____

A **Questionnaire** is _____

_____

A question can be written with no suggested answers which is known as an _____ question or it may have a set of answers to choose from which is known as an _____ question

When writing a questionnaire we must ensure that each question follows a set of rules. We use the acronym **COLE** to help us remember:

C _____

O _____

L _____

E _____

QUICK QUESTION

Rowan wants to find out how much pocket money his friends get each week. He gives them the following short questionnaire.

Use **COLE** to criticize the questions he has used

1. Do you think you should get more pocket money than you currently do?

   Yes ☐ No ☐

2. How much do your parents earn per year?

   £ ☐

3. How much pocket money do you receive each week?

   £0 ☐ £1-£5 ☐ £5-£10 ☐ £10+ ☐

Experimental Design

SCAN ME

The purpose of randomisation is to _____

This is usually done through _____ although this is often difficult to execute well

However, if you take notice of every criticism you will end up _____

When computing any experiment, it is important we have 2 groups take part in each trial for comparison

- The _____ group who have the treatment such as take a new drug
- The _____ group who do not have the treatment

  This group may take a _____ rather than receive no treatment at all

  This is where the subjects receive a treatment which should have no effect on them but are not aware of this result

It is better if the two groups are _____ so that the comparisons made can be more accurate

This does not mean that the people within the group must be similar to each other just that the two groups as a whole should be similar

A _____ is when the subject does not know which group they are in

These are done to eliminate false results.

Some patients may appear to improve because they believe that the treatment will make them better - giving false information

A _____ is when the experimenter does not know which group they are in either

An experimenter may, consciously or sub-consciously, let the subject know which group they are in – giving false information

Cleaning Data

When you have some data the first thing you need to do is to check it out and get rid of any obviously wrong or false data

Key errors to be on the lookout for:

- **Pointy Pete** _____
- **Obvious Olive** _____
- **Silly Samantha** _____
- **Devious Dave** _____

QUICK QUESTION

Clean the data below by highlighting any wrong or false data

Name	Age	Height (cm)	Shoe	Hair	Eyes	Subject	Animal	Colour
James	75	154	8	Grey	Brown	Home time	Dog	Blue
Clarys	64	153	4	White	Blue	English	Cat	Red
Frank	69	139	9	Brown	Green	Maths	Canary	Green
Iris	82	1.56	5	Pink	Brown	Science	Flamingo	Pink
Ivan	78	165	100	Brown	Brown	History	Dog	Blue
Rose	74	148	5	Grey	Blue	Music	Rabbit	Yellow
Marvin	18	184	12	Blonde	Blue	Geography	T-Rex	Blue

Pictograms

A pictogram is an example of a statistical graph which uses pictures to represent _____ data

A pictogram must have:

- A _____
- Each picture must be the _____

SCAN ME

Bar Charts

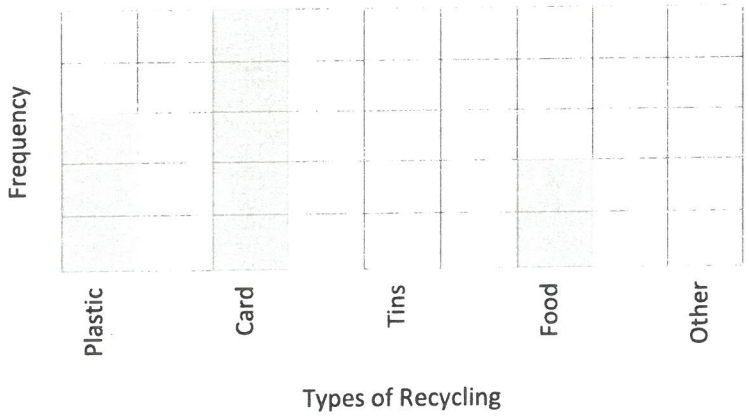
A bar chart is an example of a statistical graph which uses bars to represent _____ data

A bar chart must have:

- _____ between the bars
- These must be of _____ width
- The bars must also be of _____ width

QUICK QUESTION

The pictogram and bar chart below display the same data. Complete the missing information



Types of Recycling	
Plastic	⌀ d
Card	
Tin	⌀ ⌀
Food	
Other	⌀ ⌀ d

Key	⌀ means 20 items

Comparative Bar Charts

Comparative bar charts allow you to _____ more than one set of data for each variable

A comparative bar chart must have:

- A _____
- _____ between each variable
- NO _____ between bars of the same variable
- Bars must be of _____ width

Composite Bar Chart

Composite bar charts allow you to compare the _____ of variables between sets of data

A composite bar chart must have:

- The _____ plotted on the y axis rather than the frequency
- Bars which are _____ of one another rather than side by side
- _____ between the bars
- These must be of _____ width
- The bars must also be of _____ width

Stem and Leaf Diagrams

A stem and leaf diagrams is used to display volumes of _____ data

To the right of the diagram is the 'leaf' which is only the _____

To the left of the diagram is the 'stem' which is usually _____

Any Stem and Leaf Diagram must:

- Be in _____
- Have a _____

SCAN ME

From a Stem and Leaf diagram we can interpret data in the usual way but quicker and more accurately.

