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

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 _____

SCAN ME

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

SCAN ME

A pictogram must have:

- A _____
- Each picture must be the _____

Bar Charts

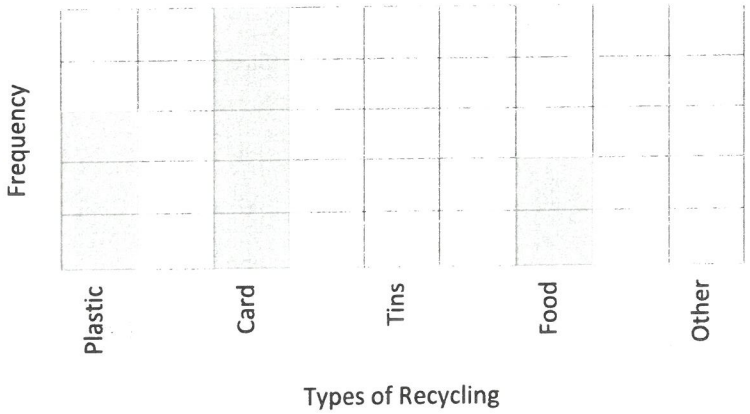
A bar chart is an example of a statistical graph which uses bars to represent _____ data

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	db d
Card	
Tin	db db
Food	
Other	db db d

