QUANTITATIVE METHODS CONTINUED

ECMH5004 – Applied Research Skills

AIMS OF THIS SESSION

- The importance of empirical research and experimental methods in informing our judgements on treatment interventions.
- The importance of surveys and questions
- Types of statistics used in experimental methods.
- Types of studies / trials.

EXPERIMENT

- The only research method which directly explores causality
- They are designed to test a hypothesis
- They aim to test a theory (hypothetico-deductive)
- Theory – some kind of general explanation for why something happens
- The aim of the experiment is to gain evidence which will support that theory, or which will challenge it.

DEFINING THE RESEARCH QUESTION

- Problem should be capable of being stated in a question form. Examples are:
  - What is the effect of….?
  - Under what conditions do….?
  - Does the effect of….?
- A **research question** defines the area of interest but is not a declarative statement like a **hypothesis**.

FORMULATING HYPOTHESES

- Stated in declarative form.
- Posits a relationship between variables.
- Ideally reflects a theory or body of literature.
- Is brief and to the point.
- Is testable.

Research Idea	Question	Hypothesis
Drug abuse and school attendance in teenagers	Is drug abuse related to school absenteeism in teenagers?	There is a positive relationship between drug abuse and school absenteeism in teenagers.

Research Idea	Question	Hypothesis
Effects of cyber bullying on anxiety in teenagers	Does cyber bullying affect anxiety levels of teenage victims?	There is a positive relationship between cyber bullying and anxiety levels in teenagers.

HYPOTHESES

- Scientific/alternative hypothesis states the ‘predicted’ relationship amongst the variables.
- Null hypothesis is a statement of no relationship amongst the variables.

HYPOTHESES

Sometimes a hypothesis will predict the direction in which the results are expected to go:

e.g. males are more aggressive than females

e.g. alcohol slows down reaction time

When a hypothesis predicts the direction in which the results will go it is referred to as a **one-tailed** hypothesis.

When a hypothesis does not state a direction but simply states that one factor affects another or that there will be a difference between the scores but does not state in which direction the difference will be, it is referred to as a **two-tailed** hypothesis.

SO…..

- A hypothesis is a testable statement of cause and effect
- A one tailed hypothesis predicts the direction (faster/slower, etc.) the results will go.
- A two-tailed hypothesis makes no prediction about which way the results will go.

EXPERIMENT – HYPOTHESIS

A criminal trial is an example of hypothesis testing without the statistics.

- In a trial a jury must decide between two hypotheses. The null hypothesis is

  H_0: The defendant is innocent

- The alternative hypothesis or research hypothesis is

  H_1: The defendant is guilty

- The jury does not know which hypothesis is true. They must make a decision on the basis of evidence presented.

HYPOTHESIS

Takes the form of '*if-then*' prediction: 'if the theory is true, then under these conditions, this will happen'.

e.g., if it is true that overcrowding leads to aggression, then if I ask children to play in a small and crowded room, they should play more aggressively.

So…when we carry out an experiment, we begin by identifying an experimental hypothesis, and then we set up situations which allow us to test it.

NULL-HYPOTHESIS

- Null hypothesis (H_0) is a statement of no relationship amongst the variables.

- After we have the null hypothesis, we have to determine whether to reject it or fail to reject it.

- Type 1 Errors

-Rejecting H_0 when it is actually true

-Concluding a difference when one does not actually exist

- Type 2 Errors

-Accepting H_0 when it is actually false

-Concluding no difference when one does exist

HYPOTHESES

- Sometimes a hypothesis will predict the direction in which the results are expected to go:

  e.g. males are more aggressive than females

  e.g. alcohol slows down reaction time

- When a hypothesis predicts the direction in which the results will go it is referred to as a **one-tailed (directional)** hypothesis.

- When a hypothesis does not state a direction but simply states that one factor affects another or that there will be a difference between the scores but does not state in which direction the difference will be, it is referred to as a **two-tailed (non-directional)** hypothesis.

OPERATIONALISING

In order to test the hypothesis, we have to carry out a number of operations/activities that are

(a) observable and
(b) performable – this is termed *operationalising.*

If we cannot do this, then it is not a scientific concept.