- The number of _____ tell us the number of pieces of data in the diagram
- The data is already in _____ which makes finding the highest values, lowest value and median easier

QUICK QUESTION

For the following set of data and steam and leaf diagram, showing the test scores of year 9 students, fill in the missing values

121 ___ 131 ___ 123 135 136 142 132 124 ___ 128 ___ 139 145

12	1	1		4	7	
13		2	5	5		9
14	1	2				

How many students were sampled for this diagram? _____

What was the average (median) test score of the year 9 students in this sample? _____

What was the range (highest – lowest) of scores for the year 9 students sampled? _____

Back-to-back Stem and Leaf

Two sets of data can be _____ on a 'back to back' stem and leaf diagram

In this case the _____ runs down the centre of the diagram and the _____ come out either side

Any Stem and Leaf Diagram must:

- Be in _____ (smallest closest to the stem)
- Have a _____ (the same for both sides of the diagram)

QUICK QUESITON

For the following sets of data showing the test scores for boys and girls, complete a back-to-back stem and leaf diagram

Boys 26 49 35 42 34 39 26 15 32

Girls 17 18 15 35 36 29 28 49 49 49 48 7

0	
1	
2	
3	
4	

Which group had the highest score on average (median)? _____

Two-Way Tables

Two-way tables are used to organise information about two separate variables simultaneously

They allow you to present a lot of data easily

We are able to calculate both row and column totals and to calculate the grand total we simply add together

EITHER all of the row totals or all of the column totals

SCAN ME

QUICK QUESTION

For the two-way able below, fill in the missing values form the table:

	Went to the match	Watched it on TV	Listened on the radio	Total
Wolves	14	32		58
Albion		16	15	
Villa	21			52
Total	54		40	

Misleading Statistics

Statistics are often used and manipulated to make the data look better (or worse) without lying to the reader.

When looking at any statistics we must look for

What is 'misleading'	What impact this may have on the reader
Missing labels	
Unequal gaps along scales	
Large breaks in the scales	
3D graphs	
Incorrectly calculated angles	
Time series without repeated seasons	
Different scales used for box plots	
Key not included	

Measures of Central Tendency

Measures of Central Tendency (also known as Measures of Location) give you information about the location of the data set in reference to a number line

SCAN ME

The _____ is the most common value or the variable with the highest frequency in a data set

When we have data presented in a grouped frequency tables this may be called the _____ class

The most common average is known as the _____ which can be found using the formula $\frac{\Sigma x}{n}$ or $\frac{\Sigma fx}{\Sigma f}$ for frequency tables

When we have data presented in a grouped frequency tables we use the _____ to *estimate* this average

It is only ever an estimate because _____

_____

The _____ is the middle number in a data set. Its position within the data set can be found using the formula 0.5(n+1)

When we have data presented in a grouped frequency tables we draw a _____ to *estimate* this average

Weighted Mean

A Weighted Mean allows us to calculate the 'average' of a set of results where one result has more weight than another.

To calculate the weighted mean we:

- _____ each value by its weight
- _____ these values together
- _____ this total by the total of the weights

Geometric Mean

The Geometric mean is the _____ th root of the _____ of n numbers

It is often used to calculate the mean of _____ rather than the arithmetic mean

QUICK QUETION

Find the geometric mean of 4, 6, 8 and 12

Transforming the Mean

When data involves large or awkward numbers you can sometimes make the numbers easier by **scaling**. this means you make fewer errors in your calculations

We can either:

- add or subtract the same constant from each value so the numbers you use to calculate the mean with are simpler.
  - We then _____ or _____ back the constant to the mean
- multiply or divide by the same constant for each value so the numbers you use to calculate the mean with are simpler.
  - We then _____ or _____ back the constant to the mean

Correlation

Correlation is the way in which the data is _____ on a scatter graph

There are three types of correlation:

1. _____ where as the x variable increases so does the y variables
2. _____ where as the x variable increases the y variable *decreases*
3. _____ there is no relationship between the x and y variables

SCAN ME

We must be careful not to mistake correlation for causality and remember that:

_____ implies a connection between 2 variables

_____ implies a direct link between 2 variables. i.e. 1 variable causes the other

Scatter Graphs

When plotting a scatter graph we must always plot the _____ variable

(the one we are in control of during the experiment) on the x axis

And the _____ variable (the one that we measure during the experiment) on the y axis

Line of Best Fit

You can draw a line of best fit on a scatter diagram if the graph shows there is _____

The line of best fit will pass through _____ of both sets of data

The equation of the line of best fit can be found using _____

Where _____ is the gradient (rise over run) and _____ is the y-intercept (where is crosses the y axis)

QUICK QUESITON

What type of correlation is shown on this scatter graph?

_____

Hence, what can you conclude about the relationship between the body length and wingspan of the birds shown?

