Introduction to Research Methods for Psychology

Dr Sheila Thomas

INTRODUCTION
TO
RESEARCH METHODS
FOR
PSYCHOLOGY

CONTENTS

Welcome to Psychology!

You will find Psychology to be a great subject, but to get the most out of it, you do need to get on top of a few basic skills. Research Methods is the foundation of everything in Psychology and this booklet is designed to give you that foundation. Your future studies will build on this, so it is really worthwhile spending some time familiarising yourself with these basic concepts.

This book has been written for students who have not studied Psychology before and who quickly need to learn about how research is carried out. Once you have read through each section, it would be a good idea to have a look at the practice questions at the back of the book which come with answers so you can see how you are doing.

I hope you will find this booklet a helpful start to your studies.

Enjoy the rest of your course!
Dr Sheila Thomas

Introduction: Psychology is the study of mind, brain and behaviour in human and non-human animals. Some view psychology as a science and there are certainly some aspects that are scientific, others are less open to scientific investigation.

Psychologists who regard the subject as a science:

✓ Look at **little bits of behaviour**

✓ Usually in a **laboratory**

✓ They look for **nomothetic laws** – this is generalised or universal laws of behaviour which can be applied to all similar people in similar situations

✓ They use **quantitative data** (numbers) as they can put them into statistical tests which can tell us if we have found significant results

✓ Because they look for nomothetic laws, they are assuming that similar people will behave in similar ways in similar circumstances so **behaviour must largely be controlled (determined) by certain factors.**

Psychologists who do not regard psychology as a science:

✓ Feel that looking at little bits of behaviour does not give an accurate picture of what it is to be human.

✓ They believe that humans in labs behave differently to how they behave in the real world.

✓ They feel it is more appropriate to look for **ideographic laws** (laws unique to each of us) since we are all unique

✓ They prefer to use **qualitative data** (descriptive data) since they are not concerned with statistics and they believe qualitative data gives a truer picture of what it is to be human.

✓ They do not believe that behaviour is determined but that we can have a high degree of free will.

THE BASIS OF SCIENTIFIC METHOD

All sciences share one basic feature; they aim to discover facts about the world by using systematic and objective (producing clear facts that are not open to debate) methods of investigation.

It is called `systematic' because there are several stages in this;

- We observe a feature in the world e.g. people who drink alcohol in excess tend to lose their balance

- This leads us to formulate a theory about it e.g. that alcohol has an effect on the central nervous system

- We then use this theory to generate a clear statement that can be tested in order to support or refute (question) our theory. This statement is called a **research hypothesis**

- We design our research to test our hypothesis (e.g. that drinking alcohol has a detrimental effect on balance)

- We carry out our test then put our results into a statistical test which will tell us whether alcohol is having a significant effect on balance. If we find we do not have a significant effect, we alter our theory.

Experiments are thought to be the most scientific since they involve the most **systematic and objective** way of testing hypotheses. In an experiment, all **variables are controlled** so hopefully we eliminate any variable which could bias our results. Experiments tend to use **quantitative data** which allows data to be **analysed statistically** which will calculate how certain we can be that one variable had this effect on the other variable.

Features of Science

Over the years various theorists have put forward ideas about what makes something a science. Very famously Einstein said "No amount of experimentation can ever prove me right; a single experiment can prove me wrong."

The key feature of science is that it deals with what can be observed and measured, not with beliefs or ideas. The following aspects have come to be regarded as being essential features of science:

1. Replicability - i.e. you should be able to repeat an experiment and come up with the same result.

2. Objectivity – the measurements taken must be completely without bias.

3. Falsification – This is a very important aspect on which Karl Popper wrote a theory. He said that we should try to disprove a scientific finding and if we cannot disprove it then it must be true.

People sometimes draw a distinction between an **inductive** approach which is when we don't know what we will find because we don't know much about the research area and so we cannot formulate a hypothesis (very typical in qualitative data analysis) and the **deductive** approach which is when we formulate a hypothesis and then test it to see if it is true or not. Both approaches are used in psychology.

It is possible to make non-experimental methods systematic and objective in their testing of hypotheses but in non-experimental methods variables can creep in which **can bias our results** so we cannot be so certain we have found facts which can be proven e.g. if you are interviewed by someone you could decide not to tell the whole truth and that would bias the interviewer's data.

One of the first things we need to decide when doing research is which method we will use. This decision should be based on several factors:

• **Ideological standpoint** – there are different views in psychology about whether it can be regarded as a science. Those who do regard it as a science will tend to use experimental method since it is the most rigorous at establishing cause and effect. Those who believe psychology is not a science because we are all unique (so why look for universal laws?) and that we have free will, tend towards more qualitative methods which describe, rather than quantify, behaviour.

• **The nature of the research question**. Psychology is the study of mind, brain and behaviour and **different methods can be appropriate for each of these** e.g. it would be inappropriate to ask people which parts of their brain are working at particular times – brain scanning techniques would be more appropriate.   We also need to take **ethical concerns** into account since it would be unethical to lock your granny in the cellar to carry out an experiment – it would be ethical to study people who have found themselves in solitary confinement, however, this may represent only a small biased sample of people.
The research question itself can determine which method is used for e.g. if we are studying children over time then we will need to decide between a **longitudinal design** (where we follow the same children over time) or a **cross-sectional design** (where we study children of different ages at the same time.)

• **The stage of the research** – if you are just beginning to plan a research project and know little about the area, it would be unwise to go ahead with a large scale experiment as your results may turn out less than reliable. It would be more sensible to carry out a small-scale **pilot study** first.

i.   Pilot studies test whether it is feasible to carry out a piece of research
ii. it should highlight any problems with your test,
iii. whether there are any sources of bias
iv. and whether your results are valid.

In the early stages of the research, psychologists sometimes use qualitative methods (e.g. interview people) in order to gain insights into what variables are involved in the area.

Cost and time – longitudinal studies will take longer than cross-sectional ones and time is money, as the saying goes. Large grants to carry out research means that more people can be employed which will shorten the amount of time required.

In psychology we take samples from larger populations as we cannot study the whole of the population. In order to be able to draw valid conclusions from the study, the sample needs to be **representative** of the larger population.

The **target population** is the wider population that the researcher hopes to generalise the results to.

Generally, the larger the sample, the better the study is as this will probably eliminate **inter-participant confounding variables** if we are using an independent design or just a biased sample if we are not. However, studies involving more P.s are more costly and time consuming.

> **There are several methods psychologists use to sample P.s**

Random sampling: the most well known method. It means that every member of the target population has an equal chance of being chosen for the sample. A list of all the people in the target population is obtained. In order to chose the correct number of P.s randomly you could put the names of all the people onto individual strips of paper, put them into a container and then pull out the required number. Alternatively, you could put numbers by the names of those in the target population and use a random numbers table to chose which of those numbers will be in the actual sample.

Evaluation of random sampling:
This is a relatively easy method to use which can produce a representative sample.

It is seldom that random sampling will produce a truly random sample e.g. you cannot get a full list of people in an area as the homeless and travellers are not listed. Also, the people chosen may then decide not to take part which means the researcher may be left with a biased sample.

Opportunity sampling: this involves selecting P.s on the basis of their availability. This method often involves using university students (in the USA students had to take part to get maximum marks).

Evaluation of opportunity sampling:
This method is the easiest to use but generally it does not produce anything like a representative sample. For e.g. student populations will be biased in terms of age, education, class etc.

Volunteer sampling: this is also known as a self-selecting sample since P.s chose to take part. Adverts can be put in papers, dropped through doors etc.

Evaluation of volunteer sampling:
This is an easy method to use, however, the sample is likely to be biased as volunteers may be more motivated to perform than non-volunteers and have certain psychological features which set them apart from non-volunteers eg more likely to be extrovert.

Stratified sample: this is the best way to obtain a representative sample. A list of factors which might affect the study are listed and the percentage of people in the target population possessing these factors is calculated. The sample is obtained randomly from the target population but in the proportions defined by the factors e.g. if the target population is school children, we would expect 2.5% to have an IQ in excess of 130 so 2.5% of our sample would similarly have an IQ in excess of 130.

Evaluation of stratified sampling:
Although this can produce a representative sample, it is very time consuming and difficult to obtain a sample in this way.

Independent variables – are what the researcher is **studying the effects of.**
It is what is altered or **manipulated** in a study (the difference between the conditions).
It is **the cause** of our test results.
e.g. if we are studying the effects of alcohol on balance, alcohol is the independent variable (I.V.)

