M1 Methodology
Questionnaires

Using questionnaires to obtain survey data | *Social Approach*

Generally, the term **survey** refers to both the idea of a **questionnaire** or an **interview**. A survey can be carried out for a wide number of reasons. Those people who are the survey-takers are called **participants**. A survey is planned with a set **aim** or aims in mind. A **hypothesis** is usually set too, which is about what is being tested and setting a level of measurability about it. For example, in a survey to investigate fashions of shoes people wear, the *aim* might be "to investigate types of shoes people wear" whereas the hypothesis might be "yellow shoes are worn more by women than men". Experiments also have hypotheses which are alternatives to the **null hypothesis**. In an experiment, the alternative hypothesis is called an experimental hypothesis.

Null Hypothesis - hypotheses which state that there will be no common relationship in the data unless by chance

Alternative Hypothesis - alternative hypotheses to the null hypothesis, stating the expected outcome of the data

Planning a Questionnaire

A questionnaire must be designed very carefully. There are a wide variety of types of questions to choose from, as outlined below:

- A good questionnaire contains both **open** and **closed** questions. An open (or open-ended) question allows the respondent to give their thoughts, opnions, reasons and justifications where they are not tied to a set of responses chosen by the questionnaire creator, for example "Explain why you like/dislike football". A closed (or closed-ended) question limits the options of responses the respondent has to choose from in answer to the question, for example, a **dichotomy** (e.g. "Yes" or "No" answers) or a **Likert**-scale question. The balance is particularly important, having too many open questions which require long answers put the respondent off completing the questionnaire, so these should be limited to only the absolute essentials you need to ask the participant
- Questionnaires all aim to gather **data**, but there can be different types. **Personal data** (e.g. age and gender) should only be asked for if it is completely relevant to the subject, as it is unethical to ask for unnecessary to ask for irrelevant personal information and also puts many people off doing the questionnaire
- The two main types of data obtained from a questionnaire are **quantitative data** and **qualitative data**. Closed questions tend to generate quantitative data, which is easy to collate and analyse because all of the answers are usually multiple-choice and therefore easy to group together and spot immediate correllations. Open questions, however, produce qualitative data which provides more in-depth information into the participants' thoughts and opinions, but because all respondents respond to the open questions differently, this data is hard to collate, and even harder to analyse as a single group

Some of the possible types of question you might find in a questionnaire are shown below:

Tick the appropriate box which corresponds with your opinion for each of the below statements:
(**Key: SD** Strongly disagree, **D** Disagree, **DK** Don't know, **A** Agree, **SA** Strongly agree)

Statement	**SD**	**D**	**DK**	**A**	**SA**
I feel confident about my school studies	☐	☐	☐	☐	☐
I like to go out with my friends	☐	☐	☐	☐	☐
I enjoy going ice skating	☐	☐	☐	☐	☐
Psychology is the most fun subject	☐	☐	☐	☐	☐

The box above shows a **Likert-scale** question. This question type involves the respondent ticking a box to correspond with their level of agreement with a set statement, where the boxes are along a measured scale. This is a closed question which will produce quantitative data which can be easily collated and analysed.

Rate yourself on the following scales by placing a cross (×) on the line where you think you fit:

Sad	0	______________________________	10	Happy
Mean	0	______________________________	10	Friendly
Selfish	0	______________________________	10	Generous

The above question is a rating scale question, where the respondent is required to mark where they think they fit among the scale. Again, this is a closed question where the participant is required to *self-evaluate*. This means that the answers may be biased based on how the respondent wants the experimenter to see them. This is called **social desirability** (more on this later on).

Circle any of the following words below which you believe apply to your personality

Horrible Polite Generous Messy
Selfish Handsome Fussy
Jealous Intelligent Sporty Caring

The question above is again a closed question in which the respondent must identify characteristics which they believe applies to their own personality. As with the others, this is open to bias, as many participants prefer to answer as they want to be seen (for example, not many people would choose to circle "Selfish" even if they think that they are selfish, purely because they don't want the experimenter to think they are selfish, as it is considered a negative quality).

Why do you think football is the UK's most popular spectator sport?

____

____

____

This is an example of an open-ended question in which the respondent is free to answer however they want to. This can be useful because the participant replies with a lot more information. This is because they are free to express opinions, reasons and justifications for their answer. The drawbacks to using this type of question are that it is harder to collate and analyse the data, and that different respondents can respond differently because these questions can often be ambiguous and be interpreted differently.

The table below displays the main advantages and disadvantages of open and closed questions:

	Strengths	**Weaknesses**
Closed	▪ All respondents give standardised answers which means numbers can be generated as answers, i.e. the data is easily collated ▪ The wording of the question can be used to make the aim clear, so that all respondents interpret the question identically – making the data more reliable	▪ Respondents are forced to choose from a set of given answers, when they may not necessarily agree with any of the choices they are presented with ▪ Certain choices may mean different things to different respondents (e.g. "unsure" may mean 'don't know' to some people, and 'sometimes yes, sometimes no' to others)
Open	▪ Respondents are not forced into specific answers and so can say what they truly want to say ▪ More detailed, richer data can be obtained ▪ Questions open to interpretation and so the data obtained is more valid as the answers are what the respondents 'really' think	▪ Responses are difficult to analyse as they are long, detailed, and different from one another ▪ Because the data obtained is qualitative, the data cannot be collated and averages cannot be calculated or displayed as graphs or in tables ▪ Sometimes respondents choose to miss out these questions as it is difficult/boring to respond to

Factors to Consider with Different Types of Data

Strengths and weaknesses of qualitative data:

- ✓ It gives detailed information on a subject and allows for in-depth analysis. This is because the level of detail on a respondent's answer extends beyond the simple "yes" or "no" choices you get with closed questions. There are certain key words in an answer which express their real opinions, also, for example, whether a participant uses 'and', 'or' or 'but' says something about their attitude towards the subject
- ✓ As with most open-ended questions, qualitative data tends to be more valid because the respondent writes down what they *really* think, rather than what they think the questionnaire wants them to think
- ✗ Answers are harder to compare against each other, and they all take longer to analyse. Answers may be very different that they are hard to categorise, and the results can be very long and hard to summarise
- ✗ The data can take a long time to gather because it often proves difficult to find respondents willing to answer the longer open-ended questions which usually provide this qualitative data; similarly, many respondents will be reluctant to provide in-depth answers as they don't benefit from contributing such long answers

Strengths and weaknesses of quantitative data:

- ✓ Answers can be fairly quickly and easily collated (categorised and averaged) and analysed. Averages, percentages and other statistics can be easily drawn from the findings with this type of data, and they can be presented easily in graphs and tables which are attractive and easy to understand
- ✓ Most quantitative data is reliable because the way in which it is gathered is controlled sufficiently for the study to be replicated in order to see if similar results can be obtained
- ✗ Because the respondents have a set group of choices which they can use as their answer in the closed questions, the answer they want to put down may not be there. They are forced to choose from one of the options there (unless an "Other" option is given in the questionnaire), which means that they are having to answer untruthfully
- ✗ The respondents' answers may be untruthful in the sense that their answers have been guided by the way that the questions have been set. The three aspects of these methods of "lying" are outlined below:

Social Desirability - respondents answering to a question how they think they ought to be seen to be "socially correct" (for example, no one will answer "Yes" to the question "Are you a racist?")

Demand Characteristics - occurs when respondents try to 'guess the purpose' of the questionnaire or piece of research and so give the answers which they think the researcher would want them to give

Response Bias - happens when questions are listed so that respondents may answer (for example) "no" so many times in a row that it becomes habit and they answer all future questions similarly

Collation of Questionnaire Data

It is very easy to collate and analyse quantitative data obtained from closed questions. Take the Likert-scale question from the above example. The scale works so that you receive a certain number of points per answer per question:

Tick the appropriate box which corresponds with your opinion for each of the below statements:
(**Key: SD** Strongly disagree, **D** Disagree, **DK** Don't know, **A** Agree, **SA** Strongly agree)

Statement	**SD**	**D**	**DK**	**A**	**SA**
I feel confident about my school studies	☐ (1)	☐ (2)	☐ (3)	☐ (4)	☐ (5)
I like to go out with my friends	☐ (1)	☐ (2)	☐ (3)	☐ (4)	☐ (5)
I enjoy going ice skating	☐ (1)	☐ (2)	☐ (3)	☐ (4)	☐ (5)
I feel self-conscious about my personality	☐ (5)	☐ (4)	☐ (3)	☐ (2)	☐ (1)

For each person that responds to that questionnaire, there will be an individual score. For example, if someone answered DK, A, D, A in order of the questions, they would receive 3 + 4 + 2 + 2 = 11 points. If ten people from the **target population** responded to the questionnaire, the data is easily **collated** by adding up the individual scores and dividing the overall score by 10. This average is then the average score for your target population based on your findings. However, in order to obtain a **representative sample**, you would need a bigger sample to test.

Another example of a closed question is the "Yes"/"No" response. 50 people were asked the question "Do you like pizza?" and their responses are shown in the table. The results showed that 70% did like pizza, 24% did not like pizza, and the remaining 6% were unsure.

Answer	Number of answers	%age
Yes	35	70
No	12	24
Unsure	3	6
Total	**50**	**100**

Drawing Conclusions

It is important that all questions asked in a questionnaire are appropriate to the topic in mind, and this can be done by making sure that each question is relevant to the aims set at the beginning of the questionnaire, before it was even written. When the data has been collated and statistics have been drawn from them (mainly the closed-ended questions), you can explain what the results show and how well they match up to your initial hypotheses.

An important part of the questionnaire process is making sure that the right questions have been asked, and once the process is finished, you need to ask yourself if you have found what you set out to find. Businesses use them all the time to develop products based around the consumers' wants and needs, and this can also be applied to psychology, in that the data can be used to *generalise* the target population.

