Key for this study notes
SS: Sum of Square
BG: Between-Group, Between-Subject, Independent-Group
WG: Within-Group, Within-Subject

1 Lec 1: Introduction

1.1 Level of measurement

Nominal

- Variable with values that are names or categories (that is, they are names rather than numbers)
  - Nominal comes from the idea that its values are names
  - Variable in name only. category, number don't necessary mean anything, just a category, e.g. religion, gender (1=male, 2=female)
  - Doesn't denote anything about the relative magnitude

Ordinal/Rank-order variables (in order only)

- numeric variable in which the values are ranked, such as class standing or place finished in a race.
- numeric variable in which values correspond to the relative position of things measured
- difference in magnitude implied, No set magnitude between the 2
- not equal intervals between ranks
- group has order, e.g. race, 1st 2nd 3rd, still a category 1st (10 seconds) 2nd (11 secs) 3rd (14 secs), magnitude
- ranks: e.g., place in class, order in a horse race
- e.g. GPA between being 2nd and 3rd in the class could be different to 8th and 9th

Interval

- variable in which the numbers stand for approximately equal amounts of what is being measured
- numeric variable in which differences between values correspond to differences in the underlying thing being measured
- has magnitude
- difference in magnitude implied
- equal intervals are assumed
- e.g., time elapsed, temperature, ages, GPA, weight, stress level
- e.g. GPA 2.5 and 2.8 means about as much as the difference between a GPA of3 and 3.3

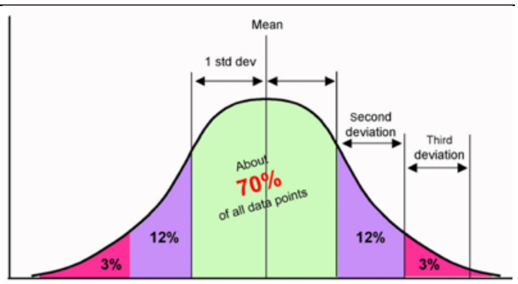
Ratio

1.2 Central tendency

Mean: arithmetic "average"
Median – mid-point
Mode – most common value

1.3 Normal distribution

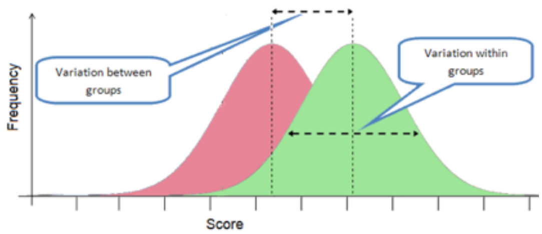
We know what the population average is, and what people vary around the average



1.4 ANOVA

ANalysis **O**f **Va**riance

- aka F-test
- Variance
  - Comparing 2 groups of people = comparing 2 different probability of distribution
  - ANOVA is testing, is there also variance between the groups in terms of their mean as is scaled by the variance within the group. Does it exceed certain amount?



- For simple (one-way) ANOVA: what is ratio of the variability between 2 group means divided by the variability of the WG variation
- Simply a ratio between 2 different types of variability
- To test if they are different to each other

$$F = \frac{\text{between-group variability}}{\text{within-group variability}}.$$

1.4.1 F-test vs T-test

- T-test: test the null hypothesis. In other words – is there a 'significant effect' in my data?
- So, – in fact, $F = t^2$
- but the ANOVA is more flexible, difference more than two groups, two dimensions
- Both the T-test and the ANOVA are specific cases of the General Linear Model (a powerful analytic tool)

1.4.2 Flexible and powerful tool

- How you use it depends on the type of question you want to ask, and ultimately this is intimately related to your **research design**
- In order to be successful, you have to make decision about how you are going to do your analysis as part of the process of designing the whole research methodology

"The general who wins a battle makes many calculations in his temple before the battle is fought. The general who loses a battle makes but few calculations."
— Sun Tzu, The Art of War

1.5 Where does knowledge come from?

These are some traditional ideas (prior to development scientific/hypothetical deductive method) about where knowledge might come from. But can we rely upon them?

- **Authority**: someone with authority that you can trust. What is a reliable source?
  - Law, Professor, Newspaper??, Hitler
- **Intuition**: just come to you, internally generated and it seems good
  - "gut feelings" - the ability to acquire knowledge without inference or the use of reason
- **Rationality**: You deduce it from the application of logical principle
  - Pure reason – the truth can be derived from first principals using logic
  - What about distorted logic? Witches burn therefore they are made out of wood, Wood floats, Ducks float, Therefore if the woman weighs the same as a duck she must be a witch.
- **Empiricism**: you saw it and measure it
  - Seeing is believing
  - Ames room is a distorted room that is used to create an optical illusion.

1.6 Scientific method

Sometimes also referred to as the **hypothetico-deductive method.** It is characterised by the development and systematic testing of **theories**.

The Method involves aspects of

- Authority: trust scholarly journals as a reliable source of information
- Intuition
- Rationality: logical rational thinking in terms of generation of hypotheses, and the structure of testing hypotheses
- Empiricism: actual gathering of evidence

Scientific method uses what is good about the above aspects, but maintain a degree of scepticism too.

It is characterised by the development of **theories** which have **explanatory** and **predictive** capacity and which must be **testable** and **refutable**

1.6.1 In a nutshell

You have a **theory** that attempts to explain a particular phenomenon of interest
That theory is used to generate **hypotheses** – if theory X is true, it follows logically that Y should occur
You test the **hypotheses**

1.6.2 Hypotheses, Theories and Laws

A **hypothesis** is a statement that can be tested. So the statement, "A watched pot never boils," is a valid scientific hypothesis because we can test it (and find that in this case it is NOT supported by the evidence).
A **theory** is a general principle or body of principles that has been developed to explain a wide variety of phenomena. It must be consistent with **known observations** and it must have **predictive**

power. As new knowledge is gained, theories are refined to better explain the data.
- Must allow you to make new prediction/hypothesis, which you can test. Then modify your theory based on the test

A **law** is a mathematical relationship that is consistently found to be true. E.g., one of the most famous laws in physics is Einstein's e=mc^2.
- There is no law in psychology, it has theory and hypothesis. So, not too worry

1.7 Theory

Induction and Deduction

1.7.1 Induction/Inductive Reasoning

- Reasoning from the specific to the general
- Taking some specifics examplas of category of things, than assume all these characteristics hold true across all these examplas
- E.g. Rainbow lorakeets, penguins and eagles all have feathers and beaks. Therefore, to induce from this, all birds have feathers and beaks.
- Making induction on things that some birds have in common, are all things birds have in common
- All the swans we have seen so far are black in Australia, therefore all swans are black (not true)
- Induction reasoning: you are hoping to generalize what you have observe

The logic of discovery: Theory Development
- Induction can be useful. Kind of logic we use when developing theory. You can have a theory based on inductive principle from the observations. It can go wrong, but it is an important part of scientific process for developing theory.

1.7.2 Deduction/Deductive Reasoning

- Opposite direction to deduction
- Reasoning from the general to the specific, using logical chains of reasoning – **syllogisms**
- E.g. all birds have feathers and beaks, rainbow lorakeets are birds, therefore rainbow lorakeets have feathers and beaks.