Dependent variables – are the **result or score** in a study.
The precise score on the dependent variable is dependent on how the researcher manipulated the independent variable.
If the I.V. is the cause, the **D.V. is the result**
e.g. if we are studying the effects of alcohol on balance, the results of the balance test are the D.V.

Confounding/extraneous variables – are any variables in a study which could lead to biased, inaccurate or false conclusions. We try to control or eliminate confounding variables.

For example, in the study of the effects of alcohol on balance, if we gave

no alcohol to a group of participants (P.s) with coordination problems and alcohol to a group of heavy drinking, tight-rope walkers and then tested them both on balance, we might find that the tight-rope walkers did better on the balance test which would lead us to conclude that alcohol improves balance, which is clearly not the case.

Differences between P.s in each condition can cause confounding variables (inter-participant confounding variables) but confounding variables can come from environmental factors (e.g. more noise or heat in one condition than the other) or when P.s see the purpose of the experiment and then change how they behave (known as demand characteristics).

Experimental condition – is the condition that receives the I.V. E.g. in the study on alcohol and balance, the experimental condition would receive the alcohol.

Control condition – is the group or condition that does not receive the I.V. so it is the baseline from which we measure the effects of the I.V.

Standardisation – means to make standard or the same. The only difference between our control and experimental group should be the I.V. so we make our instructions to both groups the same (standardising instructions) we use the same tone of voice and we control all situational variables (e.g. heat, time of day etc.) to avoid introducing confounding variables.

Randomisation – means to make random or leave order to chance. P.s may be randomly allocated to groups which means that no one P. has any more change of being in one group than another – this is to avoid bias caused by inter-participant confounding variables. P. can be randomly allocated to groups by putting all their names onto pieces of paper, putting them into a hat then picking out half the names - they then go into one group, the remaining ones going into the other group. Alternatively, you can use random number tables to allocate P.s randomly to groups.

Hypothesis – this is a statement which is tested, it states that this I.V. will have this type of effect on this D.V. There are two types

Experimental hypothesis states that the I.V. will have a significant effect on the D.V. e.g. ingestion of alcohol will have a significantly detrimental effect on balance.

Null hypothesis states that the I.V. will not have a significant effect on the D.V. e.g. ingestion of alcohol will not have a significant effect on balance.

In addition, there are 2 types of experimental hypotheses:

Directional/one tailed hypothesis states the direction of the effect of the I.V. on the D.V. e.g. alcohol will have a significantly detrimental effect on balance. (a one-tailed fish swims in one direction)

SOME IMPORTANT RESEARCH TERMS

Non-directional/two-tailed hypothesis does not state the direction of the effect of the I.V. on the D.V. e.g. alcohol will have a significant effect on balance (this doesn't say whether it will improve or decrease it). It is usually used in a pilot study when little is known about the variable being studied. (a two-tailed fish tries to swim in both directions).

Operationalising variables this is concerned with how we get our variable into a form where we can study and measure it. e.g. we could operationalise prejudice by using an interview or questionnaire, we could also observe people's behaviour when dealing with people from ethnic minorities.

Demand characteristics – this is concerned with any factor in a study which encourages (or demands) that P.s behave in a way that they would not have done otherwise. They are to do with the interaction between the P. and the researcher and involve:

I've done what I thought you wanted me to do!

When P.s guess the purpose of the experiment and try to please the researcher by providing data which fits with what the researcher had hoped would happen.

When P.s guess the purpose of the experiment and go against what the researcher had hoped would happen – known as the `screw you effect'.

P.s are scared in case they appear `abnormal' so they behave unnaturally.

P.s behave unnaturally in order to look good (**social desirability effect**)

To reduce demand characteristics you could use a **single blind procedure** where P.s don't know which condition they are in or you could use a **double blind procedure** where neither the researcher nor the P. know which condition they are in. Double blind procedure is used in drug trials where P.s are given a code number, this is looked up on sheet which also says the code number of the pot/packet of medication (they all look alike) the P. is then given the appropriate medication. The person who worked out the codes does not hand out the medication so neither the person handing out the medication or the P. know if they are getting the real drug or the placebo (a non-active fake drug)

Reliability – concerns whether a test consistently measures the variable it is designed to measure, much as a reliable car consistently does what it is designed to do.   e.g. if an I.Q. test gets really different results with the same people on different days, the test does not seem to be reliable.

Types of reliability:

Inter-rater reliability – this is concerned with how much agreement there is between a number of researchers who are scoring on the same test. If researchers are scoring on an I.Q. test with clear-cut answers, you would expect inter-rater reliability to be high. Inter-rater reliability may be more of an issue when there is a subjective element to the scoring (e.g. when counting rough and tumble play in a school playground).

   To establish inter-rater reliability you show the scorers the same behaviour etc. then compare how they have rated/scored it. If there is a difference in their scoring you need to observe the video again, this time setting the criteria for the categories/scoring. You then watch another video of the behaviour etc. being studied and compare their results again – they should now be very similar.

Internal reliability is another aspect of reliability but this time it relates to whether each of the items or parts of a test contributes the same amount to the score – that is, the test is consistent within itself. e.g. if some items on an I.Q. test or questionnaire score more than other items then the test lacks internal reliability.

   To establish internal reliability– ensure that each item on a test is of the same level of difficulty and has the same scoring as other items.

External reliability is concerned with whether a test measures consistently over time e.g. if a P. scores 112 on an I.Q. test you expect them to get a similar result is they do the test again.

   To establish external reliability you can use the **test-retest method** where you test P. then test them some time later on the same test (hoping they don't remember what they did in the first test) and if the test has external reliability then the scores will be very similar (i.e. they will show a strong positive correlation).

Validity – concerns whether a test measures what it was designed to measure and nothing else.   e.g. if an I.Q. test had items on it which drew on knowledge of instruments in an orchestra, then it could hardly be thought of as a valid test of I.Q. Real IQ tests tend to use puzzles for which you don't need previous knowledge.

Types of validity:

Internal validity – this is concerned with whether confounding variables have influenced a result so it is about the control within a study. Because experiments are highly controlled, they are normally thought to have higher internal validity than non-experimental methods.

To establish internal validity – ensure that all possible confounding variables have been controlled.

Face/content validity – this is concerned with, whether on the `face of it' the **study, test** etc. fits with what we know about the variable. e.g. if an I.Q. test had questions about political beliefs it would lack face validity since there is no relationship between political beliefs and I.Q.

To establish face validity find out as much as you can about the variables you are studying then ensure that your test is measuring that and nothing else.

Construct validity is whether the **results of the test** etc. fits with what we know about the variable being studied. e.g. the majority of the population has an I.Q. of around 100. If a new I.Q. test finds that most of their P.s have I.Q.s in excess of 120, it is said to lack construct validity.

To establish construct validity, compare your results on a variable with other tests which have been proven to test that variable accurately e.g. other I.Q. tests.

External validity is the extent to which the findings of a piece of research can be generalised to different populations, settings and conditions. There are several types:

Population validity concerns the extent to which the findings of the research could be **generalised to other groups** of people. Small, biased samples tend to lack population validity.

To establish population validity use a large sample of people that have been carefully selected to include all the different ages, genders and types of people that are present in the wider society. That way the study could be generalised to the wider population.

Ecological validity concerns the extent to which results could be **generalised to real life situations.** Results that are obtained in a lab sometimes cannot be generalised to what would happen in the real world since the laboratory is a very artificial setting and this can influence the behaviour of P.s. An aspect of ecological validity is **mundane realism** which is about how close the test etc. is to real life e.g. if a test gets P.s to remember nonsense syllables then it lacks mundane realism since we don't tend to do this in real life.

To establish ecological validity carry out the research in the real world, however this can lower the internal validity of a study since it is difficult to control confounding variables in the real world. (So naturalistic studies are high in ecological validity but low in internal validity).

To establish mundane realism you need to design a test which is close to what we need to do in real life. e.g. show a staged event rather than a video of an incident.

ETHICAL ISSUES

Psychology is governed by the BPS Code of Ethics (British Psychological Society) and all research must conform to these guidelines.

1. Informed Consent - Participants must give their written informed consent to take part in the research.

2. Avoidance of deception – Ideally you should not deceive your participants but there is a danger of demand characteristics of they know the purpose of the research.

3. Protection from Harm and Distress – you must not place your participants in a situation in which they could suffer harm or distress.

4. Right to Withdraw – you must give your participants the right to withdraw from the research at any time and have their data destroyed.