An Evaluation of Questionnaires

Strengths and weaknesses of questionnaires are shown below:

- ✓ The same questions are asked of all participants using a set procedure. There is very little variation in the way that the questions are asked of the respondents, so nothing should really affect the answers given except for their own opinions, which is what you are looking to obtain
- ✓ Provided the researcher does not interfere with the questionnaire process or affect the answers given in any way, the results from the questionnaire are valid
- ✓ Because they use the same questions and have a set procedure, the questionnaire process is easily repeated – and so the study is said to be reliable
- ✗ Most questionnaires consist of a series of fixed questions and do not allow the respondents to expand on their answers or add their own input – this may often mean that their responses are not valid

M2 INTERVIEWS

Planning and carrying out interviews to obtain qualitative data | *Social Approach*

An **interview** is the other main type of survey. They consist of an interviewer asking the participant (the interviewee) to answer a series of questions in a face-to-face verbal situation. Having an interview allows both the interviewer and the interviewee to expand on, or clarify the questions. There are three main types of interview:

1. a **structured interview** follows a set format, where a questionnaire has been set by an individual and those questions are to be asked of the participant – some clarification on the questions is possible
2. a **semi-structured interview** consists of a series of set questions, but the interviewer is able to probe the participant into expanding on their answers when necessary, and can also add in a few extra questions to obtain more information
3. an **unstructured interview** follows no set format, so the questions are not predesigned – this allows the interviewer to structure the interview entirely around the participant's answers

The data obtained from an interview is essentially *all qualitative*. There might be a certain number of dichotomies in an interview, but generally the questions are open-ended.

An interviewer is able to make notes throughout the interview, but a preferred method of recording is to use, for example, a dictaphone or video camera, so a hard copy of the interview is permanently available for the researcher's use. Either way, the notes or recordings must be **transcribed** later on – this is, writing them out in full so that the interview can be properly analysed, linking back to your aims and hypotheses. This is time-consuming but is a crutial part of the interview process.

An interview, among any other form of data collection, has to be ethical, and these are some of the steps a researcher has to follow to make sure it is:

- the respondents must see the interview **schedule** before the interview begins, so that they know what the interview is regarding and are prepared to respond to the questions asked
- the respondents must know about and agree to the chosen method(s) of recording the interview
- after the interview has taken place and it has been transcribed by the researcher, the respondent must see a copy of the transcript to confirm that is what was said or what happened

SUBJECTIVITY / OBJECTIVITY

As with any other method of research, an interview has its fair share of bias. Like in a questionnaire, there is room for the methods of "lying" such as **social desirability** and **response bias**. But also, with an interview, the researcher themselves is able to affect the answers from the interviewee. They can do this *in* the interview by using facial expressions to mentally 'judge' the respondent; and they can do it *outside* of the interview by interpreting the results using their own views and judgements.

The term **subjectivity** refers to the idea of results analysis being affected by the input of the researcher. This means that they cannot be directly verified by someone outside of the research team. Whereas **objectivity** means that there is no bias affecting the result – this means that the researcher has not included their own judgement and views in the data. An objective study has findings which are easily verifiable by somebody who is outside of the study team. All scientific studies should be objective.

Interviews can be objective by ensuring that the researcher is not affected by whether they like, or agree, with the interviewee. A full transcript should be produced so the researcher cannot choose what to include. Also, they should have one other researcher look at the results. An interview which is not objective gives findings which are not useful.

Some of the strengths and weaknesses of interview are outlined below:

- ✓ The interviewer is able to explain what the question wants, and can explore further into their opinions by asking further questions, whereas a questionnaire is limited to the questions set
- ✓ Contain in-depth and detailed data is produced which is usually valid – the respondents can talk in their own words and are not restricted – because the data is "real life" and "true" it is likely to be valid

- ✗ An interviewer can find it hard not to influence the answers of the respondent, based on how they ask questions or where they put emphasis on, e.g. by asking "You are not prejudiced, are you?"
- ✗ Interviewees may respond differently to different interviewers, for example, giving different answers to a man or woman interviewer (**researcher bias**)
- ✗ The researcher can put their own interpretation into the data analysis and can quickly form themes in their mind which they analyse all data to

A COMPARISON OF QUESTIONNAIRES AND INTERVIEWS

	Questionnaires	Interviews
Reliability	Structured questions are set the same for each respondent, so the research is easily repeated and therefore likely to be reliable	Each respondent is interviewed in a different place at different times, and often by a different person. This makes it difficult to replicate the interview process exactly as before, so testing for reliability becomes difficult
Validity	Set questions with forced multiple-choice answers are likely to gain "untrue" answers from the respondent by not offering a choice that they perhaps wish to give, making the data not valid	Questions can be explained and explored in more detail. The interviewer is able to probe the respondent for as much information into their "true" opnions as possible – making the data obtained valid
Subjectivity	Follow a structured format and are less open to researcher bias in the data analysis – these mainly consist of closed questions which do not require researcher interpretation, and the open questions yield fairly short answers, making themes in the respondents' answers more easy to spot	Definitely open to researcher bias in the data analysis because generating themes in the findings requires some interpretation. The data is very open to subjectivity, but the analysis can be objective by making sure that the steps are clear

M3 SAMPLING METHODS

Sample sizes and methods | *Cognitive Approach*

When you carry out a study, you have a **target population** – which is the group of people you are doing the research into. However, you will almost always find it is not possible to study *everyone* in that group, either because some people are not willing to participate, or even just because the group is too large, for example, you cannot test every American male between the ages of 20 and 40 just because this is your target population. Therefore, you have to use what is called a **representative sample** – this simply means including members of each type of person within the target population in the correct proportion. It is important that a representative sample contains a decent number of participants in order to obtain the best results. To decide on the sample size, you need to know three things:

- the size of the **sampling frame** (those that you are choosing from)
- the **confidence interval** (how far answers are thought can be not valid or unreliable – for example, a researcher might say that the results are true within a confidence interval of ±3 so that a score of 20 in a questionnaire truly has a score between 17 and 23)
- the **confidence level** (the percentage of the sample that is likely to represent the population)

It is common for researchers to have a confidence interval of ±3 and a confidence level of 95%. The sample size is calculated using a complicated formula you do not need to learn. An online calculator is available at www.surveysystem.com/sscalc.htm if you wish to further investigate.

RANDOM SAMPLING gives everyone an equal chance of being chosen. Each time a participant is to be chosen to be part of the sample, everybody within the target population is available for selection and has the same chance as the next person of being used. Examples of ways to use simple random sampling include putting everyone's names in a hat and drawing them out as each participant is required, or assigning everyone within the target population a random raffle ticket and drawing out raffle ticket numbers to choose the participants

- ✓ There is no bias in the way in which participants are chosen – everybody has an equal chance and no one is systematically excluded from the sample. Therefore, the sample is theoretically likely to be representative
- ✓ It is clear to everyone how the sample was chosen – it can be easily explained and understood
- × There can be difficulty in obtaining the names of everyone in the target population, which may cause bias
- × Bias can arise if certain participants are chosen but cannot participate, for example if they are busy on the day in which the study is taking place, or simply if they don't want to
- × The sample chosen by random sampling may not be useful, for example, if the study is investigating obedience within males and females between the ages of 20 and 40, random sampling may produce all male participants

STRATIFIED SAMPLING is used to ensure that certain groups are all represented by the sample. The researcher will decide what specific groups need to be tested within the sample, and will calculate how many people should be selected from each group using proportions. If you are investigating obedience in males and females between the ages of 20 and 40, you may separate them into four categories (males 20-29, males 30-40, females 20-29, females 30-40). If there were four times as many men as women, you would sample four times as many males

- ✓ Each group has to be represented by the sample, and so clear conclusions can be drawn outlining differences between the groups
- ✓ Stratified sampling ensures that the right number of people from each group is chosen to represent their group. With random sampling, people from each group may still be picked but not in the correct proportions
- × It can be difficult to know how many to choose from each group to make the findings generalisable
- × The groups chosen by the researcher may not necessarily all be the important groups. Having the groups already decided means that some people will automatically be ruled out as participants

VOLUNTEER SAMPLING (or self-selected sampling) calls for volunteers to willingly contribute their time for the study. They may respond to a letter inviting a number of people to participate, or more commonly might respond to an advertisement, which often involves payment also

- ✓ The most ethical form of sampling because the participants come to the researcher rather than the researcher seeking them out
- ✓ Volunteers have self-selected themselves and are therefore most likely interested in the piece of research, this means that they will be less likely to give biased information or go against the researcher's instructions
- ✓ The volunteers are willing to be involved in the study
- ✗ Can take a long time to get enough willing participants because the researcher has to wait for a response to their advertisement or letter
- ✗ Because the participants have all selected themselves, they might all be similar in some way which will not provide a broad spectrum of results which is applicable to many other groups

OPPORTUNITY SAMPLING is less a mathematical method of sampling, and more of a "pick whoever is available" approach. Researchers will use whoever they can find who is filling to take part in the study. The ways in which the participants are chosen are not structured. An example might be someone doing questionnaires in a town high street. They will probably not have a specific participant in mind, but instead will just attempt to use everyone, and will happily include anyone who agrees to take part in the findings