The logic of justification - Theory testing
- to test the hypothesis

1.7.2.1 Used in

- Developing hypotheses
  - If X is true, then Y should occur
  - Hypotheses should be logical consequences of the theory
    - If you have a theory that “the heart is the seat of love” a hypothesis that follows logically from this might be that if you remove someone’s heart they will no-longer be able to love, but this also means the person is dead. Need to be careful how we test hypotheses
- Testing Hypotheses

1.7.3 The Scientific Method assumes

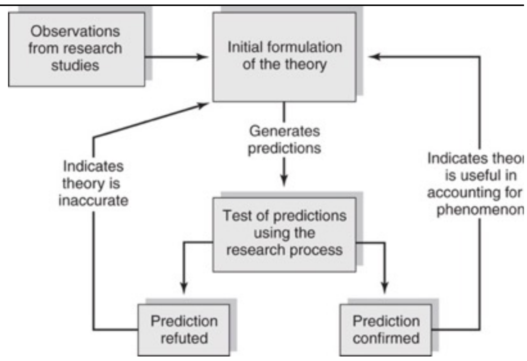
- That the universe is **ordered**, there is a structure in the universe, there is an underlining principle
- That order/structure which exists is **discoverable**

1.7.4 The Method is characterised by

- Control
- Operationism
- Replication (thing that you can show to be true and continuously showing them to be true. It is only by doing things a number of times, and be replicable demonstrateable, then we can begin to believe in them)

1.8 The Method in a nutshell

- You have a **theory** that attempts to explain a particular phenomenon of interest
- That theory is used to generate **hypotheses** – if theory X is true, it follows logically that Y should occur
- You test the **hypotheses**, then try to refute the theory
- If necessary you update your theory to accommodate the new empirical findings



The diagram is cyclical. Hopefully by having this cyclical process, we converge on knowing more and more the truth

1.9 What makes a good theory

- Theory is where prediction come from, and explains phenomena
- From the perspective of the Scientific Method
- Are these good theories?
  - Aristotle's *Theory of Geocentrism* (Theory of the world/earth is the centre of everything (the universe), the sun/star/planet all revolve around the earth)
    - Good theory but incorrect
    - Testable, therefore refutable. It generates the theory you should be able to make observations of the orbit of planets around the earth
  - Sigmund Freud's *Psychoanalytic Theory*?
    - Not a good theory
    - It doesn't generate predictions, therefore can't be refutable. On the basis on having some willy thoughts about super ego. It might be a good cultural theory, it is not a strong testable scientific theory.
  - Charles Darwin's *Theory of Evolution*?
    - Good theory and correct, have received overwhelming support
    - It generates prediction like evidence of common ancestors in the fossil records, testable and refutable.

Whether something is a good theory or not, isn't the same as asking whether you like it, isn't the same as asking whether is correct or not.

1.10 A Scientific Theory Must Be Testable

- Science proceeds by making observations of nature (by performing experiments). If a theory does not generate any observational tests (or predictions), there is nothing that a scientist can do with it. It has to generate hypotheses

Consider this theory: "Our universe is surrounded by another, larger universe, with which we can have absolutely no contact."
- Is not testable, therefore is NOT a good theory (could still be true….)

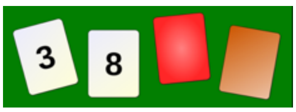
1.11 A Scientific Theory should be Refutable

Consider this theory: "There are other inhabited planets in the universe."
- This Theory is testable (we can go to another planet and see), but it is not a "good" scientific theory (not refutable).

Here's why. It may be either correct or wrong. If it is correct, there are several ways that its correctness could be demonstrated including:
1. we visit another planet and find Morbo living there.
2. radio telescopes on earth begin to receive signals from somewhere in the Andromeda Galaxy that appear to be reruns of the "I Love Morbo" show.
3. Morbo lands in your backyard and says, "I will destroy you puny Earthlings!"

But, so far this has not happened

1.11.1 The logic of Refutation



The WASON Card Selection Task
The rule is: if the card has an even number on one side, the other side must be red. Which card(s) must you turn over to test if this rule is TRUE?
- 3/8
- red/8
- 3/red
- 8/brown (not the answer people intuitively think, hence why one needs to be careful about the application of logic)

1.11.1.1 Intro to logic

A **syllogism** is a logical chain of argument, is generally structured like this:
1. A statement that declares a **rule**
2. A statement that describes an **observation** that relates to that **rule**
3. A **conclusion** that follows from that observation in the context of that rule

e.g.
1. If it is Thursday at between 10-12 there is a Research Methods Lecture
2. It is around 11:30 on Thursday
3. Therefore there is a Research Methods Lecture

- In what follows, don't get hung up on whether you think the **"rules"** are true or not.

- What is important is to understand that the **structure** of some types of argument are logically valid
- That means, **if** the rule were TRUE **then** the conclusion must also be TRUE
- Other structures or **NOT logically valid**.
- That means, even if the rule is true we do cannot trust the proposed conclusion to be true

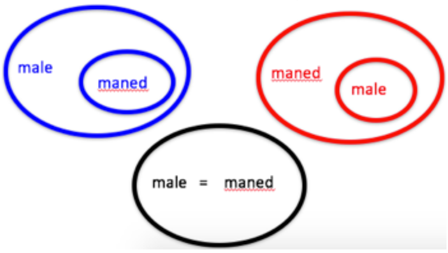
1.11.1.2 An example of a valid logical inference

Here is an example of a **valid** logical structure (modus ponens)
If P then Q
P
Therefore Q (valid structure)

If MANED then MALE
MANED
Therefore MALE (valid structure)

An important thing to note here is:
If MANED then MALE $\neq$ IF MALE then MANED

<u>MANED then MALE</u>: Maned lion is the subset of male lion (male child lion does not have maned)
<u>MALE then MANED</u>: Male lion is the subset of maned lion
Above two is not the same thing.
Male = Maned is another possible world

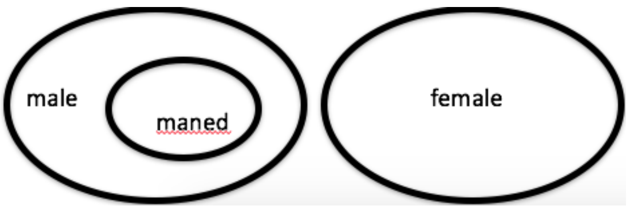


1.11.1.3 Another valid logical inference

Here is an example of a **valid** logical structure (modus ponens)
If P then Q
Not Q
Therefore Not P

If MANED then MALE
Not MALE
Therefore not MANED

<u>MANED then MALE</u>: Maned lion is the subset of male lion (male child lion does not have maned)
Male and female don't intersect, then, if it is not male, then can't be maned.
Another valid argument



1.11.1.4 Denying the antecedent

Denying the antecedent, is a **formal fallacy** of inferring the inverse from the original statement. It is committed by reasoning in the form:

~~If P, then Q.~~
~~Not P.~~
~~Therefore, not Q.~~

~~If EVEN then RED~~
~~Not EVEN~~
~~Therefore not RED~~

	~~If MANED then MALE~~ ~~Not MANED~~ ~~Therefore, not MALE~~ In other words – don't turn over the THREE – it doesn't help The rule is: if the card has an even number on one side, the other side must be red. Which card(s) must you turn over to test if this rule is TRUE? ~~If EVEN then RED, Not EVEN, Therefore not RED~~
1.11.1.5 Affirming the consequent	Affirming the consequent, is a formal fallacy of inferring the converse from the original statement. The corresponding argument has the general form: • Backwards • Fallacy of thinking ~~If P, then Q.~~ ~~Q.~~ ~~Therefore, P.~~ ~~If EVEN then RED~~ ~~RED~~ ~~Therefore EVEN~~ ~~If MANED then MALE~~ ~~MALE~~ ~~Therefore, MANED~~ In other words – don't turn over the RED – it doesn't help The rule is: if the card has an even number on one side, the other side must be red. Which card(s) must you turn over to test if this rule is TRUE? ~~If EVEN then RED, RED, Therefore EVEN~~ Turning the 3 over can't disprove it, it can't even prove it. Turning the red over, whether it's even or odd number, it doesn't help as the rule doesn't tell you anything about that. The logical structure is incorrect. So the correct answer is that we should turn over the: The EIGHT - Since if it the other side is Brown this would **refute** the "rule" The BROWN - Since if it the other side is EVEN this would **refute** the "rule"
1.11.1.6 Refute it	The fundamental point is that the way to TEST the rule is by trying to **refute** it rather than trying to **prove** it Karl Popper – all the white swans in the world cannot prove the theory "all swans are white" – but a single black swan can disprove it
1.12 Thus it is with science…	Science progresses by systematically eliminating falsehoods rather than demonstrating truths! And on that bombshell…