5. Anonymity – you must not refer to participants' names.

6. Confidentiality – All data must be treated with the utmost confidentiality.

7. Debrief – you must tell your participants the nature of the research at the end of the study.

8. In observations carried out in a public area where people would normally expect to be observed by others e.g. on public transport, you do not need to get informed consent or give a right to withdraw or debrief but you cannot invade someone' privacy e.g. observing behaviour in your neighbour's back garden, or watching people in a public toilet.

Not all psychological studies have conformed to these guidelines, in fact some of the most famous studies have broken lots of ethical guidelines but in modern day research, an ethics committee would be very rigorous in vetting any proposed study.

RESEARCH METHODS AND TECHNIQUES

There are a number of different methods that are used in psychology. The methods can be regarded as tools and different ones may suit different purposes. **Quantitative methods** use statistics to measure the effect of the independent variable (what we are studying the effects of) on the dependent variable (the score or results). **Qualitative methods** are more descriptive. We can also use **primary data** which is data that the researcher gathers themselves or we can use **secondary data** which is data that has already been gathered by somebody else.   An example of using secondary data would be a **meta-analysis** which is when you look at all the research that has already been done in a particular area and draw conclusions on what everyone else has found.

EXPERIMENTS

Experimental method

A true experiment has several key features:

1. There is random assignment of the P.s to the different conditions.

2. The experimenter has a good deal of control over the experimental situation so that all other variables which might influence the results are controlled or eliminated.

3. Because of the control/elimination of confounding variables and the manipulation of the I.V., we can infer causality (cause and effect) that is that the particular I.V. has had a significant effect on the particular D.V.

Laboratory experiments

The key feature is that the experiment is carried out in a controlled environment. This need not be an actual laboratory, just a controlled environment. Lab expts conform with the 3 criteria above.

Advantages of laboratory experiments
• Since there is a high degree of control over variables, this method is the most **rigorous at showing cause and effect** between variables

• Because of the control/elimination of possible confounding variables it is said that lab expts are high in **internal validity.**

• Lab expts can be repeated by other researchers so they are **replicable** – this is an important part of experimental method

• You can use **technical equipment** in the lab which you could not realistically use outside the lab

Disadvantages of laboratory experiments
• Because of the high level of control and the strangeness of the lab situation, it may not be wise to generalise from the findings to what might happen in real life so lab expts are said to **lack ecological validity.**

• The types of tasks people are given in lab expts may be unlike those they experience in real life so lab expts can **lack mundane realism**

• Lab expts can have **demand characteristics** which is when the experiment encourages P.s to behave in a certain way that they would not ordinary do. It includes when P.s see the purpose of the expt. and conform to expectations or when they go against them (known as the `screw you' effect.

• Lab expts can contain **experimenter bias** which is when the experimenter advertently or inadvertently brings about an effect. Tone of voice and many other features of the experimenter can influence results.

Field experiments

These are expts that are carried out in the natural environment. The independent variable can be manipulated, as in a lab, but people do not generally know they are taking part in a study. An e.g. of a field expt. is Piliavin et al (1969) who looked at helping behaviour on the subway.

Comparing this method to a true experiment, field expt. do not randomly assign P.s to groups, there is manipulation of the I.V. but there is less control over possible confounding variables so we need to be more careful when inferring causality.

Advantages of field experiments
• **Demand characteristics**, such as P.s knowing the aim of the expt, do not occur since, in general, P.s do not know they are taking part.

• A field expt has **higher ecological validity** since it takes place in real-world environments. This means we can generalise from our findings to what happens in real life.

• Since there is manipulation of the I.V. **causality can still be inferred**

Disadvantages of field experiments
• It is **difficult to replicate** field expt. because it is difficult/impossible to create exactly the same situation again.

• There is less control over possible confounding variables than in a lab so this type of study may be **lower in terms of internal validity**

• Because there is no random allocation of P.s into groups there maybe a **sampling bias** which may affect results.

• There are serious **ethical concerns** in this type of study since P.s do not generally know they are taking part so there is no consent, they do not know they have the right of withdrawal and there is no debriefing.

• Field expt may be more **time consuming and expensive** to carry out.

Natural experiments

These are experiments where we make use of a naturally occurring event for research purposes. It is sometimes called a quasi-experiment. An e.g. of one is Hodges and Tizard's (1989) study of teenagers raised in orphanages.

Compared this to a true experiment, in a natural expt. there is no random allocation to groups, there is no manipulation of the I.V. by the experimenter and there is less control over possible confounding variables so we need to be careful when we infer causality.

Advantages of natural experiments
• It is possible to do research when it is **unethical to manipulate the I.V**

• As they are usually carried out in the field they are less open to the criticisms levelled at lab expts; they are **higher in ecological validity** and have no **demand characteristics**

• Because we are not manipulating the I.V. there is less chance of us offending **ethical issues** such as those regarding causing distress.

Disadvantages of natural experiments
• We may have to wait a long time to **replicate** natural experiments

• As P.s are not randomly assigned to groups there may be a **sample bias**

• As the I.V. is not manipulated we need to be **careful when inferring causality** .

• Because confounding variables cannot be eliminated this means that natural expts may **lack internal validity.**

Types of experimental design

1.   **Independent measures/groups design** – this is where you have different P.s in the control and experimental condition. This means that P.s are only tested once.

Advantages –
    • it is a good method to use when performing in one condition could lead to P.s realising what the study was about and so alter their behaviour.

    • it is also a good method to use when there could be a **practice effect** – that is performing in one condition could lead to the P.s speeding up or slowing down in the next condition – called a **practice or fatigue effect** which can occur especially when the test is quite long.

Disadvantages –
• you need a relatively large group in order to test for statistical significance.

• also, even if you have used a random sampling technique, there still could be important differences between the P.s in each group and these could act as inter-participant confounding.

• variables and bias results.

2. **Repeated measures/groups design** – this is when P.s are in both the control and experimental condition so they are tested at least twice.

Advantages –
• it eliminates the possibility of inter-participant confounding variables since each person is in each group so they act as their own baseline.

• You need far fewer P.s in order to test for significance using a statistical test.

Disadvantages –
• this type of design can lead to practice and fatigue effects (order effects). The way to **overcome practice and fatigue effects is to counterbalance for order effects** – you split your P.s into 2 halves – half do condition 1 then condition 2, the other half do condition 2 then condition 1. Any getting slower or faster between the first test and second is therefore shared between the 2 conditions.

• Also, performance on one condition can lead to P.s seeing the point of the study so they can either comply with what is expected or go against what the experimenter expects (demand characteristics).

3. **Matched pairs/groups design** – this is where you decide what possible inter-participant confounding variables there could be in your study (e.g. in a study of alcohol on balance, possible variables would be gender, age, weight, balance, tolerance to alcohol etc.) you then test all your P.s on each of these variables. You then look for pairs that are similar on all the variables then toss a coin to randomly allocate them to the control or experimental group. With this method you have made your 2 groups as alike as possible, also, each P.s is only tested once.

Advantages –
• because you have made the groups as similar as possible, possible inter-participant confounding variables have been eliminated.

• This also means that you only need the same number of people as for a repeated measures design.

• There are no order effects as each P. is only tested once.

Disadvantages –
• the process of matching P.s can take quite a bit of time,

• also, it assumes that you know what the possible inter-participant confounding variables are for you to control and this may not be so.

OBSERVATIONS

Observation can form part of an experiment e.g. Milgram's study of obedience or it can be used as a method on its own.

Usually observational studies take place in naturalistic environments so the behaviour observed is natural and free from the demand characteristics possible in a lab.

There are 2 main types of observation:

Participant observation – this involves the researcher taking part in the observed behaviour (eg Rosenhan, 1973, `On being sane in insane places'). The researcher may tell those being observed that they are being observed and why; **overt participant observation** or they can chose not to tell those being observed; **covert participant observation.**

Non-participant observation – this is where the researcher does not take part in the behaviour but watches from a distance.

The line between participant and non-participant observation may not be clear since it may be unwise to take part in certain behaviours e.g. criminality.

There are 2 main techniques used in observational research:

1. **Event sampling**- Counting the number of times a specified behaviour takes place.

2. **Time sampling** – Counting the number of times a specified behaviour takes place at certain times e.g. every 2 minutes.