CLARIFIES TERMS AND THUS PROVIDES OBJECTIVITY

For example, if you were to read a report on the effect of *hunger* on learning in rats – you would want to know what the author of the report defined as 'hunger'.

In the methodological section, you might read that the rat or rats were deprived of food for 24 hours – this then becomes the author's <u>operational</u> definition of *hunger.*

WHAT IS A THEORY?

- A theory is a based upon a hypothesis and backed by evidence.
- A theory presents a concept or idea that is testable. In science, a theory is not merely a guess. A theory is a fact-based framework for describing a phenomenon.
- A psychological theory has two key components:

  (1) it must describe a behaviour

  (2) make predictions about future behaviours.

FALSIFIABILITY….

Not only must a theory be testable, it must also be capable of being proved wrong.

This is known as *falsifiability.*

Karl Popper argued that no matter how many times a theory predicted result occurs – there might still be another theory that is the true one.

Thus we can never **prove** a theory, we can only become 'more confident' of its applicability.

Research Idea	Question	Hypothesis
Drug abuse and child abuse	Is drug abuse related to child abuse?	There is a positive relationship between drug abuse among adults and their physical and psychological abuse as children.

Research Idea	Question	Hypothesis
Effects on your health of caring for a partner who suffers from dementia	Does caring for a partner who has a dementia affect the caregivers' health?	There is a positive relationship between degree of caregiver burden and risk of dying prematurely.

EXPERIMENT TESTING - CONDITIONS

- So…once we have our aim, question and hypothesis we set up a practical way of testing our hypothesis.

- This takes the form of at least 2 **conditions**
  - condition 1 - we do something that we hope will bring about a cause/change /result
  - condition 2 - we do nothing

For example, to demonstrate that chocolate is the factor in a child's diet causing hyperactivity, chocolate might be alternatively given and withheld.

condition 1 children eat chocolate

condition 2 children do not eat chocolate

EXPERIMENT TESTING - VARIABLE

The experiment is generally assumed to be the most reliable and effective method for demonstrating that one *variable* has an effect another *variable* to change

The term, *variable* is used to describe the empirical object or phenomenon that we are observing and it can assume a number of different values

In the chocolate example the ***chocolate*** is one variable and the **hyperactive behaviour** is another variable

VARIABLES

- **Independent variables -** these are variables that are being manipulated by the researcher (for example – we give the child chocolate (1) or withhold it (2) *(Cause)*

- **Dependent variables** – are the changes measured by the researcher – for example, the level of hyperactivity *(Effect)*

- **Extraneous variables** – are other factors which may be affecting the dependent variable (e.g. the child's intake of sugar…….

VARIABLES

In terms of the experiment, other factors, in this example, the intake of sugar could have an important effect on the child's aggressive behaviour – though it was considered that it was the effect of the chocolate which was apparently causing the behaviour.

This is also known as a ***confounding variable*** because it can produce alternative explanations of change in the dependent variable.

CONTROLLING *EXTRANEOUS* VARIABLES

In some cases, differences found in experiments may be the result of *participant variables* – that are pre-existing factors among people (e.g. intelligence).

Thus in experiments, the usual way to control for this is to *randomly* allocate all participants to the various experimental conditions – in this way any such differences would be spread evenly across the conditions of the experiment.

In studies, wherever possible, a ***standardised*** procedure should be used – this is where all participants are tested in the same place, same conditions, and with the same equipment.

WHAT ARE THE IV AND DV?

1. Alcohol makes you fall over
2. Children learn more from interactive CD-ROM teaching tools than from books
3. Frustration creates aggression
4. Men and women use different criteria to select a mate
5. People diagnosed with depression have difficulty recalling specific memories.