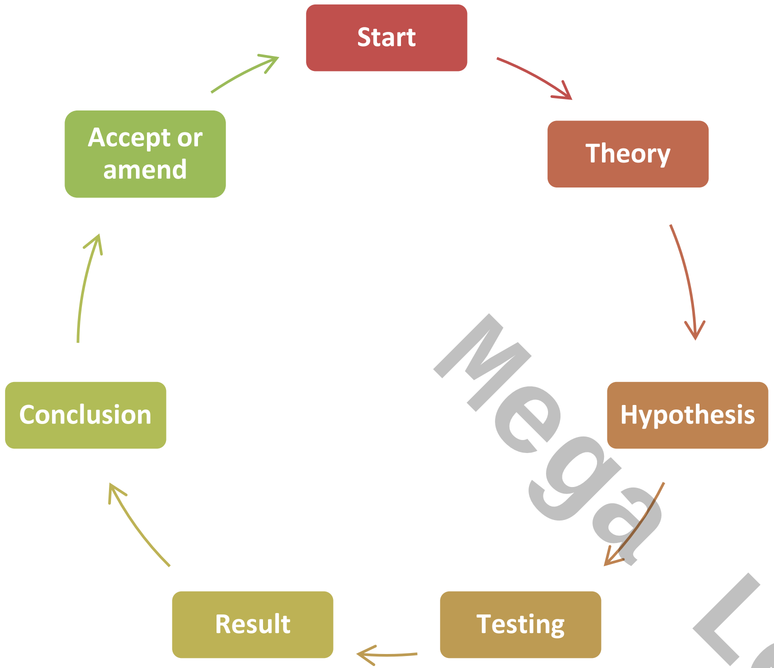
- ✓ Can be ethical, for example, if the researcher is able to judge if the experiment will upset the potential participant or if they can work out if they will be too busy to participate in the research
- ✓ The researcher has a lot of control over who is used to participate, and access to potential participants is not limited
- ✗ There is a lot of potential bias from the researcher – they may only choose people who are similar to themselves in some way, whether it be preferring people from the same sex or people of the same age as them

M4 SCIENTIFIC TESTING

The process of scientific testing in psychology | *Cognitive Approach*

Science involves developing theories which explain events.

A theory must be scientifically tested in order to be approved and accepted. Psychological theories on any of the approaches to psychology have been tested by scientific methods, trying to prove them false.



Science involves taking the theory and generating hypotheses from it, which can be scientifically tested. They are tested against reality to see whether the hypothesis is false, or if the findings support the prediction.

Theory - an idea about why something happens, based on research

Hypothesis - a statement about what the theory predicts

It is important to use strong controls when testing scientifically, to avoid bias. Concepts have to be measurable and produce quantitative data.

Science should also be **objective**. This means not letting personal (**subjective**) opinions affect the data. When a study has strong controls, objectivity and **operationalised** variables (measurable concepts), the results should be *replicable*. They can be shown to be *reliable* by repeating the study to find similar results.

Operationalisation - operationalisation of variables is making them practically measurable

It the results are reliable and support the hypothesis, the theory is supported by the study. Otherwise, the theory has to be amended, or abandoned, should the results show the theory as false. Then further hypotheses are generated.

HYPOTHESES

From a theory, an **experimental hypothesis** is generated (also called the **alternative hypothesis**). This is a statement of what is expected to come of the results. The **null hypothesis** states that any relationships or patterns displayed in the results will be down to chance, so this hypothesis predicts no relationship. When using statistical tests, it is the null hypothesis which is being tested.

Directional Hypothesis - hypotheses which predict a specific direction of the results

Non-directional Hypothesis - hypotheses which do not predict the specific outcome direction of results

A hypothesis can also predict the direction that the results are going to take. For example, a **directional** (**one-tailed**) **hypothesis** might be "Chewing gum will produce higher recall of a list of words from the participants" as it states the specific direction which the results are predicted to go by. A **non-directional** (**two-tailed**) **hypothesis** might be "Chewing gum will affect participants' ability to recall a list of words" as it does not state the specific direction, it could either improve or reduce their recall.

Variables

A variable is anything which is likely to affect the experiment. The **independent variable** (**IV**) is the variable which is changed or manipulated by the experimenter. This is to see what effect it has on the **dependent variable** (**DV**). This is what is being measured by the researcher. The DV changes as the experimenter manipulates the IV. Both of these variables must be *measurable*, this means **operationalising** them.

An **extraneous variable** is any other variable which affects the results. Experiments have strong controls to decrease the number of extraneous variables, which affect the results as well as, or instead of, the IV. There are two main extraneous variables:

- **participant variables** – for example age, gender, experience and mood of the participants
- **situational variables** – for example temperature, background noise, interruptions and lighting conditions

Extraneous variables should be controlled, but any which are not controlled and affect the results are called **confounding variables**.

Experimental Controls

In any experiment, it is important that every variable should be controlled (apart from the IV). This is because the aim of the experiment is to test how the change of the IV affects the DV. Therefore, all extraneous variables should try to be controlled. Below are some examples of experimental controls to ensure of this:

standardised instructions	Each participant is given the same set of standardised instructions before the experiment begins, to ensure all participants have the same information
experimenter effects	Cues or signals which come from the experimenter which can affect the responses of the participant

Experimenter effects occur when the experimenter gives certain cues or signals which might affect the participant's responses. These can be tone of voice, or non-verbal, such as facial expression. The most common way to avoid such effects is to have someone other than the experimental designer carry out the experiment. In a **double-blind technique**, the participants are not aware which group they are in or what the study is about, and the study is carried out by someone other than the person who knows who is doing what. Neither the participants nor the person running the study know exactly what is to be expected. In a **single-blind technique**, the participants are not aware of what is expected but the person running the study is. This technique stops the participants' expectations from affecting the results – but it does not prevent experimenter effects. The double-blind technique is preferable from the two.

M5 EXPERIMENTAL DESIGN

Types of experimental design and methods of study | *Cognitive Approach*

CONDITIONS

A study can have a different number of **conditions**. These are aspects of the independent variable, and there must be two or more conditions, so that a comparison can be made with the results. For example, in an experiment to find out how age affects IQ, you might test three age brackets: 10-19, 20-29 and 30-39. In this case, there are *three* conditions for the experiment

There are three main types of **participant design**. These will depend upon the **conditions** the experiment has.

The researcher can ask all participants to take part in every condition of the study. This participant design is called **repeated measures**. All participants do all groups.

An alternative design is to split the participants up so that they each only do one condition. This is called **independent groups** design. There are different participants in each group, and they each do different conditions. Under this design, all conditions are still fulfilled by the participants, although not every participant does each one.

A **matched pairs** design is very similar to the independent groups design. Here, again, participants only do one condition each. However, before they are split into groups, the researchers pair up the participants with factors which they think are important, such as age, gender, social class or ethnicity. This enables as similar as possible participants in each group without using the repeated measures design, so still a good comparison can be made using results.

One strength of *repeated measures* is that all the participants do every condition, and so participant variables (features of the participants which might affect data) are controlled. This is because if every participant does each condition, any participant variables present will be present for *all* conditions, so it cancels itself out. Another strength is that more data can be gathered because everyone does all conditions. Also, if you use independent groups and there are two conditions, you will require twice as many participants, which is more costly, less convenient and possibly less ethical.

A weakness of repeated measures is that **order effects** may arise. It is possible that if one participant is doing all four conditions of an experiment, they may get more tired each time the next condition happens, and therefore put less effort in: this is called **fatigue effect**. Also, it is possible that participants may get better at the conditions as they go along because they know what to do and how they could do it better, having already done the first condition: this is called **practice effect**. Order effects do affect data, and incorrect conclusions can be drawn from the findings unless they are noticed. A further weakness is the possibility of **demand characteristics**. These occur when participants respond by trying to guess what the study is about, and may want to help the researcher out by giving the response they think the study wants; or by purposely going against what is predicted. Either way, this will affect results.

Repeated Measures

- ✓ Participant variables are controlled, as all participants do all conditions
- ✓ Uses fewer participants, so more convenient in terms of costs, convenience and ethics
- × Demand characteristics are possible if participants guess the aim of the study from repeating conditions
- × Order effects such as practice and fatigue effects can affect results if there is no counterbalancing

A benefit of using *independent groups* is that there are *no order effects*. This is because different people do different conditions, no one participant does two conditions. Similarly, there is less possibility of demand characteristics being present. This is because when the participants are no taking part in every condition, they are less likely to guess the nature of the study.

A drawback of the design is participant variables can easily affect the findings, as different participants are in each condition. Also, more participants will be required for the study, because they are needed for each different condition. This is negative because it is more costly, it can be more unethical, it makes samples slower to obtain, and means that the study can take longer to complete.

Independent Groups

- ✓ No order effects to affect the findings because each participant only does one condition
- ✓ Demand characteristics are less likely to be present
- × Participant variables are very likely to affect the results
- × More participants are required, so it is less convenient, more costly and possibly less ethical

A strength of the *matched pairs* design is that (like repeated measures) participant variables should not affect the findings, if the most important participant variables have been matched well. Also, there will not be order effects, because everyone is doing different conditions, and no participant does two.

However, a weakness of matched pairs is that different people are used, and even though the experimenters attempt to match them into pairs in some ways, there are still always going to be some participant variables, which can affect the data. This is an unavoidable feature of the design. For example, it may be difficult to match education or background. Also, another weakness is that more people are needed for this design, so the same problems as before arise: more costs, less ethically-valid, etc.

Matched Pairs

- ✓ No order effects to affect the findings because each participant only does one condition
- ✓ Demand characteristics are less likely to be present
- × Participant variables are very likely to affect the results
- × More participants are required, so it is less convenient, more costly and possibly less ethical

Experimental Methods

Researchers not only have to consider the type of *participant design*, but also the **experimental method**. There are three types of experiment…

The first type is a **laboratory experiment**. These usually take place in a lab, or at least a controlled setting – this will be unnatural for the participants. These are the most scientific experiments, because only one variable is manipulated (the independent variable, IV) and all extraneous variables are normally controlled. A common design within this type of experiment is to have an **experimental group** and a **control group**. The experimental group does something the control group does not, so the control group provides a **baseline measure** (i.e. what the dependent variable, DV, would be like without manipulation). The findings for the experimental group are then able to be compared against the baseline measure. This is important, because otherwise the "normal" situation would be unknown.