Example

Let's imagine I decided to observe "violent play" in a primary school playground. I would define what I meant by "violent play" as specific behaviours e.g. one child pushing another child over and the child on the receiving end falling over, looking unhappy. This would make sure that I don't measure "playful pushing" as "violent behaviour". In event sampling I would count how many times I saw this behaviour in a specific time e.g. half an hour. In time sampling, I would look away and every 2 minutes, for example, I would observe the children and see how many   incidences of " violent play" I could observe.

Advantages/ Disadvantages - In event sampling, this techniques makes the observation more manageable as the observer couldn't possibly record every behaviour they see. On the other hand, there may be interesting behaviours which are not recorded (in this example there could be verbal violence such as swearing which is not recorded). In time sampling, this technique also makes the observation more manageable as the observer is only taking a "snapshot" observation every 2 minutes, but interesting behaviours may be lost during the time when the observer is not watching the participants.

Advantages of observational research

•   **High external validity** – since the behaviour is taking place in a natural environment, P.s will tend to behave how they normally would. In participant observation the researcher needs to be very careful not to influence the behaviour as then external validity would decline.

•   **Practical** – it can be used when manipulation of an IV would be impractical or unethical (e.g. riots). It can be used when the P.s would not normally wish to co-operate (e.g. looking for discriminatory behaviour in the police). It is also particularly useful for studying animals and children.

•   **Few demand characteristics** – when P.s are unaware of being observed there will be no demand characteristics. Even when they are aware of being watched, they usually soon get used to the presence of the researcher and behave naturally.

•   **Videos can be used over long periods** – cameras can be left in place so the researcher need not be present for the whole time.

Disadvantages of observational research

Cause and effect cannot be inferred since variables are not manipulated and possible confounding variables excluded or controlled.

Observer bias – if the observers know the purpose of the study they may interpret the behaviour in line with this. If you have more than one observer you need to check for **inter-rater reliability** by showing them the same behaviour and asking them to record it, if it is not the same to that of other observers they need to set their criteria for logging down the behaviour then show another piece of behaviour and they now should be recording it identically. You also need to check for **intra-rater reliability** where an observer may record the same behaviour differently as time goes on.

Replication – is difficult to achieve since you cannot be sure the same conditions have been replicated since there is no control over variables.

Ethics – if P.s are unaware they are being watched there are issues of invasion of privacy, no informed consent given, no right to withdraw offered etc. If P.s know they are being watched then there may be demand characteristics.

Practical problems – it may be difficult to categorise behaviour correctly, participant observation may cause the researcher to get so `sucked in' that they cannot be objective – they might even `go native'. Sometimes the researcher will be noticed even when they are covert and in some situations this might be dangerous (e.g. if studying a gang).

Design of naturalistic observations

One problem with naturalistic observations is that it may be difficult to ascertain which is the best way of recording data. It could be videoed or written down in some form. Data could simply be described or, to obtain a more quantitative method, **behavioural categories** can be used.

Behavioural categories – when this method is used, observers have a grid or coding sheet on which they record the studied behaviour. Behavioural categories are chosen according to the subject of the study. For e.g. in a study of football supporter's aggressive behaviour you might have categories such as `aggressive shouting or singing', `aggressive hand or facial gestures' and then various levels of physical violence. When you observe one of these behaviours in the person/people you are observing then you will put a mark in the appropriate box. You could assign a aggression value to each of the behaviours, starting perhaps with a score of `1' for slight verbal or non-verbal (gestural) aggression up to `10' for beating someone up.

The diary method is another technique used in observational studies. Here P.s complete a diary – it provides qualitative data in the P.s own words so this can be a rich source of data. One weakness is that P.s find it difficult to complete for long periods. The material can then be turned into quantitative data if necessary though this can be subjective.

SELF-REPORT TECHNIQUES

Questionnaires

This is a method of data collection where the respondents (the P.s in a questionnaire) give answers to a list of pre-set questions on a topic or a number of topics. Variables can be studied e.g. the hypothesis that older people tend to be more racially prejudiced, could be tested.

Questionnaires can produce quantitative data – this is through the use of **closed or fixed questions** where respondents answer `yes' or `no' or tick a box which quantifies the degree they agree/disagree with a statement etc. This quantitative data can be statistically analysed to see how significant it is. Closed or fixed answers are easy to quantify but they restrict respondent's answers.

Questionnaires can produce qualitative data – this is through the use of **open questions** where respondents are asked to respond to a question in their own words. This is more difficult to analyse but it produces data with much more depth and it does not restrict the answers of the respondents.

Advantages of questionnaires:
• **Quick and cheap** – in short period of time a large amount of data can be collected. Compared to some methods, they are also cheap.

• **Can provide large samples** – they can be completed without the researcher present. Postal questionnaires can produce large samples for just the cost of a stamp.

• **They can provide quantitative and qualitative data** – this means the method is very flexible and can be suited to the aims of the particular study.

• **Replication possible** – since questionnaires use pre-set questions or statements, replication is possible.

• **Good for operationalising attitudes and asking about actual behaviours** – this method can ask about real-life behaviours and attitudes that have not been potentially tainted by careful control of variables in an experimental setting.

Disadvantages of questionnaires:
• **Respondents can misunderstand questions** – a questionnaire is only as good as the questions and it can be difficult to ask questions in a way that all respondents would understand them and perceive them in the same way.

• **Can lead to biased samples** – a questionnaire can only be used with people who understand the language and are literate, postal surveys would not reach homeless or travelling people plus perhaps only some types of people will respond. This means that the respondents maybe a biased and non-representative sample.

• **Low response rate** – some questionnaires will have response rates as low as 5%. Postal questionnaires tend to be seen as junk mail and end up in the bin.   This is known as sample attrition.

• **Respondents may not tell truth** – they may wish to present themselves in a more positive light (the social desirability effect) particularly if the issue is sensitive, or they simply may not remember something but feel they have to give an answer.

DESIGN OF QUESTIONNAIRES

With response rates as low as 5% it is essential that questionnaires are well designed so as to not introduce other problems.

• **Aims:** it is necessary to be absolutely clear about the aim of the study when designing the questionnaire so that only questions which address these aims are included.

• **Length:** questionnaires should be short and too the point as if they are too lengthy people are less likely to complete them.

• **Advice:** when designing a questionnaire it is a good idea to look at existing questionnaires on the subject area which have shown themselves to be reliable and valid. Advice can also be sought from experts in the field.

• **Statistical analysis:** how answers are to be analysed needs to be taken into account from the beginning.

• **Presentation:** they need to look presentable to encourage people to fill them in. They need to give the impression they are not junk mail.

• **Questions:** you need to avoid leading or double barrelled questions (with 2 aspects where respondents may wish to answer differently on each part). Simple questions should come before more probing or personal ones but they need to be kept interesting so respondents complete the questionnaire. The questions need to be simple, specific and unambiguous. Lie detector scales can be included where the same question is asked in different ways a number of times.

• **Incentives:** you need to include a pre-paid envelope for reply, it is also a good idea to offer incentives e.g. entry into a prize draw, for replying.

• **Pilot study:** a pilot study is a preliminary study which seeks to gain knowledge on the variables involved and eliminate problems in the study. When doing questionnaires it is a good idea to perform a pilot study where you ask respondents open questions or ask for feedback on the questionnaire.

- **Measurement scales:** attitudes or behaviours can be measured in a quantitative manner by using measurement scales. One example is the Likert scale where respondents are asked to give their degree of agreement with a statement. It might score in the following way:

Strongly agree	Agree	Neutral/ undecided	Disagree	Strongly disagree
1	2	3	4	5

The problem with this scale is that respondents may chose a mid-way value repeatedly so you cannot assess them properly. Also, respondents may choose the same value repeatedly once they think they have the gist of the questionnaire – this is called **response-set bias** and for this reason you need to alternate the slant of the questions to make respondents stop and think (e.g.` I am keen to recycle when I can' and `I don't tend to bother with recycling') the numerical values then need to be changed accordingly.

INTERVIEWS

Like questionnaires, this is an adaptable method as it can produce quantitative or qualitative data. Questions are asked in a face-to-face manner, there are 2 broad types of interview;

Structured/formal interviews: this is when a questionnaire is read to P.s and their response is written down by the interviewer. The questions are set beforehand and do not deviate from the list – for this reason they are fairly easy to conduct. Structured/formal interviews tend to ask simple questions which can produce quantitative data.

Unstructured/informal interviews: the topic of the interview may be decided upon beforehand but the exact questions asked are not but will depend on what the P. is saying, almost like a conversation. A good rapport between the P. and the interviewer is essential and interviewers need a good deal of training to bring out quality data from P.s. This method tends to produce data that is rich in detail and qualitative but the data can be quantitatively analysed by applying categories etc. to the data.

