OCR Psychology A-level

Paper 1: Research Methods

Mathematical Skills

The specification requires students to have knowledge of and be able to apply several mathematical skills. Around 10% of the marks in the paper 1 Research Methods examination will be designated for maths, so it is important to understand each of these skills. The following document covers each skill required by the official OCR specification.

D.0 - Arithmetic and numerical computation

- **D.0.1 - Decimal and standard form**
  - ➢ The examination may require the conversion of data into or from standard form/ decimal form to another type of data to construct a data type or to demonstrate ability.
  - ➢ **Decimal form** is a form of displaying numbers using decimal places, either to represent large or small numbers.
  - ➢ For example, 200, 0.2, 0.62 ect.
  - ➢ Decimal form is precise but can be inefficient for very large numbers, such as 200000 which is difficult to read and do calculations on.
  - ➢ Therefore, **Standard Form** can be used to represent very large or small numbers in an easier way.
  - ➢ Standard form works by representing a number by how many decimal places it has: the rules are the standard form number must be above 0 and below 10, with the index/power representing the numbers of decimal places to move.
  - ➢ For example, to represent 2,305,000 as a standard form number:
    - Identify the number under 10 that can be used to represent the significant digits. In this case it will be 2.305.
    - Calculate how many decimal places this new number must move to recreate the original number. In this case, the decimal place in 2.305 must move 6 places right to recreate 2,305,000. (Add zeros where there are no more places to move).
    - If the number is very small instead of very large, the decimal place will move left instead.
    - The standard form number will then be written as **2.305 x10 ^ 6**
- **D.0.2 - Ratios, fractions and percentages**
  - ➢ A **ratio** is a way of expressing the relationship between 2 amounts.
  - ➢ For example, 3:1.
  - ➢ If there are 20 males in a class and 5 females, the ratio of males to females is 20:5 - but this can be simplified to 4:1 by dividing each side by 5.
  - ➢ Therefore, the ratio is written in its simplest form of 4:1.
  - ➢ A **fraction** is a way of expressing a number as a portion of a whole number.
  - ➢ For example, a half can be expressed as ½, because 1 divided by 2 gives 0.5 - equivalent to a half.
  - ➢ Fractions can be simplified - if a student gets 45/100 questions on a test correct, the fraction can be simplified to 9/20 by dividing each side by 5. Each one gives the same decimal number of 0.45
  - ➢ A **percentage** is a number expressed as a fraction of 100, and is usually denoted with a percentage sign.
  - ➢ The exam may require calculations such as working out the percentages from raw data collected from a given study.

➢ To calculate percentage:
- Using the following example of raw data showing the amount of people in an office who use types of transport to travel to work:

Bike	Bus	Car	Train
10	15	45	5

You can calculate that there are a total of 75 people in this study.
So, to calculate percentages of people in certain categories, divide the number in the category by the total number of workers and then multiply this number by 100 to give the percentage value.

For example, if asked to calculate the percentage of workers who use the car to travel to work, do 45/75 and then multiply this by 100.
45/75 = 0.6
0.6 x 100 = 60
So 60% of workers travel by car.

- **D.0.3 - Estimating results**
  - ➢ Some calculations can be made simpler by estimating the results rather than using mathematical formula.
  - ➢ **Estimating** can be done by rounding results to make the calculation simpler.
  - ➢ For example, calculating 2543 x 41 is quite a complicated task. However, by rounding the figures to 2500 x 40, it is a more simple to assess by sight.
  - ➢ A number that is above 5 will be rounded up, and below will be rounded down to the next grouped value.
  - ➢ For example, 103 is rounded to 100, 150 is rounded to 200, to the nearest hundred.