Key	db 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

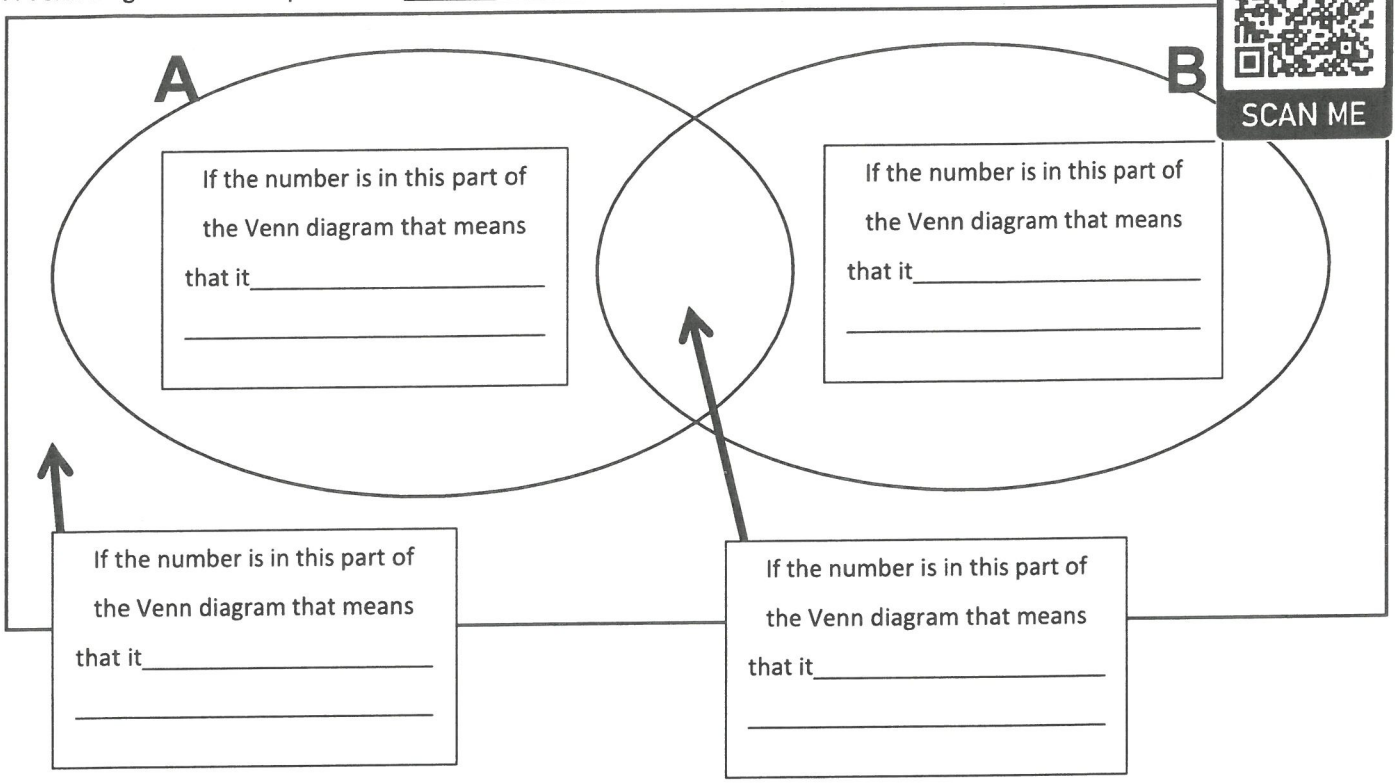
0	
1	
2	
3	
4	

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

Venn Diagrams

A Venn diagram is made up of 2 or 3 _____ surrounded by a _____

SCAN ME



The number of pieces of data in a particular variable on the diagram (E.G. A) must be the total of all the values in the _____

The number of pieces of data that appear in all the variables on the diagram (E.G. A&B) must be the value in the _____

When we draw a Venn diagram we must include the number outside the circles, even if it is _____

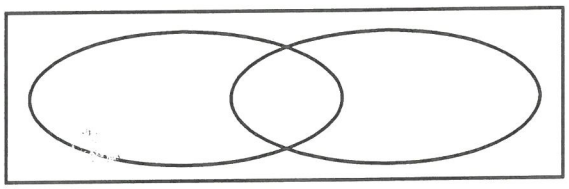
It is important when completing a Venn datagram that you start in the _____ and work your way out

QUICK QUESTION

Out of forty students, 14 are taking English Composition and 29 are taking Chemistry.

Five students are in both classes.

Complete the Venn diagram to display this information



Notation

The proper name for the overlapping part of a Venn diagram is the _____

It is denoted with the symbol _____ *E.G. the overlap of A&B is written as* _____

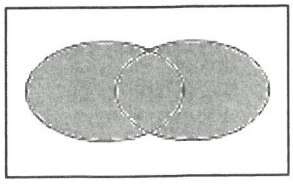
The proper name for the overlapping part and the remaining sections of the circle in a Venn diagram is the _____

It is denoted with the symbol _____ *E.G. A&B is written as* _____

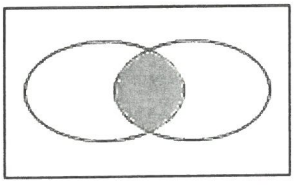
QUICK QUESITON

Match each of these notations to its diagram

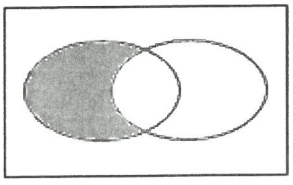
AUB



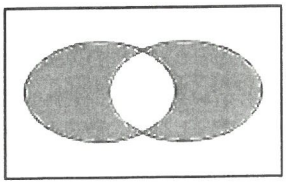
AnB



(AUB)-(AnB)



A-(ANB)



Pie Charts

A pie chart is a type of statistical diagram which is divides up a circle to represent different _____ of a population

A pie chart is in the shape of a _____ and therefore _____ degrees is an important number

SCAN ME

To calculate the angle in a pie chart we use the following formula:

$$\text{Angle} = \frac{\square}{\square} \times \square$$

QUICK QUESTION

For the data shown in the table below, calculate the angle that would need to be drawn if it were to be represented in a pie chart. SHOW YOUR WORKING OUT

House	Frequency	Angle
Hanover	289	
Lancaster	274	
Stuart	306	
Tudor	321	
York	296	

Comparative Pie Chart

As a pie chart only shows the _____ of the population, we only know how many are represented by each segment once we know the total number in the population

Comparative pie charts take this into account, where the _____ of the pie chart represents the total population

HENCE; the larger the population the larger the _____ of the circle

It is important, therefore, to remember that the area of a circle is: $\square \times \square$

To calculate a *proportionate* radius for comparative pie charts we must follow the following steps:

1. State the radius of pie chart 1
2. Calculate the area of pie chart 1
3. Find the percentage change between the two populations
4. Multiply the percentage change by the area for pie chart 1

   THIS IS NOW YOUR AREA FOR PIE CHART 2

   We must now solve the equation to calculate the radius of pie chart 2
5. Divide the area of pie chart 2 by Pi
6. Square root this answer

QUICK QUESTION

Dolly wants to draw comparative pie charts to compare how she spends her income. Last year she was a college and her income was £3000. She uses a radius of 3cm for this pie chart. This year she earns £9800. What radius should she use for this pie chart?