2 Lec 2: Experimental design, Variables and Operationalisation

2.1 Objectives of Psychological Research	To develop theories that **Describe** – portraying the phenomenon accurately • e.g., Piaget's theory of child development arose from detailed observations of his own children • describe accurately the phenomena that are of interest **Explain** – identifying the cause(s) of the phenomenon – posit explanatory mechanism – causal relationship between things • e.g., social connection and depression **Predict** – identifying risk factors of a phenomenon can help you to predict when it might happen – generate new prediction • e.g., what factors best predict academic success
2.2 How to Conduct Research	• Identify **phenomena of interest** (that interest us), that describing, explaining and making prediction about • Read the **scientific literature**, has anyone else had anything sensible to say something about ▫ Is there an established **theory** that generates **predictions** about the phenomena, that are testable? ▫ If not, what **evidence** is needed to allow a **theory** to be developed. ▫ If there are competing theoretical perspectives, ask what **evidence** is needed to establish which **theory** is correct/the best? • Formulate a **research question** • Identify best **method** to address the Research Question
2.2.1 Major Methodological Approaches (Dana interpreted Pat as saying that we don't need to know the term Positivist, etic etc just need to know the terms Quantitative & Qualitative, still need to know what they mean)	Quantitative • **Positivist** or **Etic**: Concerned with uncovering generalizable pattern and laws based on objective empirical data (tends to be deductive in nature) Qualitative • **Interpretivist** or **Emic**: Concerned with subjective interpretation, personal/cultural meaning, context specific, not concerned with generalisability but with deep understanding in line with inductive approaches. *tip to remember: - Etic, t for theory, Emic, m for me (interpreting) - Deductive sounds like reductive, reducing from general to specific - Inductive is to increase, specific to broader theory
2.2.2 Categorizing Research Approaches	**Quantitative** versus **Qualitative** Research Quantitative Studies – collect numerical data, or data that can be considered in numerical data – e.g., ratings of attractiveness, number of times a rat presses a bar in order to be rewarded, reaction times, people responses to surveys Qualitative Studies – collect non-numerical data to answer research questions, relate more to people's experience, understanding and personal meanings – e.g., pictures, clothing worn, interview statements, documents Mixed Methods – quantitative data provides an incomplete analysis of what is being investigated, numeration of phenomena – qualitative data adds additional level of understanding, layer of meanings
2.3 Quantitative	They generally work like this • You have a **hypothesis** • You collect some kind of **numerical data** to test that hypothesis
2.4 The Variable – a key	• **Variable**

concept in Quantitative Research

- something that varies
- takes on **different values** or **categories**
- e.g., gender, anxiety levels, IQ scores, on/off, heights, weights, these are all things that vary. We can numerate or categorise their level of variability

- **Categorical** versus **Continuous** Variables
  - Categorical Variables
    - varies by type or kind e.g., gender, religion, university course, type of therapy
    - e.g. 75% enrolled in psychology and 25% in law, it's categorical, one thing or the other.
    - NOMINAL MEASUREMENT
  - Continuous Variables
    - varies by degree or amount
    - Continues graded spectrum of values of a particular variable
    - e.g., reaction time, height, age, anxiety level
    - INTERVAL/RATIO MEASUREMENT

2.4.1 Which of these are variables

Interval – IQ (ratio is unknown)
Ratio – 0 (meaningful 0, means no negative number)

Variable ?	**Correct**	**Type**	**Scale**
~~Male~~	Gender	categorical	nominal
Weight		continuous	ratio (size of intervals is equal to each other, and it has meaningful 0 (no negative number), that means if something is weight 20kg, it is exactly twice as heavy as 10kg. So the different points of the scale has a meaningful relationship to each other)
Reaction time		continuous	ratio
~~6 foot 2~~	Height	continuous	ratio
~~blue~~	Colour	categorical	nominal
IQ		continuous	**interval**, the points between points on the scale are assumed to be equal and meaningfully so but there is no absolute 0 from which can be calibrated. So whilst the points on the scale are assume to be equal size interval, the ratio between them are unknown.

2.4.2 Variables in Quantitative Research

- Independent Variable (IV)
  - presumed to **cause** changes in another variable
  - the varying of IV leads to changes in DV
  - often **manipulated** by the researcher
    - therapy vs. no therapy
    - alcohol dose (1 unit versus 2 units)
    - location of learning word list (under water versus above water)
      - need to see if these changes affect the outcome

- Dependent Variable (DV)
  - the presumed **effect** or outcome of the study
  - variable that is **measured** by the researcher and influenced by the IV
  - is the thing we measure, we hope has been influenced by the manipulation of the IV
  - essentially is anything that you measure in the experiment
    - behaviours, attitudes, feelings measured through tests, monitoring, questionnaires, number of items recalled on memory task, reaction time, EEG data

2.4.3 Fundamental question

So the question that is generally asked in a quantitative research study is:
- are changes in the **IV** associated with changes in the **DV**?
- Or does changing the **IV** cause changes in the **DV**?

2.4.4 Other variables in quantitative research

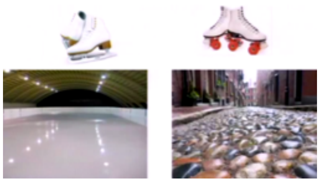
2.4.4.1 Extraneous Variables

- variable/s that competes with the IV in explaining the outcome or DV
- all of the things (you can/can't imagine) that might impact upon a person's ability to perform a task
- it is important to try to control for extraneous variables, to not allow it to be systematic variability as a function of extraneous variable

Is ice-skating faster than roller-skating?

- The thing we are interested in is what type of skates are being used in this speed test. So what are the things that might impinge the outcome?
- What kind of extraneous variables might be important to consider here?
  - The experience of the skater (uncontrolled extraneous variable)
  - Environment
  - Time of the day
  - Weights difference
  - Everything that can vary and has an impact to the outcome of the study

2.4.4.1.1 *Confounding variable*



An extraneous variable that is allowed to co-vary (to vary together with another variable) along with the levels of the IV

Is ice-skating faster than roller-skating?

- Found individuals who are confident and equally experience in the use of both type of skate, we have equipped them with the best possible skate. They are trained to peak level of fitness
- However, the condition has a systematic confound because both the skates (IV) and the course differ across the tests, in a way that is totally correlated.

Having a **confound** is pretty serious because it means that you really cannot tell whether it is the IV or the confound that is affecting performance.