ANSWERS

IV (Cause)

1. Alcohol
2. Type of media (CD or book)
3. Amount of frustration
4. Gender (M or F)
5. Depression (depressive or not)

DV (Effect)

1. Falling over
2. Learning
3. Aggressive behaviour
4. Criteria for mate selection
5. Specificity of memory

TYPES OF GROUPS

Experimental groups – can be one or more – these are the people who will, normally, be the recipients of the independent variable (e.g. drug, chocolate, etc)

Control group – using another group to provide a *baseline* against which to compare the performance of the experimental groups. For example, in a medical research programme, this group would not receive the experimental treatment or drug

RANDOMISED CONTROL TRIALS

- Randomized controlled trial: (RCT) A study in which people are allocated at random (by chance alone) to receive one of several clinical interventions.
- One of these interventions is the standard of comparison or control.
- The control may be a standard practice, a placebo ("sugar pill"), or no intervention at all.
- Someone who takes part in a randomized controlled trial (RCT) is called a participant or subject. RCTs seek to measure and compare the outcomes after the participants receive the interventions. Because the outcomes are measured, RCTs are quantitative studies.
- Considered the gold standard in intervention studies.

TYPES OF RANDOMISED CONTROLLED TRIALS

- Single Blind RCT
  - researchers but not the participants know which participants are receiving the active medication or treatment
- Double Blind RCT
  - A double blind trial is a trial where neither the researchers nor the participants know what they are getting. The computer gives each patient a code number. And the code numbers are then allocated to the treatment groups.
- Triple Blind RCT
  - A double-blind study in which, in addition, the identities of those enrolled in the study and control groups and/or the details about the nature of the interventions (experimental medications), are withheld from the statistician(s) who conduct the analysis of the data.

PARTICIPANTS IN DESIGN

M5X02252 – Applied Research Skills

PARTICIPANTS IN DESIGNS

- … how to distribute and use our participant pool
  - All participants doing all conditions?
  - Some participants doing specific conditions?

  - Design question that will depend on:
    - The manipulations we are making
    - Variables we want to investigate
    - Participants we have available

PARTICIPANTS IN DESIGNS

- There are two basic research designs associated with the experimental research strategy
  - between-subjects design/independent-measures design
    - we obtain each of the different groups of scores from a separate group of participants
    - e.g. one group of students is assigned to teaching method A and a separate group to method B
  - within-subjects design/repeated-measures design
    - different groups of scores are all obtained from the same sample of participants
    - e.g. one sample of individuals is given a memory test using a list of one-syllable words, and then the same set of individuals is tested again using a list of two-syllable words

CHARACTERISTICS

Between-subjects design

- it compares separate groups of individuals
- it allows only one score per participant (every score represents a separate, unique participant)

Within-subjects design

- It compares two or more different treatment conditions (or compares treatment and control) by observing or measuring the same group of individuals in all of the treatment conditions being compared
- it looks for differences between treatment conditions within the same group of participants

TYPES OF EXPERIMENTS

TYPES OF EXPERIMENT

- Laboratory experiments
- Field Experiments
- Natural Experiments
- Quasi-experiments

LABORATORY EXPERIMENTS

- Conducted in a special environment where variables can be carefully controlled.
- DVs and IVs can be measured with greater precision than in the field
- Participants are aware that they are taking part in an experiment.
- They may or may not know the true aims of the experiment.

STRENGTHS AND WEAKNESSES OF LAB EXPERIMENTS

☺Well controlled.

☺Permits us to study cause and effect.

☺Confounding variables minimised.

☺Can be replicated (copied/repeated)

Ethical Issues

Deception

Informed consent (is it truly voluntary?)

Psychological harm

☹Artificial or contrived.

☹Participants know they are being studied, which is likely to affect their behaviour.

☹Investigator may influence results.

☹The setting is not like real life – low in mundane realism.

☹Because IV and DV have been operationalised it doesn't represent real-life experiences e.g. trigrams to test memory

☹Low ecological validity

FIELD EXPERIMENTS

- Conducted in a more natural environment, i.e 'in the field'.
- The Independent variable (IV) is still deliberately manipulated by the researcher.
- Participants are often not aware that they are participating in an experiment.
- Not all **field studies** are **experiments**.

EXAMPLES OF FIELD EXPERIMENTS

Bickman's Study of Obedience (1974)

The effects of appearance on behaviour

- Bickman found that New York pedestrians were more likely to obey someone dressed as a guard than someone in a milkman uniform or casually dressed.
- The confederates issued orders to passers-by to **'pick up this bag for me', or 'this fellow is overparked at the meter but doesn't have any change: give him a dime.'** The researcher observed and recorded their responses.