D.1 - Handling Data

- **D.1.1 - Significant figures**
  - ➢ **Significant figures** are particularly important in the exam for expressing answers to mathematical calculations.
  - ➢ To give an answer to the amount of significant figures required, calculate the number you wish to represent first.
  - ➢ Then, round the number accordingly so that the amount of digits represents the sf required.
  - ➢ For example, if you calculate the mean of a set of data to be 3.657, to give this result to 3 significant figures, count the first 3 digits that are non-zero (unless they are place holders, that is, there are no non-zero digits before hand) and then round this number according to the remaining digits.
  - ➢ 3.657 to 3sf therefore is 3.66
  - ➢ 0.0046 to 1sf would be 0.005

- **D.1.2 - Calculating the mean**
  - ➢ The **mean is a measure of central tendency.**
  - ➢ It defines the average or central number from a data set, which in statistics is useful for determining the most common or expected value in that particular test.
  - ➢ To calculate the mean, add up all the values in the data set and divide this value by the total number of values in the set.
  - ➢ It is used for continuous data (non-nominal).
  - ➢ Using the following raw data set showing the height of 8 students in a class:

Name	Nick	Polly	Sam	John	Mike	Jaz	Kim	Ben
Height in cm	175	160	190	170	175	155	168	171

The average height of the students in the class can be calculated by adding each height up and dividing by 8 - the total number of values.
175+160+190+170+175+155+168+171 = 1364
1364 / 8 = 170.5
Therefore the average height in this data set is 170.5 cm.

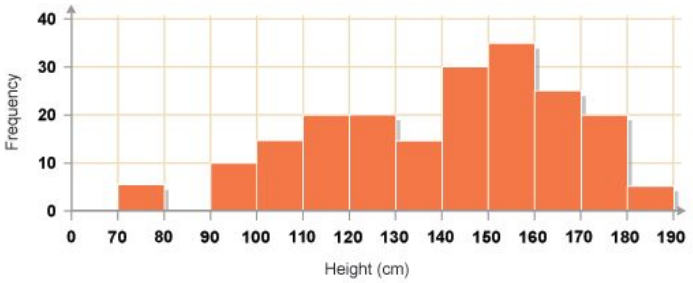
  - ➢ The mean is very **useful** for providing an average value and is **easy** to calculate, however, **it can produce values not in the original data set** and is **easily affected by outliers**, which can create skew.

- **D.1.3 - Frequency tables and diagrams, bar charts, histograms, and other graphs.**
  - ➢ A **frequency table** arranges data values, showing the frequency or amount of certain categories or groups. They are very easy to construct.
  - ➢ For example:

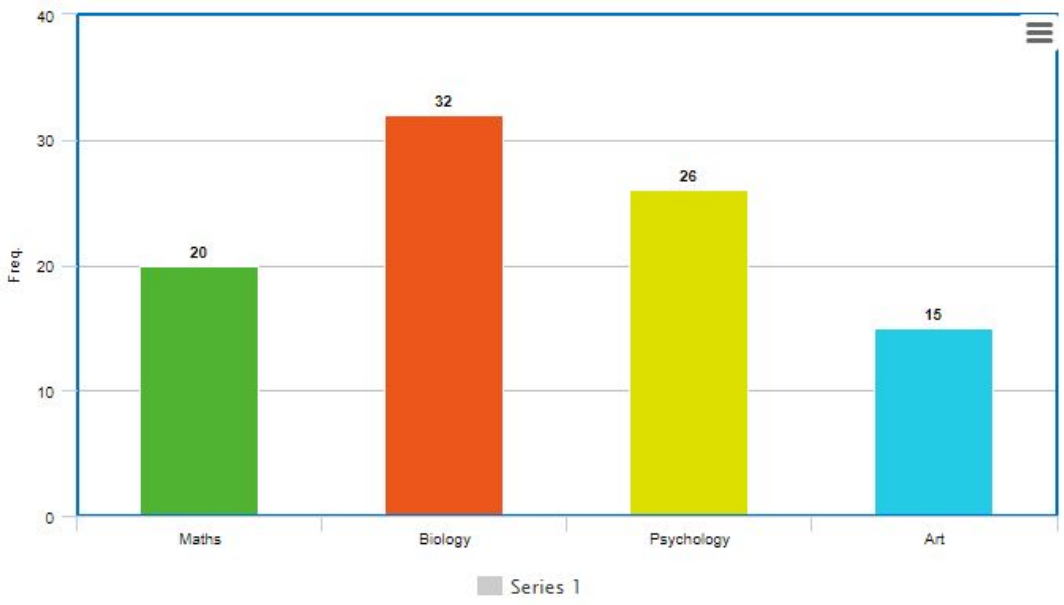
Students favourite A-Level subject	Frequency
Maths	20
Biology	32
Psychology	26
Art	15

  - ➢ A frequency table however is a visual representation of such a table, using the y axis to display the frequency of the group and the x axis to show the variable being researched (usually a continuous variable such as height or foot size).