Semi-structured interviews: these combine both the above methods by deciding on some questions beforehand but also being responsive to what the P. says and then asking further questions on what they are conveying.

Advantages of interviews:
Good with complex issues: when a topic is complicated or sensitive, interviews (particularly unstructured ones) are probably the most appropriate method since a good interviewer will make the P. feel more relaxed and hence be more open when answering questions.

Can help alleviate misunderstanding: questions can be clarified in interviews and P.s can be asked if they understand.

Method is adaptable: it can produce quantitative or qualitative data, with unstructured/semi-structured methods the questions can be adapted to the P.s. Structured methods, particularly, allow for replication.

Disadvantages of interviews:
Interviewer effect: this is when interviews inadvertently change the P.s response from what it might have been. e.g. white researchers don't tend to get the most accurate results from black children.

Training for interviewers: whilst structured interviews require interviewers to have little training, far more is required for unstructured method. Sensitivity and skill are required for dealing with sensitive issues in unstructured method.

Ethics: P.s may be unaware of the true purpose of the study, they may also divulge more than they had wanted and leave themselves feeling vulnerable.

P.s may not be able to find ways to express themselves: particularly in unstructured method. The method depends on the P. being able to access and verbalise the required material, some may not be so able to introspect.

The researcher first needs to decide whether they will use a **structured interview** where all P.s answer the same questions, or an **unstructured interview** where the researcher may vary the questions depending on P. responses. The researcher could also use a mixture of both methods.

Structured interviews can be quantified easier than unstructured interviews so this is another consideration.

Care needs to be taken over the type of interviewer used since:

• **Gender and age of interviewer** can influence the responses given, particularly when the questions are of a sensitive sexual nature (Wilson et al, 2000)

• **Ethnicity of interviewer** – sometimes interviewers have difficulty interviewing someone from a different ethnicity to themselves – you may need to understand someone's culture to be able to be able to gel with a stranger in order to get honest and open answers.

• **Personal characteristics of interviewer** – some people are easier to get on with than others. Use of language, manner and appearance may all influence the P.

Training for the interviewer is important – they need to be taught how to listen. Non-verbal communication is also important to relax the P. so they will open-up and give honest answers. It is best to ask about factual, easy to answer information first and leave the more probing or sensitive questions until later by which time, hopefully, a rapport has developed.
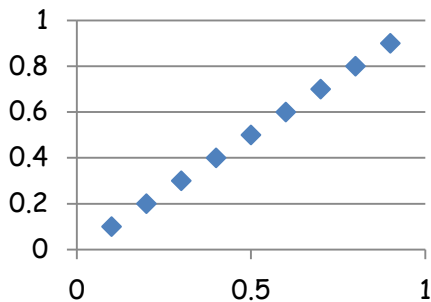
CORRELATIONAL ANALYSIS

Correlation is a method of **data analysis** which involves measuring the degree of relationship between different variables (or co-variables).

A positive correlation is when as one variable increases, so does the other e.g. heat and incidence of sun stroke.

A negative correlation is when as one variable increases, the other decreases e.g. sales of winter coats and temperature.

Correlation is measured as a **co-efficient** – this is a number which measures the degree (or strength) of relationship between the 2 variables. It runs from +1 through 0 to -1.

A perfect positive correlation would be +1, a perfect negative correlation would be -1 (note it doesn't matter whether the correlation is negative or positive when studying it's significance since whether it is + or – only depends on how you have labelled the variables). The nearer to +1 or -1, the more significant the findings are.
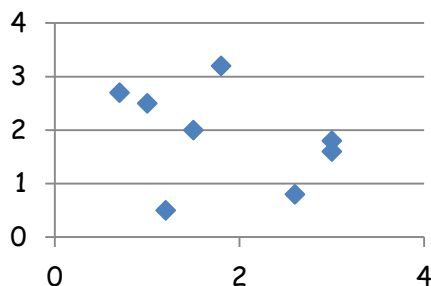
+1 perfect positive correlation
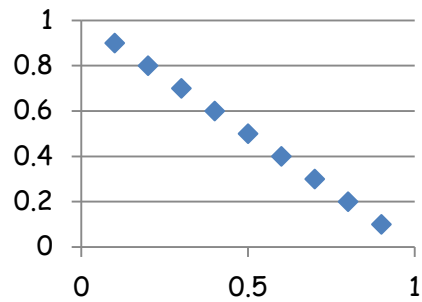
-1 perfect negative correlation

If there is no correlation then the co-efficient would be close to 0.

0 correlation

CORRELATIONAL ANALYSIS

CORRELATIONAL ANALYSIS

Advantages of correlational analysis

• **Allows predictions to be made -** when we have a correlation, we can use this to make predictions e.g. the temperature rises so a store does not order any more Winter coats.

• **Allows quantitative measurement of relationships** – co-efficients show the strength of the relationship between 2 co-variables so we are able to see how significant the relationship is.

• **No manipulation involved -** because behaviour and I.V.s are not manipulated, this can be a quick and ethical method of data analysis.

Disadvantages of correlational analysis

• **Cause and effect** – correlation does not imply causation since the results may be due to intervening or extraneous/confounding variables. An intervening variable is when another variable is influencing both of the 2 variables being correlated e.g. personality may affect people's smoking habits and their cancer rates. Since we are not manipulating the I.V. in a controlled manner, it does mean that confounding variables are possible.

• **Quantification problems** – when a sample size is large a correlation of +0.3 might be significant whereas the same value would not be significant for a small sample size. Conversely, a large co-efficient may not be significant when it involves a small sample size (so you need to look up the significance in a table of co-efficients).

• **Only works for linear relationships -** correlation only measure linear (those falling on a straight line) relationships. Curvilinear relationships would show a zero correlation even when there is an obvious relationship e.g. temperature and aggressive behaviour.

OTHER RESEARCH METHODS

Content analysis

This is a method used mainly in media research. It produces quantitative data from qualitative material such as newspapers, TV, speeches, graffiti, adverts etc. For e.g. a researcher may be interested in gender differences in the voice-overs in adverts and may feel that these reinforce traditional gender stereotypes e.g. the voice over for big cars, power tools etc. are male but those for small cars and foods are female. A grid could be drawn up, the advert watched and it scored on the grid. A researcher may carry out thematic analysis i.e. counting the number of times particular themes occur and organising them into a hierarchy of main themes and subthemes.

Case studies

These are in-depth studies of an individual or a group. It is usually the most in-depth method and the qualitative approach is usually followed. Explanations of behaviour may be written in descriptive (not statistical) ways. The case study P. may report their feelings and this is normally carried out over a long period of time.

There is argument over the value of case studies – some state they are the bedrock of psychology whilst others adhere, almost religiously, to the experimental method.

Advantages of case studies

• **Rich detail** – this method can provide the most detail and depth and lead to greater understanding than other methods e.g. Freud used case study method to record the cases of his patients. The method also shows the differences between people which experimental method does not.

• **The only method to use** when it is impractical or unethical to study people in any other way e.g. abused children. It can study people that have had unique things happen to them e.g. Phineas Cage who had a metal bar go through his head but he lived.

• **Useful for contradicting theories** – you only need one example which goes against a theory to disprove it

Disadvantages of case studies

•   **Cannot be generalised** to other people since no two people are the same.

•   **Researcher bias** – the researcher may produce researcher effects where they influence what is said. They may ask leading questions or interpret things in a particular way.

•   **Cannot infer causality** – since no variables are being manipulated. This leads to the question, should we always be looking for nomothetic (universal) laws of behaviour so is acceptable to look for ideographic (individual) laws of behaviour?

Longitudinal studies

These are studies which are carried out over a long period of time.

Advantages - it is possible to follow changes in the same participants, so there are no individual differences.

Disadvantages- these types of study can take several years and there is a danger of losing participants (sample attrition) as they may not want to take part after a while, they may move to a different area or they may die. This leaves the researcher with a small and possibly biased sample at the end of the study.

Cross-sectional studies

Instead of looking at the same participants over a long period of time, we could look at participants of different ages and compare them. For example, we could study the cognitive (thinking) abilities off children aged 5, 6, 7 and 8 and look for differences.

Advantages - There is no danger of losing participants through sample attrition, so the sample size is maintained.

Disadvantages - Because we are comparing different participants, there will be individual differences.

Measures of central tendency are used to show the typical or average score. They try to indicate the midpoint. There are 3 methods

The mean – this is the average of the scores. All the scores are added together then divided by the number of scores.