STRENGTHS AND WEAKNESSES OF FIELD EXPERIMENTS

☺Permits us to study cause and effect.

☺Less artificial so have higher ecological validity, (Easier to **generalise** from results.)

☺Avoids some participant effects (when they are unaware of study)

☹Less control of extraneous variables.

☹More time consuming, so more expensive.

Ethical Issues

Informed consent

Difficulty debriefing

Privacy

NATURAL EXPERIMENTS

- The IV has not been deliberately manipulated by the experimenter, so the situation is described as 'natural'.
- This method used when there are IV's that cannot be manipulated directly for ethical or practical reasons.
- Not true experiments – quasi experiments, because IV not changed deliberately to see effect on DV.

EXAMPLES OF NATURAL EXPERIMENTS

- Bahrick et al (1975) – year book study.
- Hodges and Tizard study of institutionalisation.
- St Helena – Charlton et al (2000) – Does the introduction of TV produce an increase in anti-social behaviour?

  - They found that there was no change in anti-social behaviour, in contrast to a study done in Canada in1885 Williams). Reasons possibly because St Helena has a strong sense of identity and no reason to be aggressive (confounding variable?)

STRENGTHS AND WEAKNESSES OF NATURAL EXPERIMENTS

☺ Allows research where IV **cannot be manipulated** for ethical or practical reasons e.g. **deprivation studies**

☺Enables psychologists to study 'real' problems e.g the effects of disaster on health – (increased mundane realism and ecological validity)

☺ Less chance of **demand characteristics or experimenter bias** interfering.

☹Cannot demonstrate **causal relationships** because IV not directly manipulated.

☹Many confounding variables (e.g allocation of participation to groups), threat to internal validity.

☹Can only be used where conditions **vary naturally**.

☹Participants **may be aware** of being studied.

Ethical Issues

Informed consent

Confidentiality

NUMERICAL DATA ANALYSIS

QUANTITATIVE METHODS

- In using quantitative measures, we then need to understand the techniques used to data, for example
  - Descriptive statistics are used to *summarise* – for example, how many people took part, their gender, ages etc.
  - Interpretive statistics – the use of statistical tests to *analyse* data

NUMERICAL DATA

- At times if your data set is quite complex and you have a large sample it can result in anexhaustive amount of data, it can be very hard to discuss and quite easy to get lost in the numbers.
- Also difficult to draw comparisons.
- One method to deal with this is use a single number that summarises or represents the set of numbers.
- You could consider using the “average”.

MEASURES OF CENTRAL TENDANCY.

The Mean

- A calculated "central" value of a set of numbers.

The Mode

- The number that occurs most often within a set of numbers.

The Median

- The middle number of a sequence of numbers.

WORKED EXAMPLE.

- Number List
- 3, 2, 5, 41, 1, 76, 97, 3, 1, 1
- This numbers might not mean anything to you, but these could be the individual scores of high earning individuals on a well-being scale.
- We will now go through a worked example of calculating the mean, mode and median of this data set.

CALCULATING THE MEAN.

- 3, 2, 5, 41, 1, 76, 97, 3, 1, 1

- Add all the numbers together.
  - 3+2+5+41+1+76+97+3+1+1 = 230.

- Divide the total number (230) by the number of responses given (10).

- Mean = 230 ÷ 10 = 23.

CALCULATE THE MEDIAN.

- 3, 2, 5, 41, 1, 76, 97, 3, 1, 1

- Place **all** the numbers in a sequential order from Lowest to Highest.
  - 1, 1, 1, 2, 3, 3, 5, 41, 76, 97

- Look at what number falls in the middle.
  - In the case of an odd number of items, there will be a single number that falls in the middle.
  - In the case of an even number of items there will be no true one figure.
  - 1, 1, 1, 2, 3, 3, 5, 41, 76, 97
  - To work out the median, you add these two numbers together and divide by two.

- 3 + 3 = 6. The Median is 6 ÷ 2 = 3.

CALCULATE THE MODE.

- 3, 2, 5, 41, 1, 76, 97, 3, 1, 1
- Place **all** the numbers in a sequential order from Lowest to Highest.
  - 1, 1, 1, 2, 3, 3, 5, 41, 76, 97
- To calculate the mode you need to note the number that occurs more frequently (the most) in the data set.
- In this case the answer is 1 as this occurs three times within the data set.