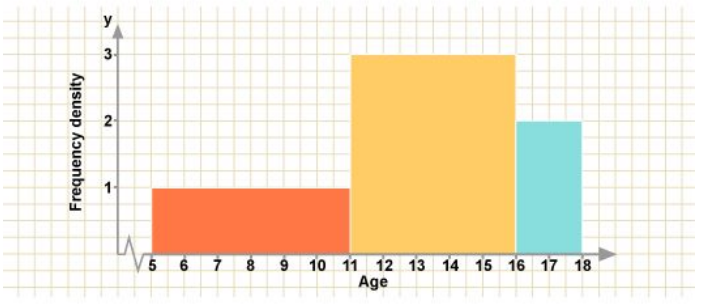
➢ For example:



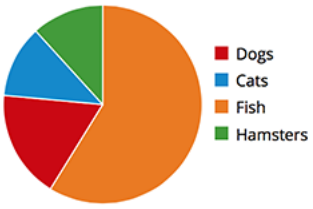
Bar charts are an easy way of representing data that is not continuous (nominal), such as in the above example where the variable is favourite subject.



➢ **Histograms** are similar to barcharts, but show frequency density on the y axis instead of the raw frequency value. This means they represent the data distribution far more accurately. The x axis is represented by an interval or group of continuous data values. The **area** of each bar, that is the height x width, is equal to the **frequency.**



- **Pie charts** use percentages to represent segments of a circle as the values associated with data groups.
- To calculate a segment of a pie chart, figure out the percentage represented by a certain group. For example, 66/100 pets sold in 1 month at a shop are fish. This is 66%.
- If you then had to draw or sketch this, you would calculate what 66% of a circle (360 degrees) is. You would then draw this onto the pie chart. (You won't be expected to draw accurate pie charts in the exam, the segments will either be large numbers like 25% or 50% so you can estimate).



- **Line graphs** display data on a graph where the data is plotted along the x and y axis.
- This method is only used for <u>continuous data</u>.
- Line graphs are similar to scatter graphs, except for on a line graph, the points plotted are all connected with a **single line.** A scatter graph however does not connect the points.

- **D.1.4 - Simple Probability**
  - **Probability** in psychology is used to show the chance of an event occuring.
  - Most importantly, it is crucial to consider the probability of an event being due to chance (gained through an experiment).
  - In psychology, researchers work to the 0.05 level of significance. This means that there must be less than a 5% chance that the results are due to chance, making the results and conclusions significant.
  - This observed value calculated in statistical tests must be either greater than or less than this value (depending on the test).
- **D.1.5 - Sampling**
  - The specification requires an understanding of the mathematical principles applied to sampling methods.
  - For example, for a **random sample**, names of people in the target population could be placed in a hat and randomly picked out. The concept of "random" means that every member of the target population has an equal chance of being picked.
  - For a **stratified sample**, the number of people in different groups are represented in accurate ratios compared to the target population.
  - If in the target population of A-Level computer science students in the UK, 20% are female and 80% are male, the sample used in the study will contain the same proportion to make the results more generalizable due to a representative sample.