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____ **ANSWER:** _____cm

Cumulative Frequency

Cumulative frequency is defined as _____ of all frequencies

To calculate cumulative frequency we simply _____ the frequencies as we go along the table

SCAN ME

QUICK QUESTION

Calculate the cumulative frequencies for the table below

Width	0≤w<2	2≤w<4	4≤w<6	6≤w<8	8≤w<10	10≤w<12	12≤w<14	14≤w<16	16≤w<18
Frequency	6	7	1	2	5	4	0	3	2
C.F									

Cumulative Frequency STEP Polygon

A cumulative frequency step polygon is used for _____ data

The variable is plotted along the _____axis

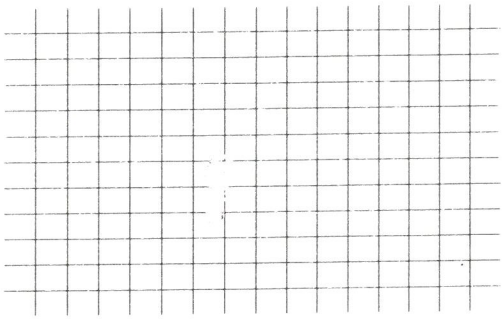
The cumulative frequency is plotted along the _____ axis

We then join the points together with_____

QUICK QUESITON

Draw a cumulative frequency diagram on the grid to the right for the table of data below

X	4	5	6	7
F	3	2	0	5
C.F				



Cumulative Frequency Polygon

A cumulative frequency polygon is used for _____ data

The _____ is plotted along the x-axis

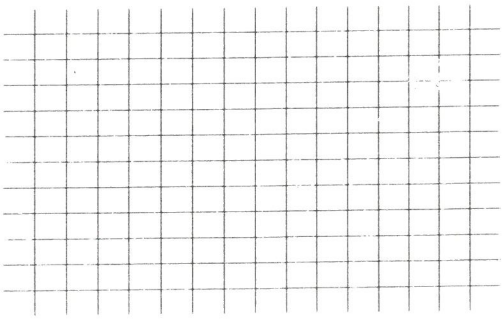
The _____ is plotted along the y-axis

We then join the points together with _____

QUICK QUESTION

Draw a cumulative frequency diagram on the grid to the right for the table of data below

X	10≤w<20	20≤w<30	30≤w<40	40≤w<50
F	3	1	2	3
C.F				



Histograms

A histogram is used when we want to draw a statistical diagram for _____ data

We use it instead of a continuous bar chart when the widths of the bars are _____

The height of the bar is called the _____

This is calculated using the formula

SCAN ME

The _____ of the bar is the frequency

Therefore the frequency of a bar can be calculated with _____ x _____

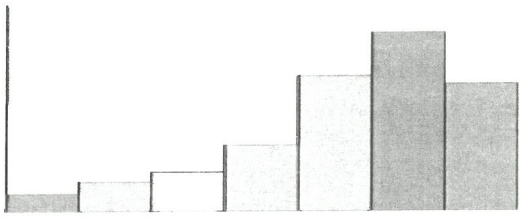
QUICK QUESITON

Calculate the frequency density for each category in the data below:

Weight (g)	Frequency	
$0 \leq w < 50$	3	
$50 \leq w < 70$	20	
$70 \leq w < 90$	21	
$90 \leq w < 110$	19	
$110 \leq w < 150$	7	
$150 \leq w < 200$	2	

We can also use a histogram to describe the ***shape*** of a distribution

A histogram which shows a distribution with two modes would be known as _____

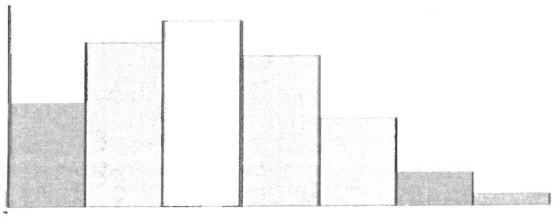


A histogram which has the majority of its data towards the right of the diagram is known as