- Uncontrolled 3rd variable is operating. If 2 variables are confounded, they are intertwined so you cannot determine which of the variable is operating in a given situation

2.4.4.2 *Mediating Variable / Intervening Variable*

- occurs between two other variables in a causal chain
  - e.g., anxiety **causes** distraction (mediating variable) which **affects** memory
  - distraction has the proximal effect on memory performance, not anxiety.
  - something that intervenes between one thing and another thing

2.4.4.3 *Moderating Variable*

- qualify a causal relationship as dependent on another variable
- qualify a causal relationship between IV and DV
  - e.g., the impact of anxiety on memory is different for men and women (sex is a moderating variable)
  - gender is moderating the effect of a relationship between anxiety and performance

2.5 The research problem/question

A good theory generates **hypotheses** – these predictions give rise to the research problem, or research question:

- an interrogative sentence that states the relationship between two or more variables or the key research question
- criteria for good research problems
  - variables should express a clear relationship
  - stated in question form
  - capable of empirical testing

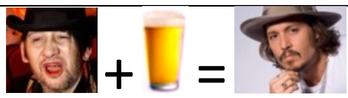
So a research question should be, specified in a way that makes clear what causal relationship is being tested.

- is number of **hours of CBT** associated with reduced **anxiety scale scores**?
- are changes in the **IV** associated with changes in the **DV**?

2.6 Work through an example

- Identifying an interesting phenomenon
- Relating it to theory
- Generating a hypothesis
- Framing a research question
- Identifying variables
- And considering different methods for addressing the research question

2.7 An interesting phenomena

Shane Macgowan ≠ Jonny Depp, but 

2.7.1 Is there a theory?

The "Inverse Cinderella" theory

	• Cinderella theory says things turns into pumpkin at midnight
	• Inverse Cinderella theory: everyone gets more attractive when the clock strikes midnight
	
	What's wrong with this theory?
	• We just need to look at people before and after midnight to know this is not true – easy to disprove.
	
	The "Beer Goggles" Theory
	• The ingestion of alcohol has a number of effects on the human brain including simultaneously increasing levels of sexual desire and decreasing aesthetic judgement with respect to the suitability of potential sexual partners
	• Shane + alcohol = Johnny Depp
	
	Is this a good theory? Does it generate predictions? Is it refutable? YES
	• Consistent with existing observations
	• Generates predictions (if you give someone alcohol, their judgement might change)
	  ○ testable
	  ○ refutable

2.7.2 Theory / prediction / question	• Theory
	  – "The ingestion of alcohol has a number of effects on the human brain including simultaneously increasing levels of sexual desire and decreasing aesthetic judgement with respect to the suitability of potential sexual partners"
	• Prediction
	  – "drinking alcohol will make people more attracted to people whom they would normally consider unattractive"
	• Research Question
	  – "does alcohol consumption affect attractiveness judgments?"
	  – do changes in IV affects DV?

2.7.3 Research question	IV: Alcohol Ingestion - we can operationalise into the following • could be categorical - YES/NO • could be continuous – number of drinks DV Attractiveness Judgements • could be categorical • could be continuous
2.7.4 How do you answer this question?	• Design a study • Find some participants • Make some measurements • Analyse the data • Write a paper explaining what you have done
2.7.5 Some ways to answer research questions	1. Naturalistic observation: simply observe the behaviour, no manipulation, e.g. animals in nature 2. Correlational study: making measurement and asking there is a relationship between different measurement 3. Internet study: online 4. Field experiment: in natural environment but with manipulation (different to Naturalistic) 5. Laboratory based experiment

2.7.6 Possible study ideas

	In choosing how best to address our research question we need to ask		
Possible study ideas	Is it possible to do the thing that we want to do? (logistics)	Is it OK to do the thing we want to do? (ethics)	Will doing what we want to do tell us anything useful? (validity)
1. Go to a nightclub and watch what happens (Naturalistic Observation)	Yes	Yes	No
2. Go to a nightclub and ask some questions	Yes	Yes	No
3. Get on facebook, encourage your friends to get drunk, then post a picture of Shane MacGowan	Yes	Maybe	No

and see how many *likes* it gets (Internet experiment)			
4. Take Shane MacGowan to a nightclub, spike someone's drink, and see what happens (field experiment)	No (we don't know him)	No	No
5. Go to your lab and perform an experiment	Yes	Yes	Yes

2.7.6.1 *Go to a nightclub and ask some questions*

Name	Drank	Rating
Sandy	1 bacardi breezer	"yuk!"
Leslie	12 schooners	"phwoar!"
Gabby	2 lemon ruskis	"meh"
Ashley	8 bacardi breezers	"bring it on!"
Pat	3 white wines	"no chance!"
Tyler	6 schooners	"not bad!"
Drew	4 mineral waters	"are you nuts?"
Morgan	3 schooners	"probably not"
Wynn	5 double vodkas	"maybe"
Sydney	7 vodka and redbull	"definitely"

This study hasn't operationalised the measures very well. They aren't stated operationalize with a degree of rigor and specificity and accuracy with which we could really get meaningful things from

2.7.6.2 *A Key Characteristic of Scientific Research*

- Operationism
  - representing constructs by a specific set of definitions or operations
  - operational definition
    - defining a concept by the operations used to represent or measure it

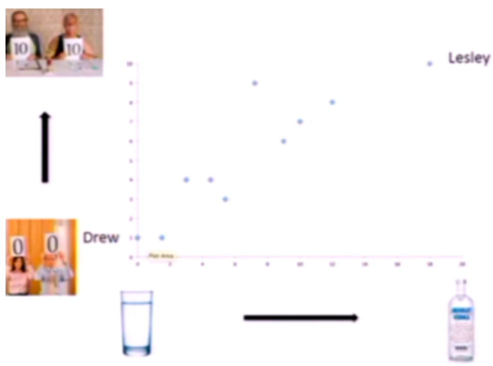
Name	Standard drinks	Rating
Sandy	1.5	"yuk!"
Leslie	18	"phwoar!"
Gabby	3	"meh"
Ashley	12	"bring it on!"
Pat	5.4	"no chance!"
Tyler	9	"not bad!"
Drew	0	"are you nuts?"
Morgan	4.5	"probably not"
Wynn	10	"maybe"
Sydney	7.2	"definitely"

How might we operationalise the concept of "alcohol intake"?
standard drinks, it is an operationalization of concept of level of alcohol intake. By a simple formula, we can transform out measurement into something operationally useful

2.7.6.3 *How about operationalising attractiveness?*

Out of 10 score / Attractiveness ratings

Name	Standard drinks	Rating
Sandy	1.5	1
Leslie	18	10
Gabby	3	4
Ashley	12	8
Pat	5.4	3
Tyler	9	6
Drew	0	1
Morgan	4.5	4
Wynn	10	7
Sydney	7.2	9



Now we have operationalize IV and DV, we want to see if there is a relationship between them.
Our data show that alcohol intake in CORRELATED with attractiveness ratings

2.7.6.4 *What valid inferences can we*

- ~~Alcohol impairs judgement?~~
  - We haven't tested judgement more generally, we have only tested a specific

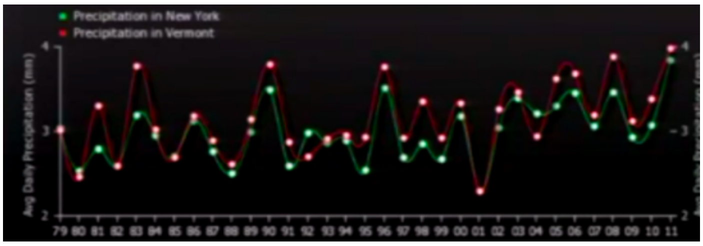
draw from this?

judgement

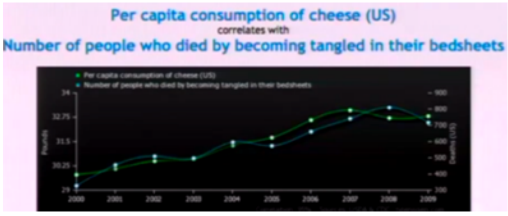
- ~~Alcohol causes people to lower their threshold for "sufficiently attractive"?~~
    - We can't say this. What we can say is,
    - People who had drunk more alcohol rated Shane MacGowan as being more attractive?