- **D.1.6 - Mean, median, mode and range**

Measure	Explanation	Strengths	Weaknesses
Mean	The mean is a measure of central tendency. It shows the average or expected value in a set of data.	Easy to calculate mathematically, gives a precise value, useful for statistical tests.	Can be affected by outliers (causing skew), can give a value not originally in the data set (unrealistic).
Median	The middle value in a set of data that is organised by increasing value.	Not affected by outliers, will always give a value originally in the data set	Not as useful for other statistical test, requires sorted and ordered data.
Mode	Provides the most frequent or common value in a data set.	A useful way of calculating the average when the data is nominal	There may be multiple modes, there may not be a mode at all, the mode may not represent the spread of the data in an accurate way
Range	A measure of dispersion - the difference between the highest and lowest value in a set of data.	Easy to calculate and gives a good idea of the spread of data.	Is affected by outliers and is not useful for other statistical calculations.

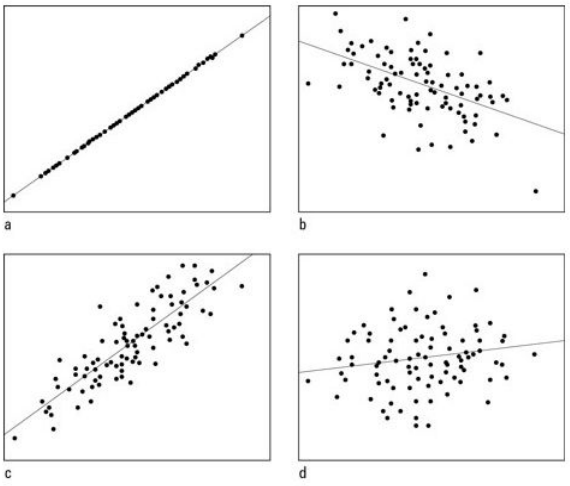
➢ Calculating the Variance:

- Calculate the mean of the data in the table
- Subtract the mean from every value in the table - this new value is called ‘d’
- Square each value of ‘d’
- Add up each value of ‘d squared’
- Divide this number by N - the number of values in the table

➢ Calculating the Standard Deviation:

- Follow the exact same steps as the variance
- At the end, square root your answer to give SD!

- **D.1.7 - Scatter graphs and correlation**
  - ➢ Scatter graphs are only used for <u>continuous data</u> and are often used to find correlations between variables as the direction and strength of the line produced can show correlation type.



(a) shows a perfect correlation. **<u>This does not mean causation!</u>** It simply means there is a definite relationship between the variables.

(b) shows a moderate negative correlation

(c) shows a strong positive correlation

(d) shows no linear relationship

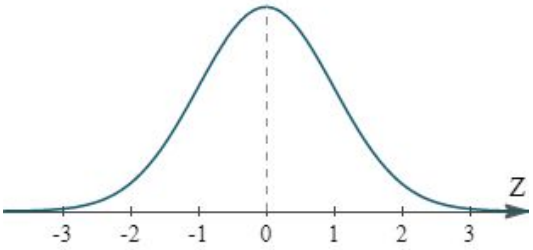
The correlation coefficient gives a value between <u>-1 and 1</u> that shows mathematically what the relationship is. <u>-1</u> is a perfect negative correlation, whilst <u>+1</u> is a perfect positive correlation. <u>0 shows no relationship</u> between the variables at all, and is displayed by (d). <u>-0.3 or +0.3</u> shows a weak relationship, whilst -0.5 or +0.5 shows a moderate. Ideally, an experiment should show a coefficient of <u>0.7 and above</u> (strong) to show a significant relationship.

A **POSITIVE** relationship means that as one variable increases in value, so does the other (e.g hours of revision and percentage score on a maths test)

A **NEGATIVE** relationship means that as one variables increases, the other decreases (e.g hours watching television and percentage score on a maths test)

  - ➢ Scatter graphs are therefore extremely useful for experiments testing for correlations.
  - ➢ They can also display non-linear (not in a straight line) relationships, which the correlation coefficient cannot show.
  - ➢ You may need to plot a scatter graph from a raw data table, but will not need to calculate pearson's correlation coefficient.

- **D.1.8 - Using statistical tests**
  - ➢ **Parametric tests** assume that the population that the data is drawn from is normally distributed.
  - ➢ You don't need to know how to conduct parametric tests for the exam, but must be aware when they are used.
  - ➢ **Non-parametric tests** are used when this assumption cannot be made. In psychological experiments, these tests are often used due to difficulty in gaining normally distributed populations.