_____

A histogram which has the majority of its data towards the left of the diagram is known as

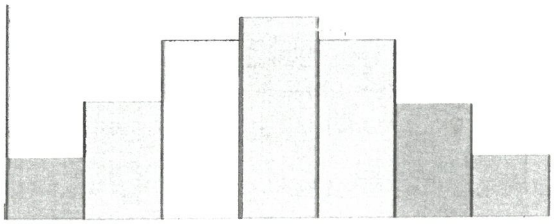
_____



A histogram which has the majority of its data

In the centre of the diagram is known as

_____



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

Measures of Spread

SCAN ME

Measures of Spread give you information about how 'spread out' the data is

The _____ gives you the width of the full data set by finding the difference between the highest and lowest observed values

The _____ is the difference between the lower and upper quartile

The position of the **Lower Quartile** can be found using the formula _____

The position of the **Upper Quartile** can be found using the formula _____

QUICK QUESTION

Find the IQR for the following set of numbers

4 5 8 8 9 10 10 12 13 15 15

Lower Quartile _____

Upper Quartile _____

IQR _____

Variance and Standard Deviation

The variance and standard deviation are a way of looking at how _____ the data is around the _____

The Variance and Standard Deviation are closely related

The Variance is the _____ of the Standard Deviation

The Standard Deviation is the _____ of the Variance

On the formula page we are given 2 versions of the standard deviation formula

$$\sqrt{\frac{1}{n}\sum(x-\bar{x})^2} \quad \text{and} \quad \sqrt{\frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2}$$

QUICK QUESTION

Find the Standard Deviation of the following data:

4 5 8 8 9 10 10 12 13 15 15

Box and Whisker Plots

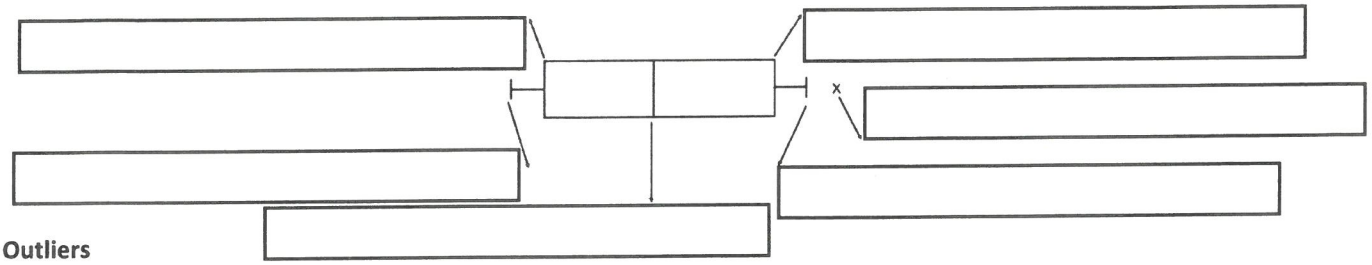
A box and whisker plot is used to display _____ data

The 'box' is made up of three values: the _____, _____ and _____

The 'whiskers' are made up of two values: the _____ and _____

And any _____ are represented with a 'x'

SCAN ME



Outliers

Outliers are values that are unusual in comparison with the rest of the data

An outlier is either:

- _____
- _____

To find outliers you use the following formula:

Anything above _____ or *anything below* _____

We can also use the formula $\mu \pm 3\sigma$

Comparing Box and Whisker Plots

When comparing box plots it is important to look for how many comparisons they want. The number of marks on offer tells you

_____

You need to do your best to comment on different aspects of the box plots rather than all on the same thing.