A second type is the **field experiment**. These have as many controls as possible, but not all of them can be controlled to their optimum because they take place "in the field" (in the real world). Field experiments take place in natural settings, such as that of Hofling et al. (1966) (see 1.7 Hofling et al. (1966) for more), which took place in a hospital – a natural setting for the nurses who were being studied. Apart from being in a separate setting, the features of field experiments tend to generally mirror those of lab experiments.

In the field - a term meaning in the participants' natural setting

The third type is a **natural experiment**. These also take place "in the field" rather than in a laboratory, and they involve a *naturally-occurring IV*. These are uncommon, because it is out of the experimenter's control if the IV is manipulated naturally. An example of a natural experiment comes from those which have looked into the effect of children's behaviour from TV. Their behaviour was studied, such as levels of aggression, in an area where TV was yet to be introduced, and then they continued to study after television had appeared in that area. The researchers themselves had not arranged for TV to not be in the area and then for them to have it – that would be neither ethical nor practical. They found a community where television was about to be introduced and decided to study it: a natural experiment. A natural experiment is a **quasi experiment**, because the IV is not manipulated by the researcher. A quasi experiment uses situations which are *discovered*, not *manipulated*, and so take place in the participants' natural setting.

	Strengths	Weaknesses
Laboratory experiments	Strong controls are present, which makes the experiments replicable, so reliability can be tested (and results are more likely to be reliable) Because of the strong controls on extraneous variables, clear cause-and-effect conclusions can be drawn from the findings	Because of the strong controls, the tasks may become unnatural, giving invalid results There is no ecological validity, because the environment is not natural for the participants and there are controls Experimenter effects could affect the results, and cause bias
Field experiments	More ecologically valid than lab experiments because they take place in natural settings to the participants Fairly replicable, and so still likely to be valid, because generally there are some controls present, as with laboratory experiments	Because of the natural setting, it is hard to control all factors, so findings may be less valid than lab experiments, also may not be replicable Experimenter effects could affect the results, and cause bias
Natural experiments	The independent variable occurs naturally, so the findings are valid because it is not artificially set up High ecological validity, experiment takes place in a natural environment, with a naturally-occurring IV	It is difficult to control variables because the IV is naturally occurring, so other variables may contribute to findings Hard to control the experimenter effects, using a double-blind technique is not straight forward, because most are carried out in a natural setting

The table above displays the main strengths and weaknesses of each experiment type.

M6 Case Studies

Evaluating case studies as a research method | *Psychodynamic Approach*

A **case study** is a particular type of research method where one individual (or sometimes a small group of people, connected in some way, such as a group of children being brought up together and deprived of parenting) is studied over a period of time, allowing data to be gathered in depth and detail.

Whilst a case study itself is a research method, we do not say that data is gathered *by* a case study. Instead, data is gathered *for* or *within* a case study. It is other research methods from within the case study that gather the data, such as interviews, observations and questionnaires. A researcher will use more than one of these types of research method within a case study to obtain sufficient data to analyse. The researcher will use **triangulation** to discover themes between the data that are gathered, and to produce the final results. Triangulation involves pooling together the data obtained from all of the research methods (i.e. everything from the whole case study) and looking for themes.

Are case studies scientific?

It is unclear as to whether a case study is scientific or not. There are points for and against it being so. One might say that a case study is *not* scientific because they tend to gain **qualitative data** which means the aim of a case study is to an understanding of meaning, which would rely on the researcher's own interpretation, making them subjective

However, another person might counter-argue that they are scientific because the researcher gathers information systematically and makes sure that there is sufficient evidence and support for any claim made within the data, some-times in the form of quotes, or percentages, etc, also the research methods can be scientific, if for example a questionnaire is used, it should be valid (i.e. by measuring what it claims to measure, such as what people think, not what they feel they should be thinking) and reliable (i.e. if shown the same questions, the same answers are given)

Case studies generate mainly *qualitative* data. This data is analysed by finding common themes from the data obtained. This is done by sorting the data into tables and flow charts, etc. Frequencies of events that were observed can be jotted down, repeating patterns identified, and such. Statistical tests may be applied to the findings to assess the **level of confidence** in the data. Sometimes, more than one researcher may be analysing the data. If so, and similar themes are found by all researchers, there is a higher level of confidence in the data, and there may be **inter-rater reliability**.

Level of confidence - this shows how confident you can be about the findings; statistical tests will produce a percentage of how confident you can feel - e.g. 97.5% certain that your findings are true

Sometimes, the research will consist of there being more than one case involved in the case study. If so, **cross-case analysis** can be used. What might happen is that the cases get divided up for analysis. Different types of data can be divided up between the cases, rather than doing case-by-case analysis. An example might be a case study involving following three people, where the research methods are interviews, questionnaires and observations. One person would analyse all the data from the interviews, one person from the questionnaires, and so on.

The table below considers some of the factors affecting the validity, reliability and generalisability of qualitative data:

	Evaluation of qualitative data (from case studies as a research method)
Validity	☺ Generally valid because detailed, rich and in-depth; and the information comes from a real person in a real situation ☹ But may be influenced by the researcher
Reliability	☹ Generally unreliable because not easily replicated ☺ However, data from different cases can be obtained and cross-analysed to spot themes
Generalisability	☹ Not usually generalisable because they come from one individual (or a small group) ☺ However, Freud did generalise his theories based on the individual case studies

M7 The Case Study as Used in the Psychodynamic Approach

The case study research method used by Freud | *Psychodynamic Approach*

Freud, the founding father of the Psychodynamic Approach to psychology, only ever used case studies. It is important to understand that he did not use his case studies to obtain research, but was actually trying to help the individuals being studied, who all suffered from **neurosis**. Freud believed the cure to someone's neurosis was in unlocking the unconscious, and only after accessing that area of the brain could any change be made. Note that Freud did not work with sufferers of **psychosis**.

Freud and others believed that the unconscious mind is inaccessible, so he could not use direct questions about it to the patients. Instead, he had to use special ways of reaching this part of the mind. He developed ways of tricking his patients into revealing their unconscious thoughts.

Neuroses - mental health problems where the individual is aware that they are suffering from the neurosis

Psychoses - mental health problems where the individual is not aware that they are suffering from the psychosis

He used a variety of different research methods within each of his case studies, but none of them used the same research methods as other case studies:

● *Dream analysis*

Freud would use **dream analysis** by listening to the content of the **analysand**'s dreams and apply the ideas and concepts from the approach to try and interpret and explain them. The content which is described by the dreamer (i.e. what the dream actually physically entailed) is known as the **manifest content**, and the underlying meaning which Freud would look to decipher is known as the **latent content**. Something called **symbol analysis** is also carried out when trying to access the unconscious through dream analysis, as the manifest content is symbolic of the latent content.

Analysand - the person being analysed in the case study

● *Free association*

Another research method Freud would use was **free association**. This is the idea of associating ideas, things and feelings by saying whatever is in the mind, without censoring your thoughts. As one thing follows another, the analyst listens to find connections which can reveal unconscious thoughts.

● *Slips of the tongue*

Also, Freud would look for **slips of the tongue** (which are often known as **Freudian slips**). This occurs when somebody says one thing but they meant to say another, such as saying "erection" rather than "rejection", or saying "orgasm" instead of "organism". Freud believed that the mistake, or slip, being made revealed repressed unconscious thoughts. They do not necessarily have to be sexual, it could be so much as calling someone by someone else's name, but Freud was focusing on underlying sexual meanings.

Freud's therapy process was called **psychoanalysis**. He had the central purpose of curing his patients of their illnesses, but he would gather data alongside to his therapies in order to improve and amend his theories.

Similarities between Freudian case studies with case studies from other approaches…
There is a strong focus on obtaining qualitative data, and all of the data is in-depth and rich about one person

Differences between Freudian case studies with case studies from other approaches…
There were different research methods found within the case studies, and he was using therapy to try and cure his patients as well as using them as analysands to help strengthen his theories

Evaluation of Freud's case studies as a research method

One strength is that the data are in-depth, detailed and rich with information. Also, his case studies use different methods to uncover unconscious wishes which are impossible to access by conventional means. Thirdly, his case studies act both as a therapy which allows the analysand to be cured, and a research method to help Freud amend his psychosexual theories.

However, the analysis involved in his case studies is personal interpretation, so very subjective and is not scientific. They cannot be replicated to test for reliability, either, because it focuses on the unique individual, and their unique unconscious desires, and the analysis is carried out by one single therapist.

The Credibility of Freud's Theory

Freud used these case studies, such as that of *Little Hans* (see 3.4 Freud's Case Study of Little Hans (1909)) to help improve and amend his theories. Freud's theory came together as the **psychosexual theory** (which is covered in vast detail in Unit 3) and as you will discover simply from reading Freud's ideas, they seem a little hard to believe.

The biggest criticism of Freud's theory is its **credibility** (how believable the findings of research are). Obviously most case studies have an element of doubt towards their credibility because the findings depend on the analyst's interpretation, but Freud's theory is particularly considered to be questionable and controversial.

As a therapist and analyst, Freud had many patients of whom he thought originally he was hearing stories of child abuse from. But according to **Masson** (1984), he later dismissed the idea that his patients had suffered child abuse and came up with the *Oedipus complex* (see 3.3 Psychosexual Development) to help explain their stories. Freud said that child abuse could not have been so widespread that so many of his patients had suffered it, and so thought that his idea of the *Oedipus complex* was a better explanation. However, Masson claims that the stories of abuse were real, and therefore Freud's alternative explanation (in this case, the *Oedipus complex* part of the theory) is not credible.