2.8 Correlation

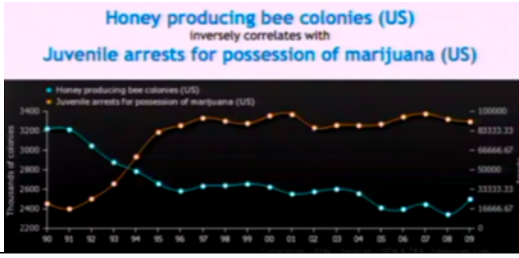
- Precipitation in New York correlates with precipitation in Vermont
- Is this one thing causing the other?
- These two things are highly correlated
- Because they are geographically close to each other and subject to same weather system



- The weather system causes the rainfall: the geographic proximal location of the regions moderates the variables/relationship



BUT, be careful of correlation
There might be a possible **mediating** variable, e.g. eating cheese may causes bad dream which caused the bedsheets tangled, but unlikely.



If one thing causes another thing they MUST be correlated
< does not equal >
~~If two things are correlated there MUST be a causal relationship~~ (random chance)

2.8.1 The issue of Causation

Causation

– a condition in which one event (the cause) generates another event (the effect)

Criteria for identifying a causal relation

– **cause** (IV) must be **related to** the **effect** (DV) (relationship condition)
– changes in IV must **precede** changes in DV (temporal order condition, cause must happen before effect)
– no other **plausible explanation** must exist for the effect
– we need these things to be true to infer a cause

the relationship between alcohol and Shane, we haven't established causality, because other explanation do exist. There are people really like alcohol and/or Shane.

2.8.2 Inferring Causality

A **well designed** and appropriately **controlled** and conducted **experiment** can allow **inferences about causality**

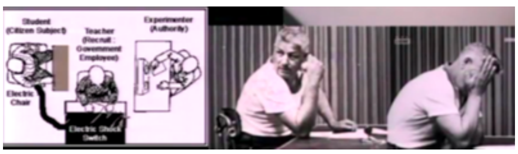
– Perform an action (**manipulate IV**)
– Measure the consequences (**changes in DV**)
– **CONTROL** for other possible explanations

2.9 An experiment should be…

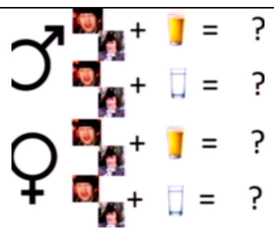
- Carefully designed
- Rigorously Controlled (try to control as many <u>extraneous</u> variables as possible, and avoiding <u>confound</u>, if we don't, we can't draw causal influences)
- Replicable (others should get the same results if copied the method and get the same results)
- Ethical

2.10 Some important ethical issues

- Informed consent (people should be asked, and consent to the participation of research)
- Right to confidentiality
- Right to withdraw
- Do not cause physical or mental anguish,



	harm/distress • Example of unethical experiment: Milgram (induce anxiety/stress to the participant, results can't be trusted)
2.11 Experimental Approach	
2.11.1 Advantages	• Causal inference – experimental approach is best method for inferring causation - causal description refers to identifying the consequences of manipulating an IV - causal explanation refers to explaining the mechanisms through which the relationship exists • Ability to manipulate variables - only scientific methodology in which variables are manipulated • Control - extraneous variables are controlled by: o holding them constant, e.g. same IQ o using random assignment o matching (method that are available to us that facilitates experimental control with which we can make causal inference, by being able to say this extraneous variable isn't the explanation as we have taken this control measure to counteract possible effects of this extraneous variable)
2.11.2 Disadvantage	• Does not test the effects of non-manipulated variables – many potential IVs cannot be directly manipulated • e.g., people's ages, gender • Artificiality or Generalisability – refers to potential problems in generalising findings from laboratory settings to the "real world" – people may behave differently in lab setting vs natural environment
2.12 Experimental Research Settings	
2.12.1 Internet Experiments	• advantages – access to diverse population – bring experiment to participant – large sample and thus greater power – cost savings • disadvantages – multiple submissions (from same person) – lack of control – self-selection – dropout
2.12.2 Field experiments	• an experimental research study that is conducted in a real-life setting – advantage – may be easier to generalize findings, cut out the artificiality of laboratory setting, therefore getting more real data on how people behaves – disadvantage – less control of extraneous variables, can be time consuming • confederate – use of deception, a person who is in league with the experimenter, unbeknownst to the participant – e.g. people are more generous and willing to give more money in the lab setting. People are less generous in real life, e.g. selling baseball cards at a convention, real life setting – less generous. – This is because in the lab, they feel the pressure of social judgement. They alter their behaviour to conform to what they think is the nice way to behave
2.12.3 Laboratory experiments	an experimental research study that is conducted in a controlled laboratory setting • advantage – more control over extraneous variables, e.g. same time of the day, temperature etc • disadvantage – less generalization related to artificiality (lab)
2.12.3.1 Different ways we could manipulate	***Experimental manipulation*** Experimenter determines which level of the **IV** a participant is tested at;

IVs	• event manipulation (e.g. presence of alcohol v absence of alcohol), complete control • instructional manipulation (e.g. drink alcohol quickly / slowly)
2.12.3.1.1 Beer goggles experiment 1	• IV: Drink Type: alcohol, water (alcohol vs non-alcohol) • DV: attractiveness of the picture of Shane
2.12.3.1.2 Beer goggles experiment 2	• IV: vary the standard of drinks: e.g. no drinks, one drink, 5 drinks • DV: attractiveness of the picture of Shane 
2.12.3.2 Different ways we could manipulate IVs	***Individual difference manipulation*** • Although we can't allocate people to be male/female, high/low IQ • **Quasi experimental manipulation** rather than true experimental manipulation ▪ **Quasi-experiments** are subject to concerns regarding internal validity, because the treatment and control groups may not be comparable at baseline. With random assignment, study participants have the same chance of being assigned to the intervention group or the comparison group. (Wikipedia) • We could try and look at the effects of individual differences across participants • Try to look at the effects of variables related to individual differences • A characteristic of the participant determines the level of the IV at which they are tested; – Computer anxious vs. non-computer anxious – Male vs. female – Level of social support received (high v low)
2.12.3.2.1 Beer goggles experiment 3 	Is there an effect for alcohol vs no alcohol based on individual's sexual preference? Whether the effects of alcohol on the attractiveness judgement are general that you will say everyone is more attractive whether you would consider them as a sexual partner or not vs whether it is moderated by whether they are the kind of gender people with who them wants to engage in sexual activity. These are sort of things we can start to make casual inference.
2.12.3.3 Different ways we could manipulate IVs	***Repeated Measure (Within Group)***: each participant tested at each level of the IV; • Same participant is contributing to more than one IV • More sensitive design (easier to detect the effect of interest) • Can't always use this design • When used appropriately, it is a really good method ***Between Group***: each participant tested at only one level of the IV; • Less sensitive design • Often forced to use this design ***Mixed Design***: • more than one IV with at least one IV manipulated BG • and at least one WG.
2.12.3.3.1 Beer goggles experiment 4	Multifactorial Beer Goggles Experiment 
2.12.4 Potential manipulations	• Alcohol vs no alcohol • Different doses of alcohol • Male vs female • Male vs female stimulus pictures • All of the above

3 Lec 3: Sampling, Validity and Reliability

How do you answer a research question?