CALCULATING THE MODE.

- A frequently asks question is what happens when there is a tie between more than one number in terms of frequency in a given data set.
- For example - 1, 1, 1, 2, 3, 3, 3, 5, 41, 76, 97
- As you can see above the number 1 occurs 3 times, the number 3 also occurs 3 times.
- Answer is you will have multiple modes – in this case the modes are 1,3.
- Do not divide the numbers – just report the multiple modes.

RANGE

THE RANGE OF A DATA SET IS A MEASURE OF THE SPREAD OR THE DISPERSION OF THE OBSERVATIONS.

It is the difference between the largest and the smallest observed value of some quantitative characteristic and is very easy to calculate.

A great deal of information is ignored when computing the range since only the largest and the smallest data values are considered; the remaining data are ignored.

The range value of a data set is greatly influenced by the presence of just one unusually large or small value in the sample (outlier).

Examples

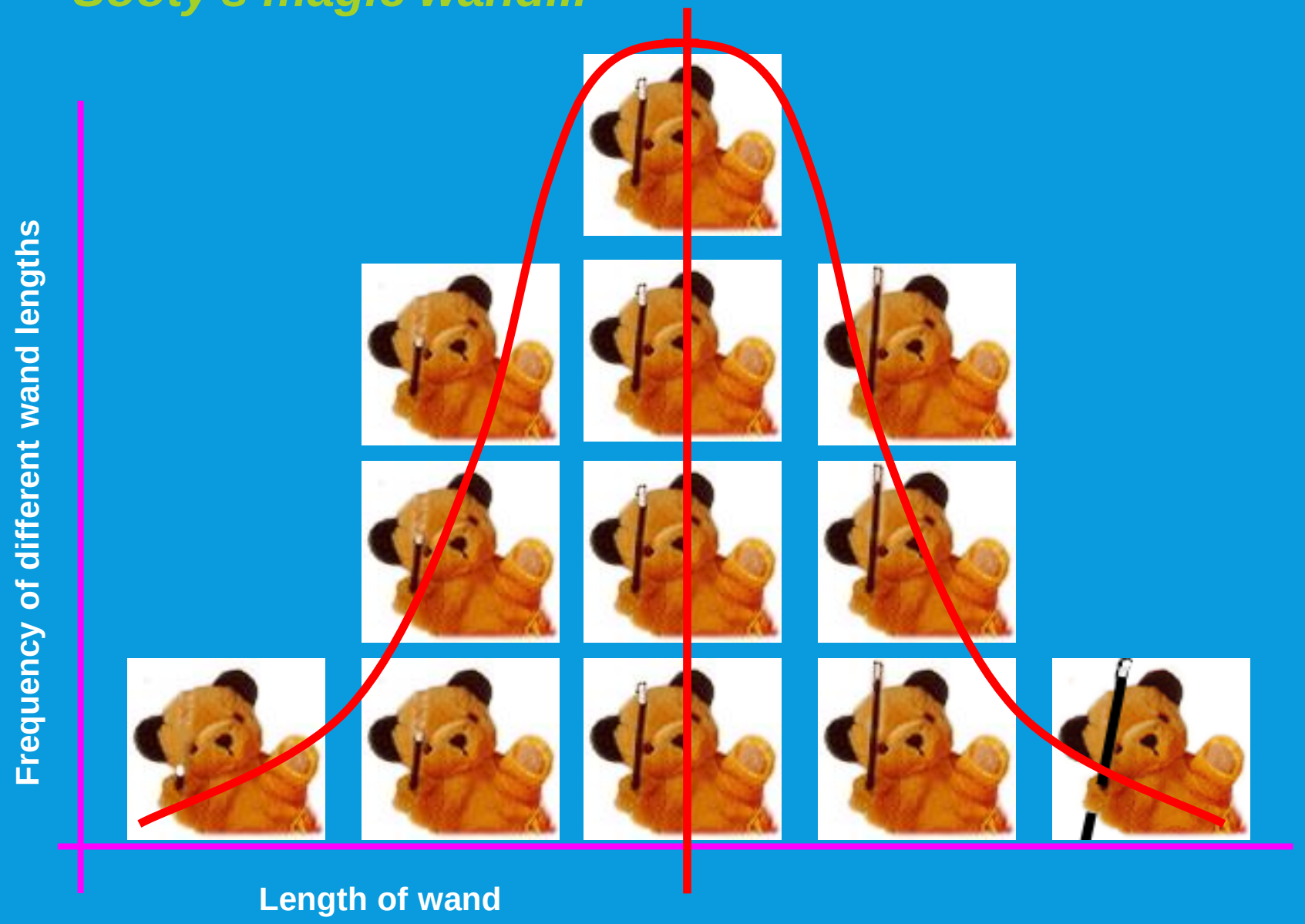
1. The range of 65,73,89,56,73,52,47 is 89-47 = 42.