Masson further criticised Freud's work and thought that there were three flaws:

- Firstly, the power of the analyst interpreting the patient's thoughts and dreams could lead the patient to accept their interpretation, whether they really agreed with it or not
- Secondly, his theory shows gender bias, because Freud focused mainly on young boys with regards to his theory, saying that boys identify with their fathers more than girls identify with their mothers, so girls have less of a moral code, and so Masson said that Freud's theory had **alpha bias**
- Thirdly, his theory was overindulged with sexual matters, which was a sensitive issue for the patients (you will notice from Freud's theory that he interprets most things to have an underlying sexual meaning, which in itself can often prove not credible)

Gender bias

If a theory emphasises one gender over another, this is known as **alpha bias**. Freud's theory shows alpha bias, because it is more centrally-focused on boys, and Freud did not go into much detail about the psychosexual development of girls. Alpha bias is usually against females, and there is a feminist argument that Freud's theories were biased against women. If a theory does not emphasise gender differences at all, this is called **beta bias**

M8 Cross-Sectional and Longitudinal Studies

Which study is better for looking at developmental trends? | *Psychodynamic Approach*

A **cross-sectional study** takes place at one specific moment in time, and compares different groups of people at that time. The participants are tested once, usually to find a simple relationship between one variable and another. This type of study is most common, as it can be done quickly and participants need only be tested once.

Cross-sectional studies are perfect if you want to test a group of participants on their memory, because you can run the experiment and use your findings to come to some sort of conclusion. However, if you want to study developmental trends, for example, to test how a group of participants' memories change with age, a cross-sectional study is not the only choice available.

In this example, of course a cross-sectional study could still be used. This would involve testing the memories of a group of 10-20 year olds, a group of 20-30 year olds, a group of 30-40 year olds, and so on. The findings of each condition could be compared to generate the conclusions. However, a **longitudinal study** may be used, whereby the study follows one group of people over a period of time. This can be as little as a month or so, but many longitudinal studies continue for years, if not decades. In this case, the memory test would be conducted at the beginning of the study period, and then again at regular intervals to see the changes found. The aim is to compare the data of each test to see how the passage of time affects whatever it is being tested for.

It may be difficult to identify a study as longitudinal, as sometimes the time period can be debatable. For example, the study of Reicher and Haslam (the BBC Prison Study) went on for six days. This wasn't a quick hour test of participants, their every move was watched over six days and developmental trends were looked for. But is six days long enough to be considered a longitudinal study? Most would consider the Prison Study to be a cross-sectional study, as it is generally accepted that a month or a couple of months is the minimum period for a longitudinal study, but this amount is variable. It depends upon what the study is looking at on what fits a 'longitudinal' study.

Strengths of cross-sectional studies	Weaknesses of cross-sectional studies
reasonably cheap, quick and practical, as participants need only be tested once and there is no follow-up study necessary participants are more easily obtained, because there is less pressure with cross-sectional studies than there is for them to stick with longitudinal studies less ethical considerations than for longitudinal studies	there is less rich detailed data collected than there is with longitudinal studies with regards to individual participant differences the data collected are from a snapshot in time, it is harder to identify and analyse developmental trends in cross-sectional studies

Strengths of longitudinal studies	Weaknesses of longitudinal studies
the same group of participants is followed throughout the entire study, so participant variables do not affect data collected these studies are the best way of spotting developmental trends as they repeat tests at regular intervals and compare the findings	certain participants from the group may move away or wish to no longer participate, which disrupts the study withdrawal of participants also means if remaining participants share a characteristic, findings are biased there are a number of practical difficulties with longitudinal studies: they can be expensive, they're very time-consuming and the data collection and analysis can vary in its strength if the researchers change over time

M9 Inferential Statistics – Spearman's Rank Correlation Coefficient

Correlation designs and using the Spearman's Rank statistical test | *Psychodynamic Approach*

Correlation designs

As part of the methodology for the Cognitive Approach, we learned of the three experimental designs: independent groups, repeated measures and matched pairs (see M5 Experimental Design). A further design, although not an experimental design, is the **correlation design**. This design of study involves comparing two different sets of data for the same set of participants. Each participant does two measures, and those measures are recorded and compared. This comparison is done by testing the relationship between the two sets of results statistically.

The relationships correlation designs might identify are **positive correlations** (i.e. as one measure goes up, the other measure also goes up) and **negative correlations** (i.e. as one measure goes up, the other measure goes down). When there is no relationship identified, the term 'no correlation' is used.

The data used for a correlation design must be numerical, and both the measures must come from the same participant. There is *no independent variable* and *no dependent variable*. There are just two variables, none of more significance than the other. The hypothesis of such a design will not be about a 'difference between' but will be hypothesising a relationship between the two measures.

Strengths of correlation designs	Weaknesses of correlation designs
good for finding relationships at the start of an investigation; also unexpected relationships; once two sets of data have been collected from the same participants, a relationship can be statistically tested the data yielded is more secure, as there are no participant variables to affect it	the findings only show a relationship between those sets of data, not a definite connection, it does not allow room for the concepts of chance or another factor causing the relationship to arise if the data are artificial or unconnected, it is not valid

Scattergraphs

Correlation data are normally displayed graphically using a **scattergraph**. The scores for each measure for each participant are used along the *x*-axis and the *y*-axis to plot a point for each participant, and then a line of best fit may be drawn to test for a correlation.

What is referred to as the 'eyeball test' is used to see if there is a correlation in a scattergraph. Simply looking at the results plotted on the chart and seeing the line of best fit should tell you if there is a correlation or not. A good measure is to compare the number of scores on or close to the line of best fit with the number of anomalous values. The more values there are on the scattergraph which don't seem to fit, the less likely there is to actually be a correlation.

Spearman's Rank correlation coefficient

The **Spearman's Rank** test (Spearman's Rank correlation coefficient) is a statistical test used to mathematically calculate if there is a relationship between two sets of data. We will use the example of IQ and income of some participants.

Participant	IQ	Income (£)
1	118	35,000
2	103	12,000
3	98	10,000
4	124	18,000
5	109	15,000
6	115	20,000
7	130	30,000
8	110	12,000

In our study, eight participants were given an IQ test and their scores were recorded. The participants then had their relative salaries noted. The following instructions outline how to use the Spearman's test.

Step 1: Rank the first variable, the lowest rank for the lowest score
In this example, this means ranking the lowest IQ with a 1 and the highest IQ with an 8. When two participants share the same score, take an average rank. For example, if three of them have an IQ of 100, occupying ranks 3, 4 and 5, allocate all of them the rank 4 (average)

Step 2: Rank the second variable, the same way as the first

Simply rank the incomes with the lowest income receiving a 1 and the highest receiving an 8

Participant	IQ	Income (£)	IQ rank (Step 1)	Income rank (Step 2)
1	118	35,000	6	8
2	103	12,000	2	2.5
3	98	10,000	1	1
4	124	18,000	7	5
5	109	15,000	3	4
6	115	20,000	5	6
7	130	30,000	8	7
8	110	12,000	4	2.5

Step 3: Calculate the difference between the ranks for each participant

Take the value of Rank 2 away from the value of Rank 1 (in this example take income rank from IQ rank). This will give the difference between the two ranks. If the resultant number is negative, don't forget to record the minus sign

Step 4: Square the differences between the ranks

Square the value you obtain for each participant from Step 3

Step 5: Total the figures from Step 4

Find the total by adding up all the values obtained from Step 4. The Greek letter sigma, Σ, is used to show add total

Participant	IQ	Income (£)	IQ rank (Step 1)	Income rank (Step 2)	IQ rank – income rank	(IQ rank – income rank)^2
1	118	35,000	6	8	-2	4
2	103	12,000	2	2.5	-0.5	0.25
3	98	10,000	1	1	0	0
4	124	18,000	7	5	2	4
5	109	15,000	3	4	-1	1
6	115	20,000	5	6	-1	1
7	130	30,000	8	7	1	1
8	110	12,000	4	2.5	1.5	2.25
					Total (Σ) =	13.5

Step 6: Multiply the value of Step 5 by 6

In our example, 13.5 x 6 = 81

Step 7: Find the value of N

N is the number of pairs of observations you have, so this will simply be the number of participants, in our case 8

Step 8: Square the value of N and subtract 1 from that number

In our example, 8 x 8 = 64 and 64 – 1 = 63

Step 9: Multiply the number from Step 8 by N

In our example, 63 x 8 = 504

Step 10: Divide the value of Step 6 by the value of Step 9

In our example, 81 ÷ 504 = 0.160714285

Step 11: Calculate rho, the result of the Spearman's test, by doing: 1 – Step 10

In our example 1 – 0.160714285 = 0.839285714

A statistical table can then be used to see if there is a correlation between the two sets of data. Statistical tables cannot be simply generated by thinking about them, they take hundreds of mathematical studies to calculate. They are published in masses in books so people can refer to them in order to see if there is a correlation in their data.

The table below is an extract from a statistical table for rho:

		Level of significance for one-tailed test			
		0.05	0.025	0.01	0.005
N =	4	1.000			
	5	0.900	1.000	1.000	
	6	0.829	0.886	0.943	1.000
	7	0.714	0.786	0.893	0.929
	8	0.643	0.738	0.833	0.881
	9	0.600	0.700	0.783	0.833
	10	0.564	0.648	0.745	0.794
	11	0.536	0.618	0.709	0.755
	12	0.503	0.587	0.671	0.727

You will come across levels of significance in more detail in M13 Inferential Statistics – Mann-Whitney U Test, but for now just bear in mind that choosing a level of significance basically means choosing an acceptable level at which you can reject the null hypothesis and that the results are due to chance, and accept the alternative hypothesis.