- Design a study

- Find some participants
- Make some measurements
- Analyse the data
- Write a paper explaining what you have done

3.1 The issue of Causation

Criteria for identifying a causal relationship

- **cause** (IV) must be **related to** the **effect** (DV) (relationship condition)
- changes in IV must **precede** changes in DV (temporal order condition)

3.2 Find some participants

This is known as **sampling**

If we would like to be able to say that our data allow us to make **generalisable inferences** it is very important to **get this right**!

3.2.1 Some key terms

Population

- A group of people about whom one would like to draw some meaningful conclusions, e.g.
  - Adolescents
  - People with schizophrenia
  - QUT Psychology undergraduates

Sample

- A subset of that population that is actually included in your research study i.e. participants
  - 150 Year 10 students
  - 30 outpatients
  - Everyone who attends wk 3 lecture

Sampling frame

- A list of members/elements of a population from which one might obtain a sample
  - Electoral role
  - Telephone directory
  - Student enrolment list

Census

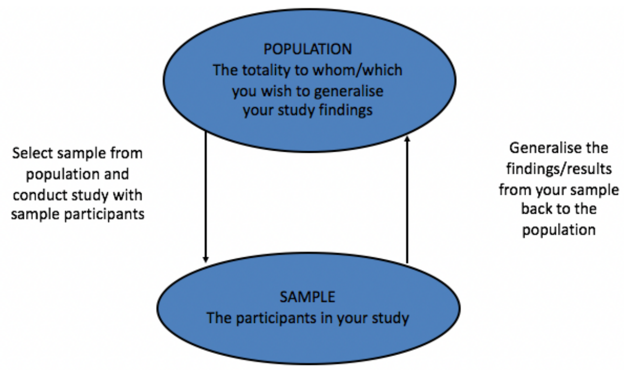
- A list of **all** the people comprising a particular population.
  - E.g. all the member of the AFL clubs

3.2.2 Aim of sampling

To make **generalisable inferences** about the **population** on the basis of measurements from your **sample**.

It is crucial that you have a **representative sample -** a sample that is like the population.

This simply means that you should select a **sample** whose **typical characteristics** are approximately **the same** as the **typical characteristics of the population**.



If you can't guarantee that this is so, you can't guarantee that your inferences generalise.

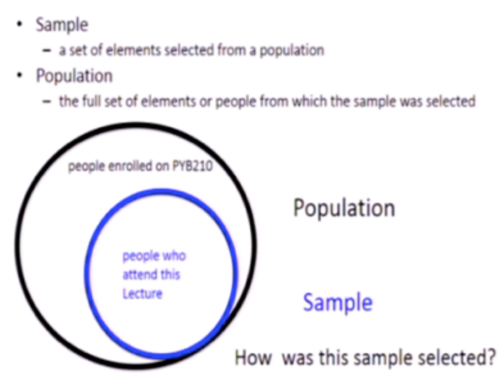
3.2.3 Representativeness

- **Sample Statistic**
  - A numeric characteristic of a **sample -** (measured)
  - Something that we measure in the sample
- **Population Parameter**
  - A numeric characteristic of the **population -** (often not known)
  - If we have a representative sample, then this sample statistics will be closely related to the population parameter what that value will be for the entire population
- **Response rate**
  - What proportion of people responded?
- **Sampling error**
  - The difference in value between the **sample statistic** and the **population parameter** (depends on sample size)

The smaller the sample, the larger the sampling error. If the sample is too small, it is not likely to reflect the characteristic of the population in general

3.2.4 Sampling bias

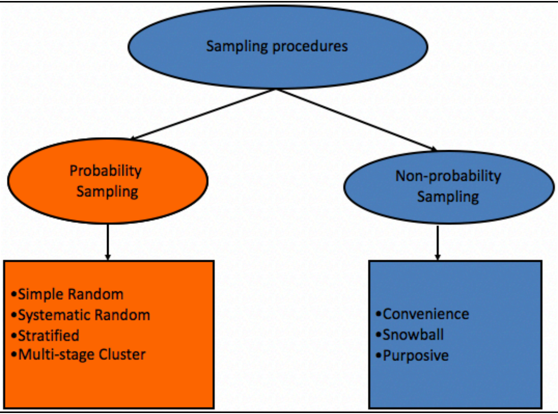
Population: People enrolled on PYB210
Sample: people who attend this lecture



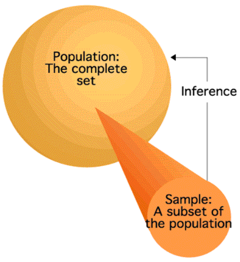
How was this sample selected?

- If it was random sampling, then students would tossed the coin when they got out of the bed, heads: go to uni, tails: back to bed
- In this case, people selected themselves to be part of the sample
- This is not a representative sample
- Self-selection: there is always a danger on people who select and who don't. Data can't trusted. Might there be **systematic differences** between people who do versus don't.
- The ones that are not in the lecture may have a fulltime job, child care responsibility etc
- Don't trust a self-selecting sample - an example of sampling bias

3.2.5 Sampling procedures



3.2.5.1 Probability sampling

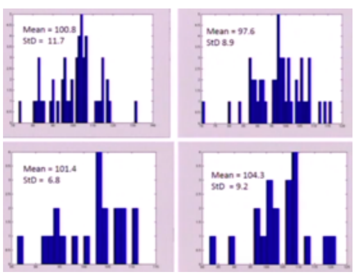


- E.g. Tossing the coin
- A way to ensure that your sample is representative of the population (on the characteristics deemed important for the study)
- Basic principle:
  - A sample will be representative of the population if all members of the population have an equal chance of being selected in the sample
  - Allows the researcher to calculate the relationship between the **sample statistic** and the **population parameter**
  - Everyone has an equal chance of being selected => representative of the population, providing you have large enough sample size

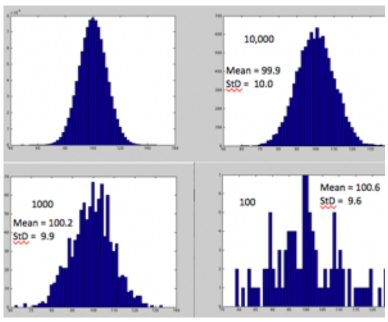
3.2.5.2 Sub-types of probability sampling

- o Simple random sample
- o Systematic random sample
- o Stratified random sampling
- o Multistage cluster sampling

3.2.5.2.1 Simple random sample



- Each member has an **equal** and **independent** chance of being selected
- Define the population, list all members, assign numbers
  - Use a table of random numbers to select, e.g. all odd numbers
  - Use a "lottery" method, pull names out of a hat
  - Use a computer program to randomly select
- Works well providing sample size is not too small



Example:
First is a histogram showing the IQ scores of a population of 1,000,000 people. The population mean is an IQ of 100 and the SD is 10 IQ points. Let's take some samples.

- Smaller samples => moves away from the population mean and SD
- Simple random sampling works really well provided that your sample size don't get too small.

3.2.5.2.2 Systematic random sample

Every **Kth** person

- Systematic is more historic when computer wasn't accessible to use for randomization.

- Randomly select the first person then divide the size of the population by the size of the desired sample, and use this to determine the interval at which sample is selected.
  - e.g., to select a sample of 1000 people from a list of 10,000, randomly select the first person and start the list with them - then select every 10th person from the list
- Need to ensure the list of elements is not arranged in a way that means systematic sampling could lead to a biased sample (e.g., student list in GPA order!).
  - e.g., different results if you start with the 2nd person and sample every 10th person beyond that than if you start with the 8th person and sample every 10th person
- Whenever people don’t have the equal and independent chance of being picked, you are introducing possible factors of things going wrong.

Which should we prefer? Simple, or systematic random sampling?

