Experimental Design and Analysis Revision Notes

TOPIC 1: The scientific method in psychology

- The basis of psychology is to determine reasons behind human behavior
- Science involves establishing relationships between an event and a set of preceding circumstances

THE SCIENTIFIC METHOD

- The scientific method has four key aspects to it

1. **Confirmation** of findings (replication)
2. **Objectivity** (not to let biases get in the way)
3. **Self-correction** (ability to alter conclusions)
4. **Control** (making sure your findings are the best they can be)

THE EXPERIMENTAL METHOD

- This involves manipulating the preceding circumstances and observing the change in the final event
- This establishes a **cause and effect** relationship
- The basis of the experimental method is that all variables other than the dependent and independent are CONTROLLED

KEY TERMS

- **Independent variables:** these are what is manipulated by the researcher and are independent of the behavior
- **Dependent variables:** This is the behavior of the participant and is dependent on the state/level of the IV
- **Hypothesis:** is a statement of the relationship between the dependent and independent variable

WHY PSYCHOLOGY IS A DIFFICULT SCIENCE

1. There are many possible circumstances concurrent to, or preceding a behavior
2. The differences inherent in an individual

TOPIC 2: Research Hypothesis

HYPOTHESIS

- A hypothesis is often created following a simple observation
- Hypotheses, and resulting experiments, are also often "thought of" after reading about previous research in the field
- Next we want to test it! But in order to do so the experiment must consist of 3 characteristics:
  1. **Repeatable** (findings can be replicated)
  2. **Observable** (this and #3 involve the ability to quantitatively measure variables
  3. **Testable**

STATEMENT TYPES

- There are 3 different ways of formulating your hypothesis statements

  1. **Synthetic**
     - These are statements that can either be **true or false**
     - They are good for a research hypothesis
     - I.e. "Abused children have lower self-esteem"
     - Allows the research hypothesis to be expressed in the "genera implication form"
  2. **Analytic**
     - These are statements that are **always true**
     - They are bad for a research hypothesis
     - I.e. "I am making an A or I am not making an A"
  3. **Contradictory**
     - These are statements that are **always false**
     - They are bad for a research hypothesis
     - I.e. "I am making an A *and* I am not making an A"

INDUCTION AND DEDUCTION

- **Induction=** the taking of particular examples of a situation and creating a general, theoretical statement about that situation (the hypothesis)
- If the hypothesis is derived through an initial observation, then such a process is termed induction

PARTICULAR→GENERAL

PIG

(Particular **I**nduction **G**eneral)

- **Deduction=** used to test a hypothesis
- When you have a general case (the hypothesis), and you test it with the prediction of a particular outcome

GENERAL→PARTICULAR

GDP

(General **D**eductive **P**articular)

- We can never really say a theory has been proven
- If the experiment **does not** support the prediction, then the hypothesis must be rejected by this one observation alone!

TYPES OF THEORIES

1. **Descriptive theories**
    - Name/describe events
    - E.g. Freudian repression preventing disagreeable thoughts into consciousness (but are expressed in other ways)
    - However, these theories don't necessarily explain why these events occur or what causes them

2. **Analogical theories**
    - Attempt to explain a relationship using an example or analogy from another field of science
    - I.e. The pathological manifestations of repression are the equivalent to **pressure** building up in a vessel, and cracks appear in the weakest points
    - However, ultimately an explanatory limit will have been reached (humans AREN'T vessels under physical pressure!)
3. **Quantitative theories**
    - Explain a relationship using precise values and mathematical expressions
    - This type of theory is the objective of many psychologists

WHAT MAKES A GOOD THEORY

- A theory should account for most of the data (contradicting data tends to disconfirm the theory)
- Recall the theory should also be testable- those that aren't are sometimes called falsifiable, and are considered "pseudoscientific"
- A theory should also make predictions
- Descriptive, analogical and quantitative theories achieve this with different degrees of success

TOPIC 3: Basic statistical concepts

SCALES OF MEASUREMENT

- **Nominal**
  → A set of category labels, identifying some things as different or the same to other things (e.g. male/ female, football jersey numbers
- **Ordinal**
  → Orders a variable, however the difference between the orders doesn't necessarily mean anything (e.g. range of colors could be ranked in order of liking)
- **Interval**
  → Orders a variable and the difference between the orders has a legitimate meaning (e.g. temperature; the difference between 5

degrees and 10 degrees is the SAME as the difference between 10 and 15)

- **Ratio**
→ Orders a variable, the difference between the orders has a legitimate meaning and it has a true "zero point" (unlike temperature, where 0 degrees is assigned arbitrarily)

TERMS WE USE WHEN COLLECTING DATA

- **Population:** the entire set of "data" you're interested in collecting
- **Sample:** A subset of the population, which can be practically obtained. Samples are used to make **inferences** about characteristics of the population
- **Random samples:** best estimate the population characteristics
- **Inferential statistics:** estimating the actual population parameters (or qualities)
- **Discrete variables:** possess only a limited number of levels or states (e.g. male/female)
- **Continuous variables:** possess many different levels (e.g. height in mm)