Using our example study, we have a rho value of 0.893285714. We used 8 participants in the study, giving us an *N* value of 8 so that row has been highlighted in the table.

The critical value shown for the appropriate level of significance in the table has to be LESS THAN rho for the hypothesis to be proven. If the rho value is less than the critical value in the table, the null hypothesis is retained, as it cannot be rejected. This means your rho value must be MORE THAN the value in the table (called the critical value)

So back to our example. We can definitely accept the alternative hypothesis as proven for significance of 0.05 (given that 0.643 < 0.893). This means we can say that the results obtained showing a correlation are less than or equal to five per cent due to chance. We can also say this for 0.025, 0.01 and 0.005 as the rho value is larger than all of the critical values for *N* = 8. Because we can do so for a significance level of 0.005, we can actually say that the relationship is less than or equal to half a per cent due to chance.

M10 Scanning Techniques as Research Methods

Using MRI and PET scans to investigate brain activity | *Biological Approach*

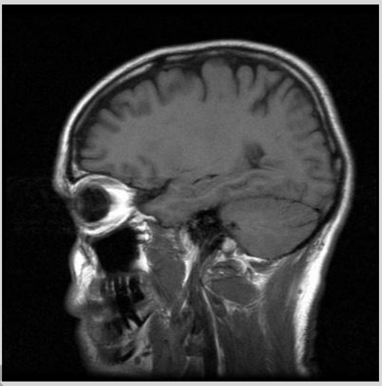
A wide variety of scanning techniques exist. These are used to provide *biological* data, rather than psychological. Scanning techniques include PET, MRI, fMRI, CAT and MEG scans. Only **PET scans** and **MRI scans** are studied as part of the methodology for the Biological Approach.

Scans have scientific purposes. They are commonly used to investigate for possible tumours, strokes or other abnormalities. However, they can be used as research methods too, such as aiding psychologists into understanding of how information is processed. Psychologists and scientists are also using brain scans as research methods, to investigate both **normal differences** between brains (such as differences between a male and female brain) and **abnormal differences** (such as differences between the brain of a murder and a non-murderer).

Possibly the main drawback of these scans is that they are expensive and fairly hard to access. Scanning machines costs tens of thousands of pounds and using them is not cheap. For this reason, they tend to be reserved for hospital needs primarily. However, when used, the main strength is that they offer scientific, reliable and valid findings.

Before brain scanning was made possible, **corpses** were the only brains available for scientific research.

MRI scans

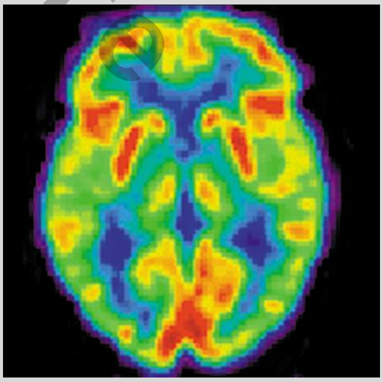


An MRI scan (**magnetic resonance imaging**) is used to look at **structure**. It studies the tissues, looks for abnormalities and can measure the blood flow. It involves injecting a dye into the body to help show organs and relevant areas. A strong magnetic field is passed over the body to pick up radio waves from hydrogen atoms in water molecules, to build up a detailed image of the brain

Different areas of the brain emit differing amounts of radio waves, producing different densities on the image produced of the cross-sectional views of the body

The MRI scan allows the comparison of the structure of brains that are performing normally versus abnormally; belonging to males versus females; and belonging to the younger versus the older people

PET scans



A PET scan (**positron emission tomography**) is used to look at **function**. It studies brain activity levels and can be used to look for evidence of a stroke. It involves injecting a radioactive tracer into the bloodstream with a chemical used by the body, such as glucose, to see where most of the blood is flowing

The radioactive particle emissions (**positrons**) from the tracer give signals which are recorded so levels of activity in different parts of the brain can be detected. Greater levels of brain activity appear on the scan as different colours

Participants are scanned in two conditions – when inactive (to provide a **baseline measure**) and when performing an activity. The difference between the two scans shows which part of the brain is being used

The PET scan allows the study of areas of activity within the brain when stimuli such as faces or names are shown; but also the study of memory and looking at sufferers of schizophrenia or epilepsy

You do not need to know strengths and weaknesses of each scan as part of the course.

M11 Twin and Adoption Studies

Using twins and adoptive families as the subjects of experiments | *Biological Approach*

The Nature-Nurture Debate

When we talk about the **nature-nurture debate**, we mean '**nature**' as in what people are born with, and '**nurture**' is what is learned through interaction with various environments. Therefore nature refers to what is inherited, and nurture refers to what is picked up or learned

A **twin study** can be used to look into the nature-nurture debate. This is because nature is *what we are born with* and what is controlled by our inherited genes. This can be compared to the nurture, i.e. what we develop. There are two types of twin which are studied in the Biological Approach.

Identical twins are **monozygotic** (**MZ twins**), meaning they come from only one fertilised egg. Their DNA is always 100% shared and so they are also always the same sex. Anything which is totally genetic and is inherited by one twin will also be inherited by the other twin. No characteristic is entirely genetic though, as environment is a depending factor on everything, no matter how small its effect. Features of MZ twins:

- MZ twins do not always share the same environments, even in the womb, so they will develop somewhat differently in certain respects, even though they share 100% of their DNA
- there are a few physical differences between MZ twins, including their different fingerprints
- some genetic characteristics are triggered by environment, and so MZ twins may become less identical over time

Non-identical twins are **dizygotic** (**DZ twins**), which means they come from two different fertilised eggs. This means that the DNA is never 100% the same, but is only as similar as that of any sibling pair. DZ twins are expected to share an inherited characteristic to only an extent, not as much as MZ twins.

A twin study compares certain characteristics possessed by both MZ and DZ twins. This is to see if it is genes which influence whether or not they share certain characteristics, or environment. If two MZ twins share the same characteristic, but only one DZ twin does (the other doesn't), it is likely to be an inherited characteristic. When two twins share a characteristic, there is said to be a **concordance rate**. For example, if the concordance rate for schizophrenia in MZ twins is 70% then studies have found that of those tested, 70% of MZ twins both developed schizophrenia when one twin had it. Only 30% would have had the condition in only one twin. Concordance rates are studied in both types of twin for things such as schizophrenia, alcoholism, IQ, depression and anorexia.

- ✓ Both MZ and DZ twins are born at the same time and share the same environment, but MZ twins have identical DNA, whereas DZ twins only share 50% so they help to identify how inheritance of genes influence certain characteristics
- ✓ There should not be significant environmental differences with regards to treatment of twins, because generally, most people will treat all twins as twins, not separately

- × MZ twins are the same sex and identical, therefore are more likely to be treated alike than DZ twins
- × Although the variation between MZ and DZ twins are useful for finding differences, but **epigenetic modification** can also have an effect on the findings (this is the term to describe how over time different environmental influences affect which genes are switched on and off

Because the environment is not the same as their biological parents', **adoption studies** are used studying adopted children in their adoptive families. These are useful because they share their genetic information with their parents, even though the environment is different. This could be used to study, for example, schizophrenia. If a parent of a child has schizophrenia and they are adopted (brought up in another environment) you can study the likelihood of the child having schizophrenia later in life – again, finding out the extreme that genes or environment affect it.

- ✓ Adoption studies are the best way of separating genes from environment so the two can be tested
- ✓ The studies are **longitudinal** and so developmental trends can be identified from them

- × It is possible that the environment of the adoptive families is not as different as it could be from the biological family's
- × Children requiring adoption are often placed with families similar to their own, so the differences are minimal

Many studies have looked at MZ twins **reared apart** (separated at birth). This is useful for looking at two or more twins which are identical, and share 100% of their DNA, and grow up in separate environments. This is usually down to adoptions, where an adoptive parent(s) do not want all twins.

When two twins are identical and brought up in separate environments, we can draw conclusions based on whether or not they share characteristics. For example, when the twins share one certain characteristic, it is most likely going to be down to a genetic basis, because they are still growing up in separate environments. When they don't share a characteristic, but only one twin does have it, it will be most likely down to the environment they are growing up in.

- ✓ Environmental conditions are controlled, and because their environments are controlled to be different, reliable conclusions can be drawn stating if concordance is more likely down to genes or environment
- ✗ Not many MZ twins are reared apart, so the number of studies which can be conducted of this type are minimal and so the conclusions may not be as strong as they could be

M12 ANIMAL EXPERIMENTS

Using animals in laboratory experiments for research | *Biological Approach*

A large number of psychological studies use animals for research. They can be used to help study language usage, memory and learning. The animals used mostly are rodents and birds (mainly mice, rats and pigeons). Very rarely might you find studies on primates, cats and dogs.