- Simple, less chance of anything systematic going on.

3.2.5.2.3 Stratified sampling

- If you want to make sure the profile of the sample matches the profile of the population on some important characteristics e.g. ethnic mix, gender.
- Divide population into subpopulations (strata) and randomly samples from the strata

Why use stratified sampling?

- When there is heterogeneity within the population, and you want to end up with a sample whose characteristics reflect the proportional heterogeneity of the population
- Can reduce sampling error by ensuring ratios reflect actual population (e.g., ratio of males to females)
- To ensure that small subpopulations are included in the sample

NB:

- can have proportional representation or disproportionate representation
- but disproportionate sample would not be used to generalise to entire population, only the subgroups

3.2.5.2.3.1 Simple Random Sampling Versus Stratified Sampling

Our population is “Animals of West Queensland Savannah” – a census reveals that the entire population consists of 60 lions, 30 tortoises and 10 rabbits.

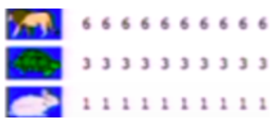
<u>Simple Random Sampling</u>

- Not a good inference of population level
- Because the stratification of the population hasn’t been reflected in the sample

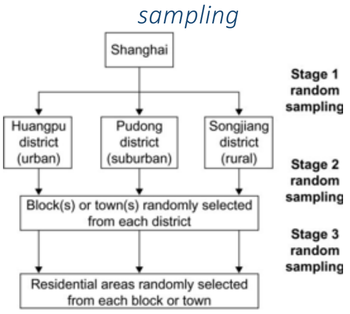
More example of Simple random sampling, sometimes we get it right

<u>Stratified Random Sampling</u>

Regardless how many times we do the stratified random sampling, we are always going to end up with this figure, reflects proportions in the population.



3.2.5.2.4 Multi-stage Cluster sampling

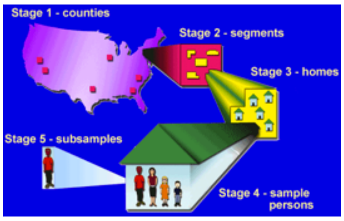


Begin with a sample of grouping and then sample of individuals

e.g. Rural sample

- Define rural townships as those with population < X
- Get listing of all relevant townships
- Take a random sample of townships
- Randomly sample people from within the randomly sampled townships

- If all the sample are from the same town, there might be something systematically different about that town, e.g. high with unemployment
- Better to select sample from bunch of different town, randomly selects the town, then randomly selects the people (multi-stage of processing going on)

	When might you use this? - When you have different region, different characteristics - e.g. Hunger game
3.2.5.2.5 Multi-Stage/Multi-Phase Sampling  	• Type of random sampling whereby • Larger sample obtained first • in order to identify members of a sub-sample • Sub-sample then randomly chosen from for study • Good (but costly) way to identify not readily identifiable subgroups E.g. using Australian Mental Health and Wellbeing Survey to identify people with psychotic illness • large scale survey • Need people with psychotic illness, “Low prevalence” (1% of population) disorders study • from AUS Mental Health and Wellbeing Survey to identify low prevalent disorders study • This is how (psychotic illness) people are randomly selected based on their previous involvement (Australian Mental Health and Wellbeing Survey)
3.2.5.2.6 Advantages of Probability Sampling	- No systematic bias - Helps overcome **sampling bias** - Ensures representativeness!
3.2.5.2.7 Problem with probability sampling	- access to list of people - costly - difficult - you can randomly select someone but there is no guarantee they will agree to participate in your study. - Is there a systematic difference between people who agree to participate and those that don't agree to participate? - **Self-selection**: asking people’s agreement to participate in research is a form of self-selection, and this can cause bias

3.2.5.3 Non-probability sampling	Not every member of the population has an equal chance of being part of the sample
	Why use then?
	- There are no lists for some populations under study,
	- Logistical or cost related problem
	- e.g.
	    o The homeless
	    o Certain occupations (e.g., farmers)
	    o Hidden populations (e.g., people involved in “clandestine” activities)
	    o Convenience / resource restrictions

3.2.5.3.1 Convenience Samples	- Most used in psychology
	- People happen to be available
	• A sample of available participants, e.g.,
	    – students enrolled in a particular course
	    – People passing a particular location
	        • Self-selecting, non-random
	        • systematic difference on who you might be exposed to, e.g. standing outside centrelink vs casino
	• Advantages:
	    – Easy, inexpensive
	• Disadvantages:
	    – No control over representativeness

3.2.5.3.2 Snowball Sampling	• Like a snowball running down the hill and gather more as it goes • Used mainly for hard to study sub-populations • Identify one member for the study, then asking for their friends to participate – e.g., Gay men, Homeless young people, Illegal immigrants • Involves collecting data with members of the population that can be located and then asks those members to provide information/contacts for other members of the population

	Problems - people tend to associate with people similar to themselves o e.g. hipster bearded guys, are bearded men more desirable than other men? Depending on the study, may or may not be a problem. - People have the same network of people who may just be like them, bearded men having more beard friends.
3.2.5.3.3 Quota Sample	- Non-probability sampling equivalent of a stratified random sample o U know there is strata within your population, and you want to reflect relative proportion of those different strata population in the sample - But you don't/aren't able to sample randomly from each strata as you do in stratified random samples - So you use non-probability sampling Problem - can't guarantee representativeness

3.2.5.3.4 Purposive/judgment sampling	- Clear purpose to the sampling strategy: select key informants, atypical cases, deviant cases or a diversity of cases.
	- Sampling in a way trying to find particular characteristic, to get particular information
	- Often used to:
	  – Select cases that might be especially informative
	  – Select cases in a difficult-to-reach population
	  – Select cases for in-depth investigation
	
	Examples:
	– Studying the problems experienced by new immigrants
	  – Interview key people involved in agencies that help immigrants such as ethnic welfare groups, community immigration legal aid groups
	  – Interviewing people with extensive experience with immigrants likely to provide rich data
	– Comparison of left-wing and right-wing students
	  – May not be possible to sample *all* left-wing and right-wing students
	  – Instead, you could sample the membership of left (e.g., Socialist Alliance) and right-wing groups on campus (e.g., young liberals)

3.2.5.4 Which Sampling Method?	- As a major aim of quantitative research is the ability to generalise results the ultimate method is a probability sampling one. Representative - However this is often not workable or feasible given resources, time, the specific target population. - Sampling method used should be fully explained to participants - and caveats about the likely generalisability of results made accordingly so that the reader can review your results in an informed way. - We will always have non-optimal sampling method as we can't just have the census of the whole population and select the sub-population from it. Therefore a research paper needs to state clearly what has been done and the problem associated with it.
3.2.6 How many people should you test?	**Sample Size** - as we have already seen the size of your sample can influence how representative it is of the **population** It is therefore important to ensure that your sample size that is appropriate
3.2.6.1 Determining Sample Size 1	How many participants do I need for my study? - Largely determined by the analysis you plan to conduct with the data derived. How are you going to treat the data? - Generally the more complex the analysis the larger the sample you require - Increases in sample size bring with them increases in accuracy/precision/reduces sampling error. - Greater heterogeneity of the population, greater variation in the population, the larger the sample should be to capture and reflect the heterogeneity in the sample size. - There are many texts which will provide you with sample size requirements for any

	given statistical test as well as calculation tools which will provide you with a sample size given a number of parameters. *Heterogeneity – being diverse in content.