Strengths – it is easily calculated and is the most sensitive of the measures of central tendency as it is working with all the raw scores.

Weaknesses – it can give a misleading result if the scores are skewed and it might not represent an actual observed score.

The median – this is the score in the middle if we put the scores in order. If there are an odd number of scores then it is the one in the middle, if there are an even number of scores then take the average of the 2 middle values.

Strengths – it is not affected by freak/outlying scores and is easier to calculate than the mean.

Weaknesses – it is not as sensitive as the mean as it does not use all the raw data and it may be misleading in small sample.

The mode – is the score that occurs most frequently. There can be 2 modes in our data which is a bi-modal distribution.

Strengths – it is not affected by freak/outlying scores and it does represent an actual score which might make more sense than the mean eg what is the average dress size of UK women?

Weaknesses – if there is a bi-modal distribution it will not give a central value and it is not very sensitive.

Measures of dispersion – this shows the spread of the scores. There are 3 types

The range – this is the difference between the highest and lowest scores + 1

Strengths – it is easy to calculate and included extreme/freak values

Weaknesses – it can be misleading when there are freak values, also 2 sets of data maybe skewed in opposite ways but the range will not show that.

Semi-interquartile range – this shows the middle 50% of scores. Place the scores in rank order and the first quartile (=Q1) is the first 25% of scores, the third quartile (Q3) includes the first 75% of the scores. The semi-interquartile range is the Q3 – Q1 divided by 2.

Strengths – it is not affected by outliers/freak scores and is fairly easy to calculate.

Weaknesses – it is inaccurate if there are large intervals between scores and it doesn't take extreme scores into account.

Standard deviation – this is essentially the average of all the deviations from the mean. The larger the number, the larger the dispersion

Strengths – it is a more sensitive measurement of dispersion than range as all the scores are used in the calculation. If we know the `standard deviation' of the test used we can calculate the number of people who fall between certain scores.

Weaknesses – more complicated to calculate and data needs to fall on a normal distribution to be meaningful.

Below – a normal distribution for I.Q. where the test has an s.d. of 15.

0.13%   2.14%   13.59%   34.13%   34.13%   13.59%   2.14%   0.13%

55   70   85   100   115   130   145

Normal distribution
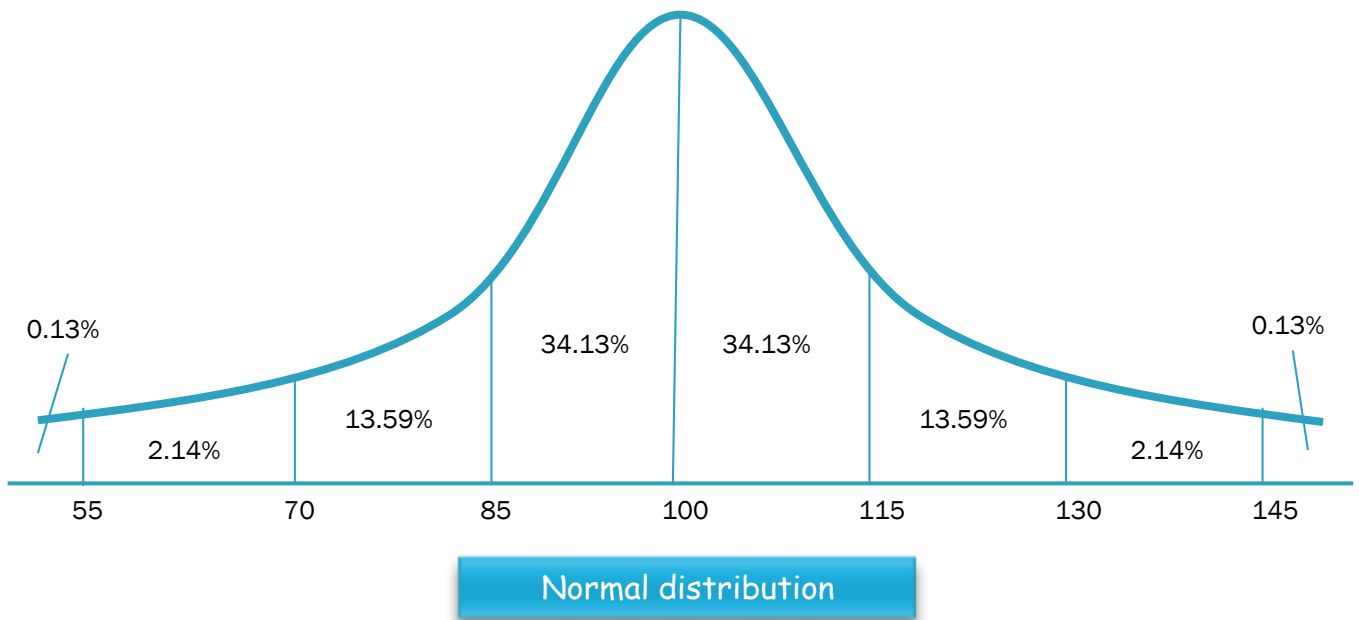
We may also want to display our descriptive in graphs or charts to make it easier for the reader to see what is going on in the data. There are various ways we can do this:

1. Bar charts – these are used for categories to be compared such as males compared to females.

2. Histograms – these are similar to bar charts but are used for continuous data such as test scores.

3. Line graphs (sometimes called a frequency polygon) – these are similar to a histogram in that the data on the x-axis are continuous and you just draw a line from the middle part of the top of each bar in a histogram and you have a line graph.

4. Pie charts – these are circular and different sizes of the slices in the pie represent different percentages.

Here are some examples of the different charts and graphs you can use.

Male/female differences in chocolate consumption by age
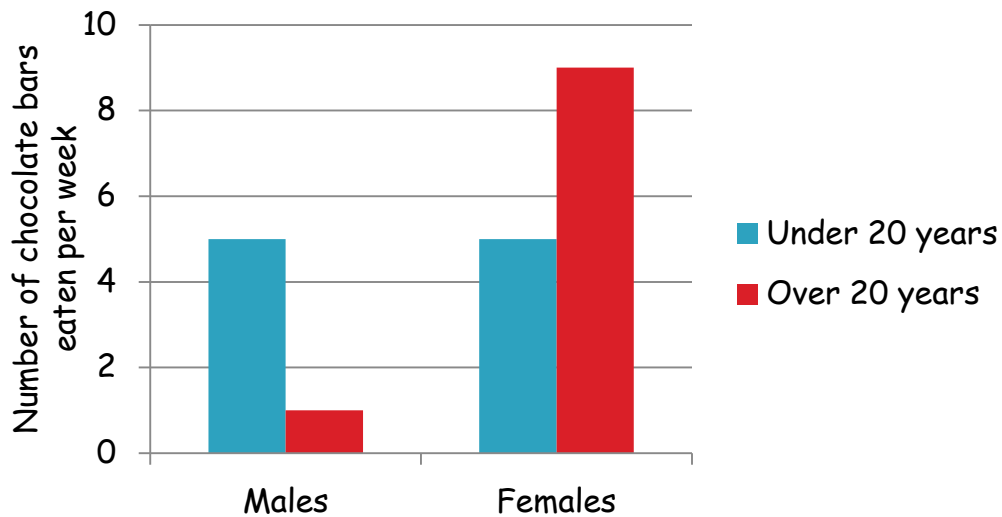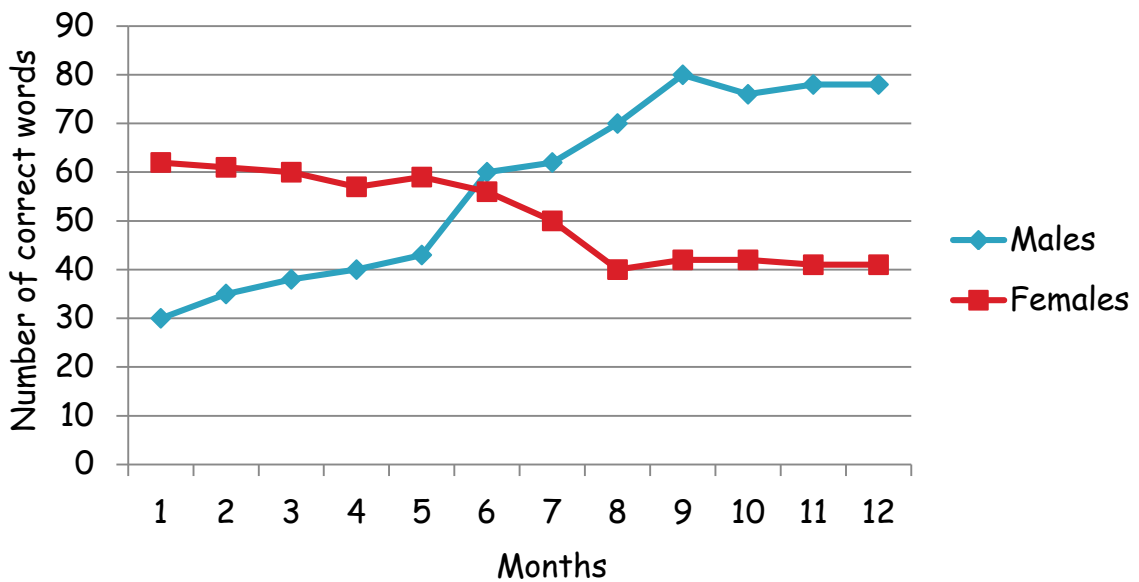