Studying genes

Mice have been used to find out how certain genes affect behaviour. Mice are useful for experiments because the breed quickly and the arrangement of genes along their chromosomes is similar enough to humans for the studies to be meaningful. Genes can cause abnormalities in humans; mice are tested to see if that particular gene causes the same abnormality in mice. For example, mice have been used in experiments on deafness

Rats have been used to study Parkinson's disease and gene therapy. Researchers used drugs to replicate in rats the disease. They then used gene therapy to try and reverse the symptoms

Studying the central nervous system

Rats have been used in investigations of the effects of antipsychotic drugs on the brain structure and on the nervous system within the brain. The changes caused by antipsychotic drugs appear to be:
- increased size of the straitum – it is thought that this increased size due to increased blood flow
- increased density of glial cells in the prefrontal cortex
- increases in the number of synapses and changes in the synapses

Studying brain function

Research has been carried out into the way that antipsychotic drugs affect the brain and its nervous system. Most of this research has been carried out on rats and the findings have only been generalised to humans. The research needs to be replicated in humans because of the differences in brain structure and function between them and rats. Some of the findings have come from MRI scans of humans, however, so animal studies are not the only way of researching the area

	Advantages	Disadvantages
Practical advantages and disadvantages	Relatively small; easy to handle; short reproductive periods; some have brain structures similar to humans; short lifespan; strict environmental control	Brain differences in structure and function; different genetic structure; human lives are complex and rarely occur in isolation; diseases being studied have to be artificially replicated in animals which mean they might be different
Ethical advantages and disadvantages	Procedures can be carried out which cannot on humans; **pro-speciesism** suggests we should do whatever to protect our own species; drugs have been developed which would not have been otherwise; knowledge obtained from studies can also improve animal treatment	Animals feel pain too; animals are in isolation in unusual conditions and so feel distressed; animals are not sufficiently different to humans to be treated as objects, and should be treated ethically similar to us

There are a set few guidelines which any experiment involving animals must adhere to. These include:

- the researcher(s) must have a Home Office licence and certificates
- anaesthetics must be used appropriately by someone who knows about them
- caging and social environment must suit the species
- a deprived animal must be monitored and its suffering kept to a minimum

COST-BENEFIT ANALYSIS

It is said that a **decision cube** should be used to see whether a study should be carried out or not. It weighs up the potential benefits from running the experiment, and measures them up against the costs of doing so. This is particularly important when deciding whether or not to do an animal study, as these are among the most controversial (ethically).

The three dimensions to consider in the decision-making process are:

1. what benefit (either for animals or humans) the findings of the study are likely to have
2. the cost of the study in terms of pain or suffering
3. scientific quality (going from *poor* to *excellent*)

If the benefits are not considerably higher than the costs of the study, and the study has not been planned well, it shouldn't be conducted. Similarly, when the benefits highly outweigh the costs and the study is well-thought out and controlled and monitored well, it should be carried out.

M10 Inferential Statistics – Mann-Whitney U Test

Levels of significance and the Mann-Whitney U test | *Biological Approach*

There are two types of statistic covered here. A **descriptive statistic** is used in all areas of the course. This includes measures of *central tendency* (e.g. mean, median and mode) and measures of *dispersion* (e.g. range). This section of the methodology, however, focuses on **inferential statistics**, those looking to draw inferences about the data, rather than just describe the results. In M9 Inferential Statistics – Spearman's Rank Correlation Coefficient you were introduced to the Spearman's test, which is one form of inferential statistical test, belonging here to the Psychodynamic Approach.

You will here learn about the **Mann-Whitney U test**, the statistical test of the Biological Approach; and later will meet the Chi-squared test of the Learning Approach. To see information on the Chi-squared test as well as a guide on how to choose the correct statistical test for a given scenario, see M15 Inferential Statistics – Chi-Squared Test.

Levels of significance

Statistical tests are designed to see if the null hypothesis (which says that the results of a study are due to chance) is true. So the tests are used to assess whether the findings were found by chance. Every study will have chance factors. The idea is to be able to decide what is down to chance and whether a relationship is significant.

In Psychology, anything that occurs due to chance in more than 1 in 20 cases is not accepted. For example if a study looks at whether or not women gossip more than men and you only study 20 women and more than 1 woman goes against the hypothesis, the null hypothesis cannot be rejected. Similarly, testing 100 people and more than 5 people going against your hypothesis means your alternative hypothesis cannot be accepted.

The **level of significance** shows the probability of the results being down to chance. This will be shown as a decimal value. Anything in Psychology must be at most 1 in 20 due to chance, so the maximum level of significance is 0.05 (the same as 1 in 20). Sometimes, acceptable levels may be even higher, such as 0.025 (1 in 40), 0.01 (1 in 100) or 0.005 (1 in 200). We use the letter p to show probability, so $p \leq 0.05$ means the probability of the results being due to chance are less than 0.05 (basically, 5%).

The choice of the researcher in which level of significance to use depends on what is being tested. A level of 0.05 is fairly lenient, whereas a level of 0.01 is quite strict. It a study has been done previously and the results proved not to be due to chance at 0.05, it might be worth repeating the second study at a stricter significance level. However, a new study which has not been seen before may wish to begin with a fairly gentle level of 0.05. The choice of level of significance can also depend on other factors, such as the seriousness of the consequences of the findings, such as with a new drugs trial or for a new educational scheme.

Mann-Whitney U test

The Mann-Whitney U test is used to see if findings are statistically significant for studies where an independent groups design has been used, and the type of data collected is *ordinal* or *interval*. You will meet more on types of data and when to choose each test in M15 Inferential Statistics – Chi-Squared Test.

Whereas the Spearman's Rank correlation coefficient is used to test for a relationship between two variables, both the Mann-Whitney U test and the Chi-squared test look for differences between two groups.

The formula for the test looks rather complicated, but as with the Spearman's test we can break it down into steps:

$$U_A = N_A N_B + \frac{N_A(N_A+1)}{2} - \Sigma R_A \qquad U_B = N_A N_B + \frac{N_B(N_B+1)}{2} - \Sigma R_B$$

In our example, we will use a study looking at males and females completing jigsaw puzzles. Eight males and nine females were asked to complete a jigsaw puzzle. The hypothesis is that males will be faster completing the jigsaws because they are better at visuospatial tasks. Each participant's 'score' was taken – the score is the amount of time, in seconds, it took the participant to complete the puzzle.

Participant Group 1 (Males)	Score	Participant Group 2 (Females)	Score
1	95	1	100
2	78	2	123
3	102	3	89
4	79	4	140
5	84	5	97
6	93	6	110
7	62	7	150
8	92	8	104
		9	96

Step 1: Rank the scores for the participants as a whole
Unlike with Spearman's where you rank each series of data by itself, here, we are going to rank the scores of all seventeen participants as if they were from one group. The lowest score (i.e. fastest time) gets rank 1, and so on

Step 2: Label the groups N_A and N_B and work out the value of *N* for each group
If one group is smaller than the other, the smaller group will be N_A

Step 3: Taking the groups separately, add together the ranks for each group
This is described as ΣR_A *for Group A, and* ΣR_B *for Group B*

Participant Group 1 (Males) $N_A = 8$	Score	Rank	Participant Group 2 (Females) $N_B = 9$	Score	Rank
1	95	8	1	100	11
2	78	2	2	123	15
3	102	12	3	89	5
4	79	3	4	140	16
5	84	4	5	97	10
6	93	7	6	110	14
7	62	1	7	150	17
8	92	6	8	104	13
	$\Sigma R_A =$	43	9	96	9
				$\Sigma R_B =$	110

Step 4: Use the formula to calculate a Mann-Whitney U test result for Group A
$U_A = N_AN_B + (N_A(N_A + 1))/2 - \Sigma R_A$ $U_A = (8 \times 9) + (8 \times 9)/2 - 43$ $U_A = (72 + 36) - 43$ $U_A = 65$

Step 5: Use the formula to calculate the result for Group B
$U_B = N_AN_B + (N_B(N_B + 1))/2 - \Sigma R_B$ $U_B = (8 \times 9) + (9 \times 10)/2 - 110$ $U_B = (72 + 45) - 110$ $U_B = 7$

Step 6: Take the smaller of U_A and U_B and label that value as *U*
In our example, 7 is smaller than 65, so U_B *becomes U (so in our example, U = 7)*

The value for *U* can then be checked against the critical value tables to see if the findings are statistically significant. *U* must be less than or equal to the critical value in the table. An exemplar table is shown here.

Level of significance $p \leq 0.05$

N_A ▼ / N_B ►	8	9	10	11	12
7	13	15	17	19	21
8	15	18	20	23	26
9	18	21	24	27	30
10	20	24	27	30	33
11	23	27	31	34	37

Our value for *U* was 7. The critical value is 18 (which has been highlighted in the table) as N_A was 8 and N_B was 9. Because 7 is less than 18, we can say that the results are statistically significant, and that they support the alternative hypothesis, that males are better at jigsaw puzzles than women.

M14 Observations

Different observation types and inter-observer reliability | *Learning Approach*

You may be familiar with observation being included as part of the data collection for experiments. But in an experiment where the investigator manipulates the IV and controls as many extraneous variables as possible, the observation itself is not a research method. For **observation** to be a research method, it must be the main method of gathering data, and this is done in a natural environment where nothing is manipulated or controlled by the experimenter – behaviour is simply observed, and recorded as natural.

Sometimes, an observation can be in the form of a **structured observation**, where the same situation is repeated with different groups of participants and researchers observe what happens to each different participant in that situation. The IV is not manipulated by the researcher, but the setting and environment are well-controlled, making it structured.

However, structured observations are very rarely used in Psychology. They may be used in child psychology, but otherwise a **naturalistic observation** is used.