3.2.6.2 Determining Sample Size 2	Larger sample sizes are needed if population is:
	- Heterogeneous
	- you want to breakdown the sample into multiple subcategories
	    ○ e.g., look at males and females separately
	- when you expect a small effect or weak relationship
	- when you use less efficient methods of sampling
	    ○ e.g., cluster sampling
	- for some statistical techniques
	- if you expect a low response rate
	
	if you are using not representative sampling method, then err on the side of having larger sample

3.2.6.3 Determining Sample Size 3	Five simple rules for determining sample size 1. if population is less than 100, use entire population 2. larger sample sizes make it easier to detect an effect or relationship in the population 3. compare to other research studies in area by doing a literature review 4. use a power Table for a rough estimate 5. use a sample size calculator (e.g., G-Power) - what sample size is needed for an effect of a particular size
3.3 Make some measurements	
3.3.1 Operationalisation of IVs and DVs	**Operationalisation of IVs** – How are you going to manipulate it? How is it manipulated (if you can't)? **Operationalisation of DVs** – How are you going to measure it? – What measurements are you taking? How might we measure intoxication? Example, DV: using alcohol consumption - Their looks, or their ability to walk straight line is not good enough - Breathalyzer and blood test may also not be good measurement for alcoholic, as their body is used to toxication.

3.3.2 Reliability and Validity	**Reliability**
	- Does our measurement <u>instrument</u> behave sensibly?
	- Does it always measure the same thing in the same way?
	
	Validity
	- Are we measuring what we think we are measuring?
	    ○ Are we measuring intoxication when we measure blood alcohol level?
	    ○ Does the blood measure gives the same result every time we use it?
	
	If we look at someone, is my view of his intoxication the same as yours?
	These are all questions we need to ask for reliability and validity.

3.3.3 Reliability and Validity	Applies mostly to indexes/scales – In psychology, we try to numerate the reliability and validity of things like questionnaire, survey How do we assess whether our measures/operationalisations are good? – Are they valid? – Are they reliable? Snag is that you can't assess these until AFTER you have developed your questionnaires and used them – This is why a pilot test can be so beneficial

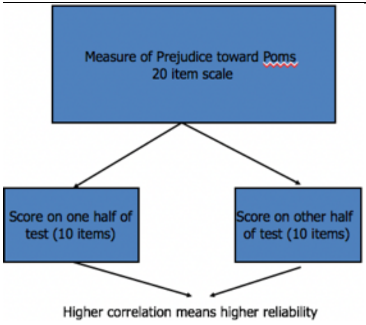
	– This is why many people chose to use established measures rather than develop their own and take the risk – E.g. use IQ test Examples - Not reliable / not valid: birthday and star signs - Reliable / not valid, judge people's intelligence based on their looks (we think people wear glasses are smarter) - Measuring with a ruler (reliable, valid) - Not reliable / valid: not possible scenario, if you can't rely on your measurement, you can't trust its validity
3.3.3.1 The relationship between reliability and validity	Can a measure be reliable but not valid? – Yes! You could have a consistent measure that does not actually measure the construct Can a measure be valid but not reliable? – No! If your measure doesn't consistently and dependably measure the construct it cannot possibly be measuring what it says it's measuring Physical measurements clear and easy to see that they are reliably and valid i.e. we can see. Psychological measurements are a little bit more tricky.

3.3.3.2 Reliability	• The consistency or repeatability of the measurement
	    o Say I weight myself on some scales at one point in time and then weigh myself 5 mins later and it says I'm 5 kilos heavier.
	    o Conclusion: dodgy scale, don't use it.
	    o Scientific conclusion: the scales are an ***unreliable*** measurement instrument

3.3.3.3 Type of Reliability test	• Stability of the measure (**Test-retest**) • Internal consistency of the measure (**Split-half**, **Cronbach's alpha**) • Agreement or consistency across raters (**Inter-rater reliability**) • Across different people making the judgement or mends
3.3.3.3.1 Test-retest reliability	Does your test measure the same thing every time you use it? • Addresses the stability of your measure • Same answers every time • You administer the measure at one point in time (Time 1) • then give the same measure to the same participants at a later point in time (Time 2) • Hoping there will be a correlation between the two times • You correlate the scores on the two measures • If it is high correlation, then it has high test-retest reliability • If it is too low, then it is not worth using it as the test-retest liability is low

3.3.3.3.1.1 Problem with Test-retest	Imagine that you want to test whether giving people vitamin supplements can improves a persons IQ. Two main problems:
	1. **Memory effect**
	    – you might remember the questions and look up the ones you didn't know
	2. **Practice effect**
	    – Performance improves because of practice in test taking
	
	• If too short there's a greater risk of memory effects
	• If too long there's a risk of other variables (e.g., additional learning) influencing results

3.3.3.3.2 Split-half reliability: is your measure internally consistent	Psychology test is not simply a one question survey, e.g. are you an extrovert? Test will have a set of items. by endorsing a set of items, it will lead you high in, e.g. extrovert category. So in the test, there are set of sub-items that relates to the construct of extrovert, introvert. In order to define the personality. Each of these personality traits is assessed by different set of items. **Split-half reliability:** Are the different items constant to what they are measuring? • You administer a single measure at one time to a group of participants • But, for your purposes (of understand of psychometric quality of your experiment) you



split the measure into two halves. Odd item is going to pool A, even items is going to pool B.

- and you correlate the scores on the two halves of the measure (higher correlation means greater reliability)
- e.g. IQ test, 2 set of questions in one test. If both set have high correlation. This suggests 2 halves are measuring the same construct.
- This way, you don't need to do test-retest reliability, rather, you ask internal consistency of the test
- Strength: eliminates memory & practice effects
- Limitation: Are the two halves really equivalent?
    - Use Cronbach's Alpha (measure of internal consistency. It is considered to be a measure of scale reliability.)

3.3.3.3.2.1 Cronbach's Alpha

- Assesses the 'internal consistency' of your measure
    - i.e., tells you how well the items or questions in your measure appear to reflect the same underlying construct
- You would get good internal consistency if individuals respond in approximately the same way to questions on your survey
    - Different items of the same test measuring the same construct
- Mathematically it's the equivalent of the average of all possible split-half reliabilities
- Coefficient alpha can range from 0 to 1.00
    - The closer the alpha is to 1.00, the better the reliability of the measure

3.3.3.3.3 Inter-rater or inter-observer reliability

- Do different raters measure the same thing?
- Rely the judgement of the observers.
- Checking the match between two or more raters or judges
    - E.g. people observe the behaviour of young babies
    - coding videos for infant "looking time" – need to check the agreement amongst the coders
    - Coding the length of time an infant is looking at one particular object vs another
- There is a degree of subjectivity of interpretation in these kinds of measures.
    - Was the infant directly looking at the object or close to the object?
- Where there is a possibility for subjectivity, what people are interested in is the inter-rater reliability
    - If people are trained properly, different people should be highly correlated with each other with respect to the subject judgement they make
- Are the different people making the judgement behaving similarly in the set of judgement they are making?
    - High correlation = High degree of reliability

3.3.3.3.3.1 Calculation of inter-rater reliability

- nominal or ordinal scale
    - the percentage of times different raters agree
- interval or ratio scale
    - correlation coefficient

3.3.3.4 Validity

- Are we measuring what we think we are?
    - Is our measure credible, is it believable?
- Why is validity an issue?
    - For reliability, we can come with these clear measures of the degree to which measurement is reliable
    - For validity, many (if not most) variables in social research cannot be directly observed. You have to infer on the basis of something
        - e.g., motivation, satisfaction, helplessness
        - need to use instrument such as questionnaire
- The challenge:
    - We can quantify reliability
    - We can't quantify validity
    - to make a judgment call about whether we are measuring what we think we're measuring

3.3.3.4.1 Types of Validity

- Face validity