Bar chart
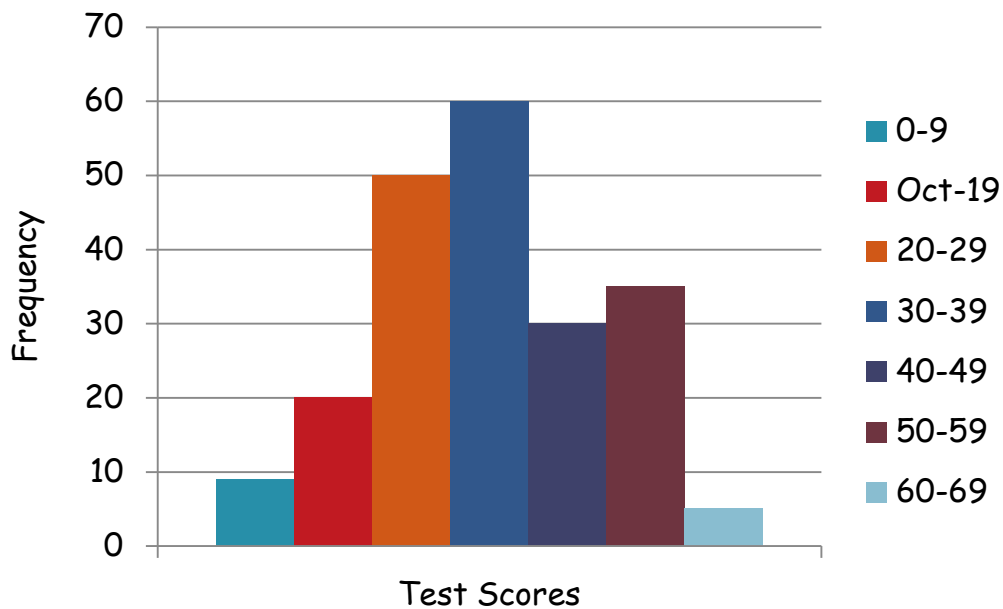
Histogram

ANALYSING QUANTITATIVE DATA

Effect of reading programme on male and female school readers

Line graph

Different ways British people accessed the internet in 2010

- mobile
- dial up
- cable
- no internet access

22% · 12% · 47% · 19%

Pie chart

Skewed and normal distributions

In the example we used for standard deviation, we used a normal distribution. Not every dataset turns out to be normal though- some are skewed and when they are shown in a graph they look uneven. For example a positively skewed distribution would have the peak on the right of the graph and would have a lot of high scores, whereas a negatively skewed distribution would have the peak on the left of the graph and would have a lot of low scores. For some statistical tests (the parametric ones), it is essential to have a normal distribution or at least close to it.

Inferential statistics

Inferential statistics – these tell us what the data mean.

Statistical tests are used in order to ascertain whether our experimental hypothesis has been supported or whether the null hypothesis must be accepted. One of the first things you need to do when choosing a statistical test is work out what level of data you have. It is therefore really important that you understand what the different levels are. It's easy to remember as the initial letters of each type of data spell out **NOIR**, which is French for black!

LEVELS OF DATA

One of the first things you need to do when choosing a statistical test is work out what level of data you have. It is therefore important that you understand what the different levels are. It's easy to remember as the initial letters of each type of data spell out NOIR – French for black!

N OMINAL

O RDINAL

I NTERVAL

R ATIO

Nominal data – This is data which is in categories. Example: If I observed the choice of drink purchase in the canteen I might draw up a chart like this:

	Males	Females
Tea	3	5
Coffee	9	6

So I am counting the number of participants who fall into each category.

Ordinal data – Is data in which the differences between the data points are **NOT** equal. For example if I asked you what your 3 favourite colours are you might say 1. Blue, 2. Yellow, 3. Red. But, you might like blue a whole lot more than you like yellow and you like yellow just a little bit more than you like red. So the distance between your 1, 2 and 3 would look like this:

large ⟶ 1
small ⟶ 2
3

You are putting things in order but the differences between the data points are not necessarily equal.

Interval data – is data which the differences between the data points ARE equal. It applies to things which are measured on an accurate scale of measurements such as time, distance, weight, decibels etc. For example, if you measured the reaction times of your participants it may be that you had data for participant 1 of 1 second, participant 2 did the test in 2 seconds and participant 3 did the text in 3 seconds. So your 2 and 3 would look like this:

Same difference ⟶ 1
Same difference ⟶ 2
                    3

Because this is an accurate scale of measurement the distance between 1 second and 2 seconds is exactly the same as between 2 seconds and 3 seconds. Contrast this with the example of ordinal level data in which the 1, 2 and 3 are not on an accurate scale (it was just based on opinion). You can see that interval data is much more precise.

Ratio Data – is just the same as interval data (i.e. on an accurate scale of measurement in which the data points are equal) but with ratio data it is possible to have a '0' measurement. For example, if I ran a test to see how much weight participants lost when following a healthy eating programme, it could happen that a participant would stay the same weight and therefore score 0. Contrast this with the example of if I measured Usain Bolt's reaction time coming off the blocks in the 100 metres sprint. It might be 0.001 seconds but it could never be 0 and so could never be ratio data.

Statistical tests – different types

Statistical tests look at the degree of similarity or difference between 2 or more sets of data. Which test is used depends on the type of data obtained, whether you are testing for an association, a correlation or a difference.

Which test is going to be used, whether it is a one or two tailed (directional or non-directional) hypothesis and the significance level are all set out before data is collected.

Statistical tests are described as "parametric" or "non-parametric". Don't be afraid of these names! Just think of parametric tests as being more powerful and sensitive tests, but because of this they need the most accurate levels of data (interval or ratio data) and your data need to have a normal distribution (or at least be close to it). Nonparametric tests are less powerful and so you can use just ordinal data and you don't need to have a normal distribution.

USING STATISTICAL TESTS

How to Choose The Right Statistical Test

Use the flowchart to guide you through the decision making process when choosing a statistical test. You just need to ask yourself 4 questions.

Have I got NOMINAL data? — **Yes** → Chi² (Independent Measures)
                                  → Sign Test (Repeated Measures)

No

Am I looking for a correlation (relationship) between 2 sets of data? — **Yes** → Spearman's rank correlation coefficient (NON PARAMETRIC)
                                  → Pearson's product moment (PARAMETRIC)

No

Am I looking for differences between 2 sets of data?

Yes

What experimental design do I have?

Repeated Measures /Matched Pairs

Non Parametric
Wilcoxon Matched Pairs Signed Ranks test

Parametric
Related t-test

Independent Measures

Non Parametric
Mann Whitney U test

Parametric
Unrelated t-test

Tips for remembering the flowchart

1. Pearson's Product Moment is used for parametric data (say this aloud!)

2. The parametric tests of difference are both t-tests, one is <u>un</u>related (for independent measures design in which the participants are <u>un</u>related to each other ie totally different people) and the other is <u>related</u> (for repeated measures design or matched pairs design in which the participants <u>are related</u> to each other because they are either the same people (repeated measures) or matched with each other (matched pairs design).

3. The non-parametric test of difference for independent measures design is a Mann Whitney U test. Imagine Mr Mann in one group and Mr Whitney in the other group. Different men in different groups.

The data is applied to an equation which is different for each test. The number which this gives is called the **observed value.** To find the correct table to see whether this figure is significant, it is necessary to look at whether the hypothesis is 1 or 2 tailed and what significance level is. Some tests will have separate tables for one and two tailed tests, others will have them on the same table but in different columns. You can find these tables on the internet.