You must also make sure that you use the statistical language

The four main comparison areas are:

SP _____

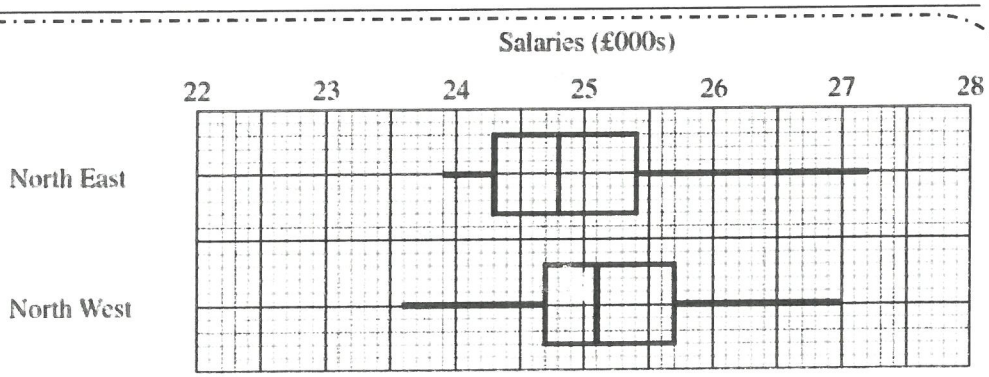
A _____

S _____

I _____

QUICK QUESITON

Compare these two box plots showing the salaries of samples of employees in the North East and north West



_____

_____

_____

_____

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?

_____

_____

_____

_____

_____

_____

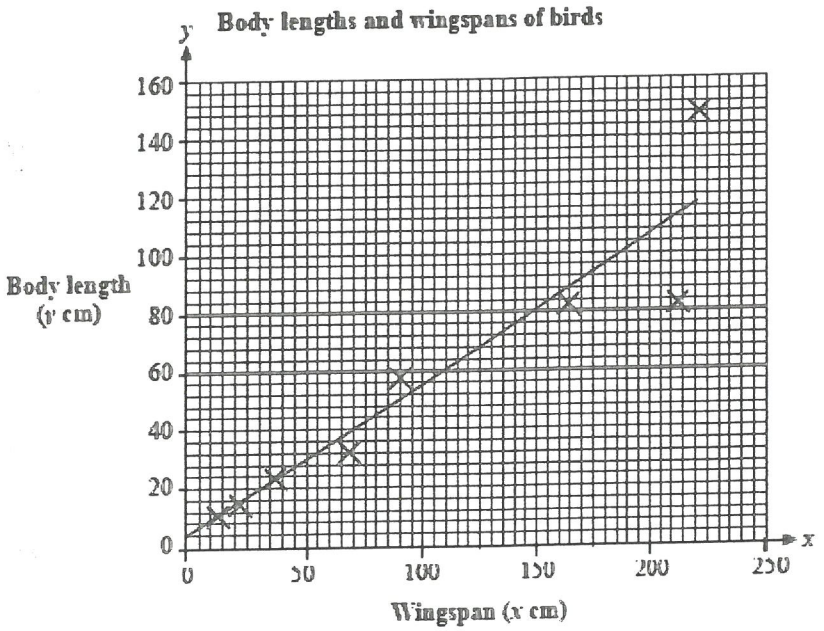
_____

_____

_____

_____

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

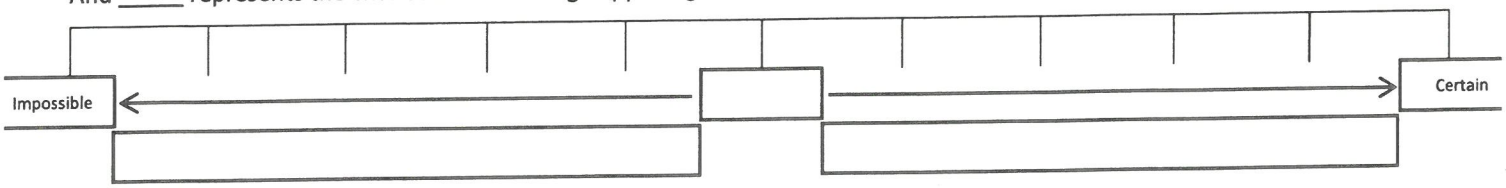
The Probability Scale

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?