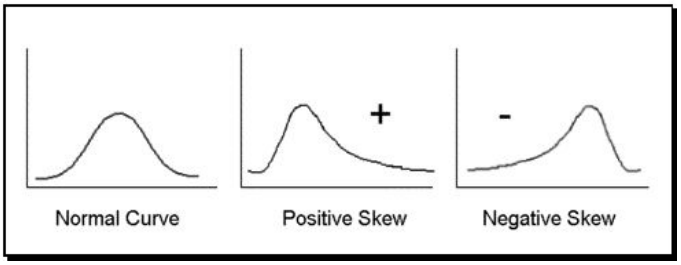


Normal distribution refers to the bell-curve shown below in which the mean of the data is central, with all other data creating the shape displayed.

- ➢ The non-parametric tests you need to learn for the exam are **Chi-Squared**, **Binomial Signs Test**, **Mann Whitney U**, **Wilcoxon Rank Sum Test** and **Spearman's Rho.**
- ➢ You will be expected to explain how to conduct each test, know how to use the observed and critical values to calculate significance and sub in given values to the formula (which would be provided in the question)
- ➢ If significance is found, then the alternate hypothesis can be accepted and the null rejected.

- **D.1.9 - Order of magnitude**
  - ➢ An order of magnitude is the scale of which an amount differs from another amount.
  - ➢ For example, if a value is 2 orders of magnitude bigger than another value, it is around 100 times larger.
  - ➢ They are useful for making approximate comparisons for very large differences - for example, if a very large group of participants was involved in a study.
- **D.1.10 - Levels of measurement**
  - ➢ **Nominal data** is used to label or describe variables, but have no quantitative value.
  - ➢ For example, hair colour is nominal data as the categories can be named with descriptors such as brown, blonde, ginger.
  - ➢ This is a form of qualitative data.
  - ➢ **Ordinal data** is used to quantify variables, but the difference between each value is not known or quantifiable.
  - ➢ For example, if participants had to rank on a scale of 1 to 10 how much they enjoy school, it is easy to see that the data is quantitative and that 10 is better than 1, however, there is no quantifiable scale between 1, 2, 3 ect.
  - ➢ **Interval data** is quantitative data in which the differences or scale between the values is precise and measurable. For example, time, temperature or distance.
- **D.1.11 - Normal and skewed distributions**
  - ➢ **Normal distribution** is shown by a symmetrical bell curve, in which the mean, median and mode are all the same value.
  - ➢ **Negative skew** indicates that the mean < median < mode and **positive skew** indicates that the mode < median < mean.

- You may have to identify skew either from a graph as displayed below or from the values themselves (which you may be asked to calculate first).



- **D.1.12 - Choosing statistical tests**
  - The easiest way to remember which statistical test to use for a given experiment is to use a table:

	Independent Measures	Repeated Measures	Correlation
Nominal	Chi Squared	Binomial Sign Test	x
Ordinal/Interval	Mann Whitney U	Wilcoxon	Spearman's Rho

  - A useful way of remembering the order of the tests in the table is with the saying ***"Crispy Bacon (Can't) Make Will Sick"***
  - Each letter corresponds to the test in the table, going right from the top left.
  - You cannot perform a non-parametric statistical test of correlation on nominal data, so this cell is left blank.
  - This method is very useful for being able to assess an experiment in the exam for which test needs to be applied!
  - When explaining why you would choose this test, refer to the column/row you have chosen. For example, "the experiment uses repeated measures design and collected interval data, therefore I would conduct the Wilcoxon statistical test".
- **D.1.13 - Using statistical tables to determine significance**
  - Each of the five statistical tests that you must learn will have its own table of values to determine whether the results are significant.
  - You must know how to **calculate the observed value**, and then use the table to find the critical value which will determine whether the alternate or null hypothesis should be accepted.