Once the correct table is found, you look for your correct sample size (`n') and follow the appropriate columns until the points intersect and there is a single number; this is the **critical value.**

For some tests the observed value needs to be above the critical value, in others it needs to be below in order to be significant – if a table is given in an exam it will tell you if the observed level needs to be above or below the critical level for it to be significant. If it is significant the experimental hypothesis has been supported, if it is not then the null hypothesis has been supported and the experimental hypothesis rejected.

It is then necessary to formulate the statement of results. This needs to relate to the original hypothesis.

When data is nominal and there are independent groups, the test is Chi square and the word `**association**' is used in the hypothesis.

A non-directional (2-tailed hypothesis) might be that `there will be a significant association between birth order and career'

A directional (1-tailed hypothesis) might be that `there will be a significant association between students achieving high grades in GCSE and staying on for `A' levels. The results might be written as `the hypothesis that there would be a significant association between birth order and career has been supported..

When dealing with tests of **correlation** either a Pearson Product Moment (if data fulfils the criteria for being parametric) or a Spearman rank correlation (if data is non parametric and at least ordinal). In this case the word `correlation' needs to be in the hypothesis and statement of the results.

A non-directional hypothesis might be `there is a significant correlation between eating cheese late at night and the number of nightmares experienced by participants'.

A directional hypothesis might be `there is a significant negative correlation between the amount of alcohol consumed and amount of time P.s can balance on one leg'.

A non-directional hypothesis might be `there is a significant correlation between eating cheese late at night and the number of nightmares experienced by participants'.

A directional hypothesis might be `there is a significant negative correlation between the amount of alcohol consumed and amount of time P.s can balance on one leg'.

When testing for **difference,** a Mann-Whitney (for non-parametric data) or unrelated t test (for parametric data) will be used if using independent groups/measures design and data is at least ordinal. When using a repeated/related design a Wilcoxon or related t test should be used (for non-parametric and parametric data respectively).

A non-directional (2 tailed) hypothesis might be `there is a significant difference between the number of P.s who have nightmares after eating cheese late evening and those who do not eat cheese late evening'.

A directional (1 tailed) hypothesis might be `P.s eating cheese late evening will have a significantly higher number of nightmares compared to P.s who do not eat cheese late evening.'

Remember: any experimental hypothesis needs to inform the reader of both variables involved (the IV and the DV or the 2 variables in a correlation) contain the word `significant' and the words `association', `correlation' or `difference' according to the design of the study. Words like `relationship' or `affect' will not gain marks.

The Sign Test

In the exam you may be asked to work out a Sign Test. Don't worry! This is the easiest of all the tests and only requires basic addition and a bit of common sense. Let's imagine we have 10 participants and they have each said whether they prefer Snickers or Twix to eat. 2 of the participants said they liked them both the same, 7 said they preferred Twix and 1 said they preferred Snickers.

You simply add up the lowest number of participants, in this case it is 1 as only one person said they liked Snickers best whereas, 7 said they liked Twix best. You then compare this figure (called the observed value) with the critical value in a stats table and at $p < 0.05$, N=8 (we omit the 2 scores that said they liked them both the same), the critical value is 0. In a Sign Test the observed value has to be below the critical value to achieve significance so, we have a non-significant result.

USING STATISTICAL TESTS

Levels of Significance

When we find an association, correlation or a difference between 2 sets of data we need to be able to work out what exactly this means. We do this by using levels of significance. In Psychology we use 2 basic levels of significance – 5% and 1% which are written as $p< 0.05$ and $p<0.01$.

If we get a significant result at $p<0.05$, what this means is that there is a less than 5% likelihood of the results being due to random chance and there is a 95% likelihood that the results are due to the effect of the independent variable (or the co-variables if it is a correlation).

If we get a significant results at $p< 0.01$ what this means is that there is a less than 1 % likelihood of the results being due to random chance and there is a 99% likelihood that the results are due to the effect of the independent variable (or the co-variables if it is a correlation).

It is therefore harder to get significance at $p<0.01$ than at $p<0.05$, but if you do manage to achieve this, then you have a really meaningful result. This is why we include the word "significant" when we write hypotheses, because if we get a significant result (at $p< 0.05$ or $p<0.01$) then we accept the experimental hypothesis which states "There will be significant difference between the scores on condition A compared to condition B" (2-tailed) or Participants in Condition A will score higher than those in Condition B (1-tailed).

If we don't get a significant result, then we have to accept the null hypothesis which states "There will be no significant difference between the scores on Condition A and Condition B"(2-tailed) or " Participants in Condition A will not score significantly higher than those in Condition B" (2-tailed).

Hopefully, you can see now why we need 2 hypotheses – an experimental and a null, because either we get a significant result or we don't.

Type 1 and Type 2 errors

When you understand the concept of levels of significance, you are ready to move on to Type 1 and Type 2 errors. First of all, here is the easy way of remembering which way round they go:

1 comes before 2
O (for optimistic) comes before P (for pessimistic)

 Let's imagine you get a significant result from your test, you might think " Wow, this is great!" but you might have only tested 10 people, so finding a difference between only 2 groups of 5 people doesn't really mean anything as the sample is so small. Alternatively, you may find that you have a significant result when you look at the $p<0.05$ level but not when you look at the $p< 0.01$ level. In other words you have made a Type 1 error, you have been too **optimistic** about your results.

On the other hand, you may look at the $p<0.01$ level and think "Oh no! This is terrible!". However, when you look at the $p<0.05$ level you find that you can get significance. In other words, you have made a Type 2 error, you have been too **pessimistic** about your results.

Sometimes in exams, the question may ask "Why were the researchers confident that they had not mad a Type 1 error?". The answer to this will be found by looking at the critical value for the $p< 0.01$ level and if the results are significant even at the very rigorous level of $p< 0.01$, then the researchers can be confident that they have not been overly optimistic about their results.

Peer Review

Peer review is a necessary process that has to happen in research. This is when new research is submitted to research journals and experts in the field (usually two) try to find fault with it, e.g. looking for possible confounding variables or wrong conclusions etc. The review can be double blind (both the reviewers and the author remain anonymous) or single blind (the reviewers remain unknown to the author) or open (both the reviewers and the author know each other. They can either :

1. Accept the paper without any changes (this never happens!)
2. Accept the paper with amendments to be done.
3. Reject the paper.

 It is vital to have this process as it protects the public against wrong research getting into the public domain (Think of all the children who were not vaccinated as a result of the wrong research which frightened parents into believing that the MMR vaccine causes autism). It also ensures that only research of the highest quality is published. It also prevents plagiarism (copying other people's work) and falsifying data. It can however, be negative in the sense that it can be exploited e.g. reviewers using information for their own research, or it could be seen as elitist e.g. not allowing radical new ideas to go forward.

How to write a Psychology report

There are some strict guidelines on how a Psychology report should be written. The following sections have to contain specific information:

1. Abstract – this gives a summary of the research for someone who doesn't want to read the whole paper.

2. Introduction – this gives a review of the literature in the filed up to the point where the current research is being done. It should conclude with the statement of the aims and hypotheses. The idea is it should start broad and then narrow down to the specific hypothesis being tested.

3. Method – this should contain sections on 1. Experimental design, iv.d.v. etc 2. Participants profile, sampling method etc 3. Apparatus/ Materials used.4. Procedure including how confounding variables were controlled 5. Ethical considerations.

4. Results – Descriptive and inferential statistics.

5. Discussion – this relates the findings to the literature, offers criticism both positive and negative of the current study and suggestions for future research.

6. Conclusion – very short, like a mirror image of the abstract.

7. References – using Harvard referencing system e.g. Thomas S. (2019) "DLD in preschool children ", Journal of Cognitive Psychology, 15 (1), 405-420.

8. Appendices

USING STATISTICAL TESTS

Implications of Psychology for the Economy

In recent years there has been a growing appreciation of the role which research in Psychology plays in the economy. Research is important for practical applications which can benefit society in the following ways:

1. Developing effective therapies

2. Effective therapies enable people to return to work sooner and contribute to the economy, pay taxes etc.

3. We can export treatments to other countries

4. A healthier workforce also puts less strain on the police, social services, legal system etc.

Some Advice on Research Methods

You have learned all the theory for Research Methods now and this is the foundation of everything in Psychology. Well done!

You now need to get really good at applying your knowledge to some questions. You will find it gets easier and easier the more practice questions you do.