_____

_____

_____

_____

_____

_____

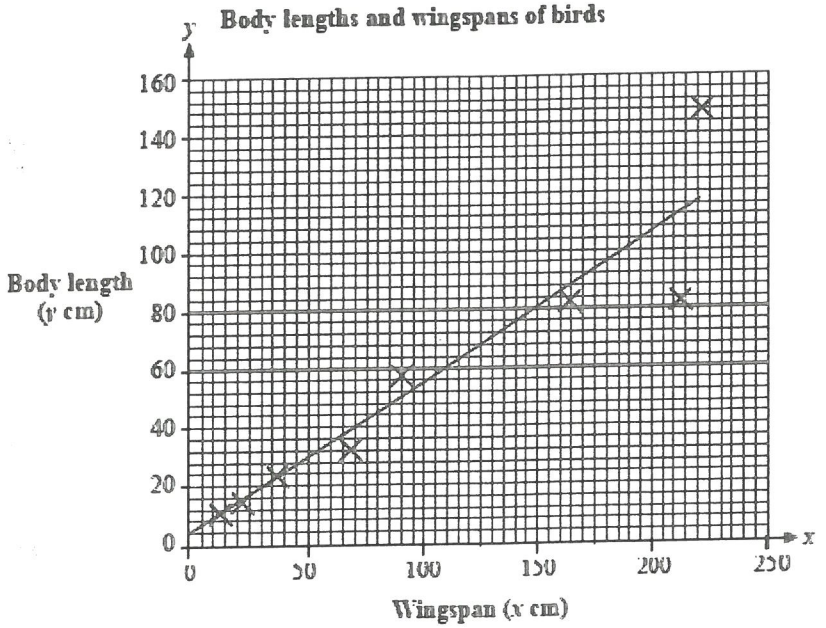
_____

_____

_____

_____

The scatter diagram below shows the body lengths and wingspans of the same birds.



Find the equation of the line of best fit drawn on the scatter graph

_____

Hence, estimate the body length of a crow with wingspan 100cm

_____

Probability

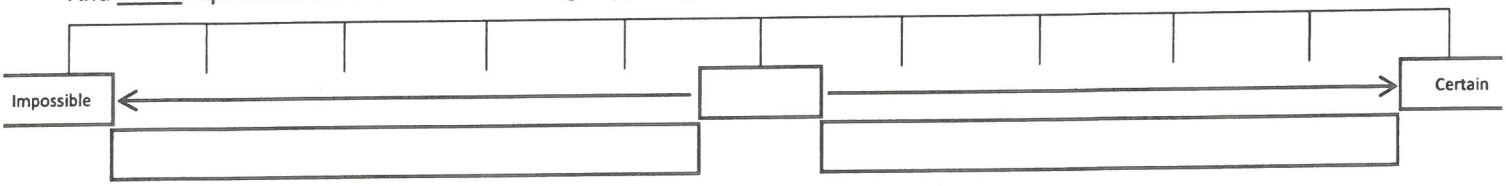
The Probability Scale

The *Probability Scale* is a scale which goes from _____ to _____

Where _____ represents the chance of something happening to be impossible

And _____ represents the chance of something happening to be certain

SCAN ME



Theoretical Probability

Theoretical probability is calculated based upon previous knowledge.

It is usually calculated with the assumption that there is no _____ present or if there is, this is taken into account in the calculations

It will give you a *guide* as to what should happen in an experiment but will don't give you a definitive answer of what to expect.

Probabilities are usually display as _____ or _____

We calculate Theoretical Probability as: $\frac{\quad}{\quad}$

Experimental Probability

Experimental probability (also known as _____) is calculated based upon the results of an experiment.

It is considered to be a more _____ representation of the probabilities of the outcomes of an experiment as it takes into account _____ which may be present

It will give you a *guide* as to what has happened previously but will don't give you a definitive answer of what to expect in the future.

Probabilities are usually display as _____ or _____

We calculate Experimental Probability as: $\frac{\quad}{\quad}$

QUICK QUESTION

Complete the table below of the probability of obtaining each of the following outcomes when rolling a fair die

x	0	1	2	3	4	5	6
P(X = x)							

Julie rolls her die 200 times with the following results:

x	0	1	2	3	4	5	6
P(X = x)	0	9/40	4/25	29/200	37/200		1/5

a) Complete the missing relative frequency from the table above

b) Does this experiment suggest the die is biased?

Index Numbers

An index number is a number showing the _____ compared with the value at a specified earlier time

Index Number =

SCAN ME

Price Relative

Price relative is the most commonly used index number. It allows us to compare values from a starting point - we can look at if there has been an _____ or _____ compared to the original price

Price Relative =

QUICK QUESITON

The table below shows the average price of a Mickey Mouse cushion in the Disney Store over the last 5 years.

Year	2014	2015	2016	2017	2018
Average Price	£12	£16	£14	£15	£12

(a) Without doing any calculations, what would you expect the price relative for 2018 to be? Why?

_____

_____

(b) Using 2014 as the base year, calculate the price relative of the cushion for the following four years.

Chain Base Index Numbers

To find out how the price of an item has changed over a year, you use the _____ as the base year

A Chain base index tells you the annual _____

QUICK QUESITON

The table below shows the average price of a Minnie Mouse cushion in the Disney Store over the last 5 years.

Year	2014	2015	2016	2017	2018
Average Price	£8	£12	£10	£12	£14

Calculate the chain base index numbers of the cushion for each year