Type of observation	Description	Strengths	Weaknesses
Naturalistic in general	An observation is carried out 'in the field' (in a natural setting), and the participants may be aware or unaware that they are being studied	There is ecological validity because they take place in a natural environment for the participants They gather in-depth and detailed data that is usually qualitative, but is still quite rich even when quantitative	It is possible that the observer is subjective because they have to choose what to observe and what to record Data and findings are not generalisable to all people at all times, as the study is a cross-section of one moment in time
Participant	The observer takes part in the study and takes part in all activities – the observer is one of the participants of the study	There is ecological validity because the study takes place in a natural setting The observer is likely to gather valid data which is obtained from a natural setting with natural activities	The observer may become too involved with having both the observer and participant role and may not be able to record all the information needed Difficult to replicate as it's hard to find someone who can do both roles
Non-participant	The observers are not part of the study, they sit away from the activities and do not get actively involved	Findings can be objective and therefore more reliable, because the observers stand back from the study and have more time to record findings Time-tallying can be used which is very difficult when also taking part in the study	The observers are more likely to affect the situation lost from their presence The observers might miss the relevance of some interactions or misunderstand something due to not having an active part in the activities
Covert	Participants do not know that there is an experiment taking place and that they are being studied, the study is being carried out secretly	Studies have high ecological validity because normal behaviour is observed (participants are unaware of the study) The observation is easier because the observer can carry out the study without the participants worrying about the observer	There is no informed consent, so they may not be ethical Participants cannot help the observer (e.g. by finding a suitable location) which causes problems as it is hard to observe in secret as the observer must do something different from the norm
Overt	Participants are aware that they are being studied and are completely aware of all aspects of the study	They are ethical because the observers have informed consent and right to withdraw The observers can ask for assistance from the participants (e.g. where to study from)	The participants are aware of the study and so normal behaviour may not occur It might be difficult to carry out because the observers themselves would be watched to see what they are doing

To summarise from the table above, there are four types of naturalistic observation, which fall under two conditions:

- **participant observation** or **non-participant observation** – either the observer takes part in the study as one of the participants, or they remain a sole observer to record information only
- **covert observation** or **overt observation** – either the observation takes place secretly without the participants knowing they are being studied, or the observer tells them fully about the study to gain informed consent

Whilst one of the main weaknesses of observations in general is that they tend to be hard to replicate, and therefore more often than not lack much reliability (as it cannot be tested for), an observation can have high **inter-observer reliability**. This occurs when there is more than one observer allocated to the study, and they each record their own data separately. After the data collection has taken place, the findings from each are compared and if there is a clear correlation in the data then the observation is said to have inter-observer reliability, which is a strength.

Data Collection in Observations

An observation does not only collect *qualitative data*, but also *quantitative data*. Whilst less in-depth and rich with interpretable information, quantitative data (numerical) are useful when it comes to analysing the results obtained from the observation as a whole. There are two methods explained here for collating quantifiable data:

Tallying

This involves making a mark each time a specific behaviour is observed. For successful **tallying**, there should be an initial observation, preferably with more than one observer, in which categories of behaviour are recorded so that all the researchers know what behaviour should be tallied

Time-tallying

Time-tallying involves using a tally table to show behaviours being observed, but rather than giving one tally for every time a behaviour is observed, it means putting down a tally mark for each interval of time (set by the observer) that the behaviour *remains to be done for*. For example, if you are observing the types of toys a child plays with, and they play with some play-dough, are you going to wait until he's finished playing with the play-dough until your next tally? The child could play with it for a long time. Instead, mark off one tally for every minute (or other period of time) he continues playing with the play-dough. When he's finished, the tallying stops, and the next toy goes up

M15 Inferential Statistics – Chi-Squared Test

The Chi-squared test and when to use each statistical test | *Learning Approach*

The Learning Approach requires that you do a practical (and also that you understand for the exam) using the **Chi-squared test**. This document will take you through the Chi-squared test in a step-by-step process, but first, we must look at *when* to use each test.

There are a number of factors which decide on whether a Spearman's Rank correlation coefficient, Mann-Whitney U or Chi-squared test should be used for a given set of data. Factors which must be taken into consideration are:

- the experimental design
- whether the study is looking to find a difference between the behaviour of groups (an experiment) or for a relationship between two variables (a correlation)
- the level of measurement

Experimental design

This section will briefly remind you of the experimental designs available, but for more detailed information, including an evaluation of each design, see M5 Experimental Design.

An **independent groups** design uses different participants for each condition in the study. A **repeated measures** design uses the same group of participants for each of the conditions. A **matched pairs** design uses different participants for each condition, but they are matched into pairs (or equivalent) based on important participant characteristics.

Each of these three experimental designs look to find *differences* between two variables, usually two groups of people in an experiment. A fourth design, although not an experimental design, is the **correlation design** which is used to try and find a *relationship* (correlation) between two variables of equal importance.

Levels of measurement

In AS Psychology, there are only three levels of measurement:

- **nominal data** is the lowest level of measurement
  data sorted into categories or sets, there are no scores or numerical values, for example, responding "yes" or "no" to a questionnaire
- **ordinal data** is the middle level of measurement
  ranked data – i.e. that that can be put into a hierarchy, e.g. rank out of 10
- **interval data** is the highest level of measurement
  *also called **ratio data**, although not exactly the same, they are considered to be the same for the purpose of your course, these are mathematical data – so statistical figures which have numerical values*

Choosing the correct statistical test

These factors must all be taken into consideration. Before you can choose which test is to be used, you *must* know the experimental design (or if it is a correlation design) and the level of measurement. The table below explains when you would use each test.

	Spearman's Rank	Mann Whitney U	Chi-squared
Experimental design:	correlation design	independent groups	independent groups
Testing for:	a relationship	a difference	a difference
Level of measurement:	ordinal or interval	ordinal or interval	nominal

Degrees of freedom

The only other component of statistical tests that you need to understand are *N* and *df*. You will already know by now that *N* is just the number of participants (or scores). In the Mann-Whitney U test there are two different *N* values.

The letters *df* stand for **degrees of freedom**. Whilst the Mann-Whitney U test and the Spearman's Rank test use *N*, the Chi-squared test of the Learning Approach does not, it uses *df*. This refers to the number of values in the final calculation that are free to vary. This is found by using the formula:

$$df = (\text{number of rows} - 1) \times (\text{number of columns} - 1)$$

for a table of data the Chi-squared test is being used for. You will come across degrees of freedom in a while.

Chi-squared test

We know by now that the Chi-squared test is chosen when the study uses an independent groups design, tests for a difference and has nominal (i.e. in categories) data. The experiment will use more than one condition on separate groups of participants. The example we will use for our Chi-squared test is people making donations. People entering a cathedral might be observed to see if they are more likely to make a donation if they believe someone is watching them. So the independent variable (IV) is whether the participants enter being observed or think they are unobserved and the dependent variable (DV) is whether or not they make a donation.

A Chi-squared test uses a contingency table called a **two-by-two table** which pairs the various conditions. It is common for a tally system to be used during the experiment. The two-by-two table for our study may look like the one below:

	observed	**thought unobserved**
donated		
not donated		

Each person would fall into one of the four conditions: they either donate knowing they are observed, donate unaware of being observed, don't donate knowing they're observed or don't donate not knowing they're being observed. The categories must be **mutually exclusive**, meaning they cannot fall into more than one of the categories.

The following instructions outline the Chi-squared test. The test is designed to see if the combinations of categories occur more frequently than would be expected by chance. It looks for an association between each combination of categories and predicts the value that would be expected for each cell in the contingency table. It then calculates how different the observed values were from the expected values, providing a probability for this occurring randomly.

Step 1: Put the raw data into a contingency table

	observed	**thought unobserved**	**TOTAL**
donated	50	30	80
not donated	20	60	80
TOTAL	70	90	**160**

Step 2: Transfer the totals for each cell into a table

Category	**Observed value (O)**
observed, donated	50
unobserved, donated	30
observed, not donated	20
unobserved, not donated	60

Step 3: Calculate the **expected value** of each cell from the original contingency table
The expected value is calculated by multiplying the row total by the column total of the cell in question, and then dividing that number by the grand total. For example, the expected value for "thought unobserved" and "not donated" would be the row total (80) multiplied by the column total (90), then that figure divided by the grand total (160) which would give us an answer of 45

Step 4: Minus the expected value from each observed value

Step 5: Square the figure for each answer of Step 4

Step 6: Divide the answers from Step 5 by the expected value

Category	Observed value (O)	Expected value (E)	O – E	$(O – E)^2$	$\frac{(O – E)^2}{E}$
observed, donated	50	(80 x 70) / 160 = 35	15	225	6.43
observed, not donated	30	(80 x 90) / 160 = 45	-15	225	5
unobserved, donated	20	(80 x 90) / 160 = 35	-15	225	6.43
unobserved, not donated	60	(80 x 90) / 160 = 45	15	225	5

Step 7: Use the formula for Chi-squared to calculate the answer of the Chi-squared test
The formula for Chi-squared is as below:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

This just means totalling all of the values in the last column of our table above. So in our example, $\chi^2 = 22.86$

Step 8: Calculate the degree of freedom for your study
This is the number of rows – 1 multiplied by the number of columns – 1, so (2 – 1) x (2 – 1) = 1

With the answer to the Chi-squared test and the degree of freedom calculated, we can then use the statistical tables to assess whether or not our findings are statistically significant. An example of one is shown below.

		Level of significance for one-tailed test			
		0.05	0.025	0.01	0.005
		Level of significance for two-tailed test			
		0.10	0.05	0.025	0.01
df =	1	2.71	3.84	5.41	**6.64**
	2	4.60	5.99	7.82	9.21

The Chi-squared test can be used to test either one-tailed or two-tailed tests, as the table shows.

The critical value must be equal to or less than our Chi-squared value (i.e. the result of the Chi-squared test must be the same as or more than the figure seen in the table).

Our hypothesis was two-tailed, as we did not specify which way the outcome would go, we merely hypothesised that the idea of observation would affect the donations being given. So we use the level of significance for a two-tailed test.

If we wanted to test to see if our findings are statistically significant for $p \leq 0.01$, we could do so. The critical value for a two-tailed test at $p \leq 0.01$ is 6.64 (remember the degree of freedom is 1). This has been highlighted in the table above. Our value was 22.86, which is certainly higher than 6.64, and so we can reject the null hypothesis for $p \leq 0.01$ and accept the alternative hypothesis.