Sample Space Diagrams

A Sample Space Diagram is a way of displaying all of the _____ of 2 events occurring at the same time

BE CAREFUL: We do not use these to show _____ (for this we use a two-way table)

These can then be used as a visual aid to calculate _____

SCAN ME

QUICK QUESITON

A fair die is thrown and a fair coin is flipped.

a) Draw a sample space diagram to show the possible outcomes of this experiment

b) Calculate the probability of scoring an even number and a head

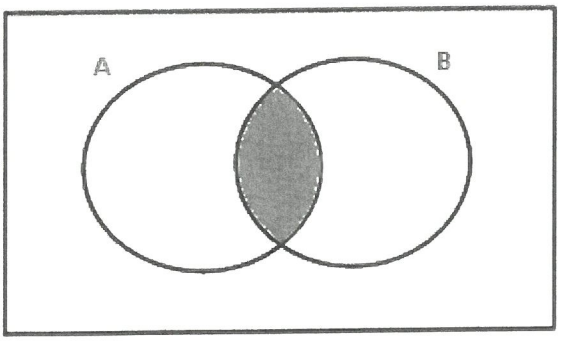
c) Calculate the probability of scoring multiple of 3 and a tail

d) Given that we score a prime number, calculate the probability of obtaining a head.

Venn Diagrams

A Venn diagram is a diagram representing mathematical or logical sets pictorially as circles or closed curves within an enclosing rectangle (the universal set), common elements of the sets being represented by intersections of the circles.

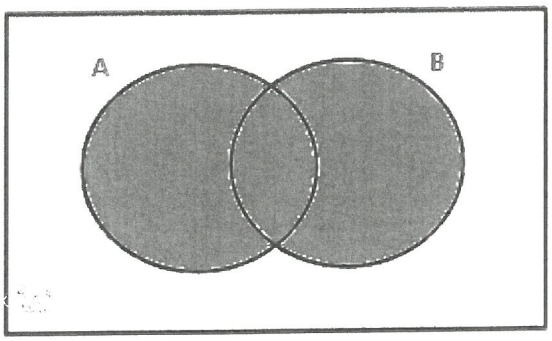
We use a Venn diagram to represent _____ or _____



This is called the _____

It is represented as _____

= _____



This is called the _____

It is represented as _____

= _____

Tree Diagrams

A tree diagram can be used to represent either _____ or _____

SCAN ME

Probabilities on the branches of a tree (in a column) must add up to _____

As we go along each branch or a tree diagram we _____ the probabilities

When combining more than one total probability, as we go down the tree, we _____ the probabilities

We can use a tree diagram to show

- _____ events (events which happen over and over again but the probabilities stay the same)
- _____ events (where several different events occur without effecting the outcome of one another)
- _____ events (different events occur and the probability of each depends upon the outcome of the last)

QUICK QUESTION

There are 12 counters in a bag. 3 are red, 2 are blue and the rest are green

Khalil draws a counter from the bag, records the colour then replaces it. He then draws a second counter form the bag and records the colour.

a) i) Draw a tree diagram to represent this experiment.

ii) What is the probability that Khalil draws the same colour both times?

b) Jenson does the same experiment but does NOT replace the counter after the first selection. What is the probability he draws the same colour both times?

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

Weighted Index Numbers

To calculate a weighted index number you:

- _____
- _____

Weighted Index number =

Population Averages

These are a way of expressing an estimate for important statistics such as _____

They are expressed as a _____ of the population

Crude (birth/death/...) rate =

Standardized Rates

Standardized rates take account of the age distribution of a population

To calculate a standardized rate:

- _____
- _____

Standardized Rates =

Retail Price Index

The retail price index is a _____ of the price relatives of goods and services

Weightings are chosen to reflect the spending habits of an _____ household

---

Quality Assurance

Statistical process control may be used when a large number of similar items are being produced.

Its purpose is to give a signal when the process mean has moved away from the target value or when item-to-item variability has increased.

The most common method of statistical process control is to take samples at regular intervals and to plot the sample mean on a control chart

For control charts for means:

Sample mean between warning limits : _____

Sample mean between warning and action limits : _____

Sample mean outside action limits: _____

_____

The *warning* limits are set at _____ and _____% of the data should lie within the warning limits

The *action* limits are set at_____ and _____% of the data should lie within the action limits

Where _____ is the target value, _____ is the short-term standard deviation and _____ is the sample size.