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09

- For example, the above table is an extract from the Chi-Square significance table. You would calculate the degrees of freedom by multiplying the number of (rows x columns) in the experiment data table.
- You will not be expected to learn or memorise the values in these tables, however, it is important to practice using the tables as this can come up in exam questions.

- **D.1.14 - Measures of dispersion**
  - The measures of dispersion required by the specification are range, interquartile range and the standard deviation. Data must be ordinal/interval/ratio.
  - Unlike the measures of central tendency, these measures give a value that identifies the spread of the data.
  - The **range** is the difference between the highest and lowest value. It is easy to calculate but does not give an indication of the type of spread within these values.
  - The **interquartile range** is the difference between the upper and lower quartiles (shows the spread within data once separated into 4 equal groups).
  - The **standard deviation** shows the average difference between the data values and the mean. The standard deviation can therefore indicate the overall dispersion as a large standard deviation highlights that each value varies quite significantly from the mean. The standard deviation is a better measure than the range if there are outlying scores as it considers all the items in the distributions and is less affected by anomalous data.

- **D.1.15 - Qualitative and quantitative data**
  - **Quantitative data** is data that is measured and expressed numerically, such as a value, amount or quantity. It is usually associated with units.
  - **Qualitative data** describes the attributes and qualities of data rather than defining it with measurements. This could include hair colour, name, job title etc.
  - **Quantitative data** is extremely useful for using directly with statistical tests and usually involves a very structured form of data collection, for example, measurements of temperature.
  - **Qualitative data** cannot be directly used in statistical tests, but is extremely useful at providing depth of information and can therefore develop a more meaningful understanding.
  - Data types can be translated into other types to make them more useful.
  - For example, a qualitative data type such as an interview transcript can be translated into quantitative data by categorising responses and recording frequency of words.
  - This is useful in experiments for providing both depth and statistical, objective evidence, but is much more time consuming and therefore will generally involve a smaller sample.

- **D.1.16 - Primary and secondary data**
  - ➢ **Primary data** is data collected directly and originally from the researchers themselves with the purpose of using the data in the experiment itself.
  - ➢ **Secondary data** is research data that was collected by a third party at a previous time, possibly for use in another experiment, but that is still useful for the research purpose.
  - ➢ A researcher dealing directly with participants is collecting primary data.
  - ➢ A researcher that uses brain scans of psychiatric patients at a hospital and does not interact with the participants themselves is using secondary data for their own research.
  - ➢ Primary data is beneficial as the researcher can ensure the collect the exact data they need for the research, however, it is a more time consuming and expensive process.

D.2 - Algebra

- **D.2.1 - Understanding mathematical symbols**

Symbol	Meaning/Use
=	Equal to
<	Less than: 23 < 56 (Value on the left is smaller than the right)
<<	Much less than
>>	Much more than
>	More than
∝	Proportional to (as the left number, the right number increases in proportion)
~	Roughly equal to
Σ	Sum of (appears in statistical equations)
$\bar{x}$	The mean of all x values

- **D.2.2 - Algebraic equations and units**
  - ➢ The exam may require you to substitute numerical values into equations, most likely using formulas for statistical tests.

$$SD = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

For example, here N represents the total number of values, x represents each value - you may be asked to sub given values in to find the SD.

- **D.2.3 - Solving simple algebraic equations**
  - ➢ The equation skills required here are similar to algebra you will have done at GCSE level maths.
  - ➢ Multiplication, addition, subtraction and division are required and could be given in a number of formats, including statistical tests.

D.3 - Graphs

- **D.3.1/D.3.2 - Translate information between graphical, numerical and algebraic forms and sketching graphs**
  - ➢ **Graphical form**: data represented on a graph, such as line graph, pie chart or bar graph.
  - ➢ **Numerical form**: raw data as numbers in a table
  - ➢ **Algebraic form**: a way of presenting data so that values can be substituted in to give an answer.
  - ➢ You need to be familiar with how to construct different types of graphs (described above) from these different data forms.
  - ➢ Always ensure to provide:
    - A name for the graph
    - Labels on the X and Y axis
    - Accurate data points