HOW TO CATEGORICALLY EXAMINE SAMPLE DATA

- **Frequency distribution**
- **Stem and leaf plot**
- **Histogram**
- **Bar graph**

NOTATIONS USED TO DESCRIBE DATA

X=VARIABLE

Σ= THE SUM OF

X_i= A GENERAL EXPRESSION THAT REFERS TO ANY VALUES

MEASURES OF CENTRAL TENDENCY

- **Mode:** the most common score
- **Median:** the 'middle' score
- **Mean:** the average of all scores; widely used

MEASURES OF VARIABILITIY

- A measure of variability lets us know to what extent X values (our collected scores) are distributed around a mean
- We can measure variability using the **RANGE** (the distance from two most extreme scores)
- A better measure of variability is the sample variance

$$\text{Sample Variance} = s^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1}$$

- What is the **variance?**
  - **It is an indication of how "spread out" the data is from the mean**

TOPIC 4: The Normal Distribution and Standardised scores

- If we square root the variance we get the standard deviation
- The standard deviation indicated the mean total data that deviates away from the mean
- A more tighter, closely knit graph and a spaced out one may have the same mean but the tighter one would have a smaller variance and standard deviation
- If you square root the computational formula for variance to get the computational formula for standard deviation

THE COEFFICIENT OF VARIATION

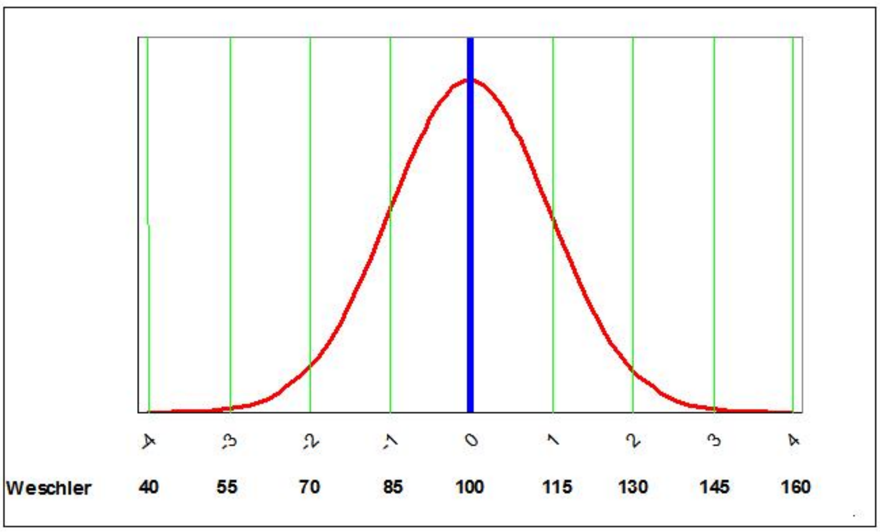
- To compare the "tightness of two different sets of data around their means, we could calculate the coefficient of variation for each of them using the mean and variance (S) value
- CV= S/MEAN
- A lower CV is tighter than a higher CV

NORMAL DISTRIBUTION

- Often, the scores of a population are assumed to be distributed in a **"normal"** (bell shaped) fashion
- We start by measuring the y axis aka **ORDINATE** (frequency of score)
- We then look at the x axis aka **ABSCISSA**
- Putting them together we get a frequency distribution
- From the normal distribution we can calculate the **mean, variance** and **standard deviation**
- **These are population not sample parameters**
- These parameters may change the look of the plot but it is still considered "normally distributed"
- It could look

1. **Mesokurtic**
2. **Leptokurtic**
3. **Platykurtic**

- Regardless of the original shape of the normal distribution we can still **standardize** it
- This means converting our original distribution into a standard normal distribution to do this the mean is converted to 0 (point) and the standard deviation is 1



- Therefore if a score of 15 **is 1 stdev** then a score of 115 is **1 stdev above the mean**
- The standard normal distribution is important because tables expressing the area under this curve are widely available
- The area under the curve is important because assuming you have access to a population's data with regards to a particular variable, these tables allow us to compare an **individual's score** on the variable with the rest of the population
- We can work out where an individual lies in the population
- *The area under the curve corresponds to the proportion or percentage of people higher or lower than a particular score*
- On a distribution curve the 0 point is the MEAN so in effect, 50% of scores are above the mean and 50% are below
- We can express a normality curve as a percentile i.e. the mean is the 50^{th} percentile and 1 stdev above the mean is the 84^{th} percentile
- Summary: if you know how many standard deviations away from the mean a score is, you know the **proportion** of the population *less than,* and *greater than,* this score!

Z SCORES

- How we do this EASILY is by calculating the **Z SCORE**
- **Z score** is a way to express any score's "distance" from the population mean
- **Z=3.9 means they're in the NEXT population**
- A Z score greater than **1.64** has an even *less than* 1/20 (5%) chance of occurring!
- This 5% chance has very important implications for inferential statistics