2. If the highest score in a 1st year statistics exam was 98 and the lowest 48, then the range would be 98-48 = 50.

THE NORMAL DISTRIBUTION

- What is it? Why do we care?
- The important thing is that distributions are tied to probabilities, and it is the probability which will be of interest to us
- If we know something about the distribution of events, then we can estimate the likelihood of our particular event of interest (data)
- We believe that the variables of interest to us are normally distributed in the population
- Assuming a normal distribution allows us to take advantage of its properties and make inferences from our sample to the population

An example of a normal distribution - the length of Sooty's magic wand...



CHARACTERISTICS OF THE NORMAL DISTRIBUTION

- Smooth curve
- Symmetrical (number above or below mean)
- Highest point is the mean
- The more data you collect the more likely you are to approach a normal distribution
- Rare to find a *perfect* normal distribution in a Psychological sample (or most samples)
- But, we should often find *approximate* normal distribution in samples (e.g. intelligence)

STANDARD DEVIATION AND VARIANCE

- Deviation = How far from “normal”.

The Standard Deviation is a measure of how spread out numbers are.

Its symbol is **σ** (the greek letter sigma)

The formula is easy: it is the **square root** of the **Variance.**

So now you ask, "What is the Variance?"

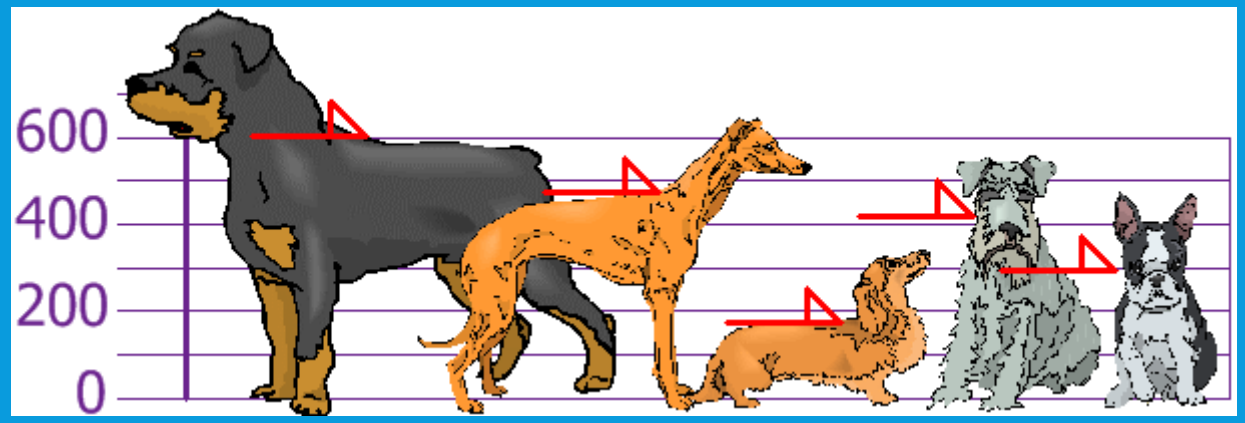
The Variance is defined as:

The average of the squared differences from the Mean.

To calculate the variance follow these steps:

- Work out the Mean (the simple average of the numbers)
- Then for each number: subtract the Mean and square the result (the squared difference).
- Then work out the average of those squared differences.

YOU AND YOUR FRIENDS HAVE JUST MEASURED THE HEIGHTS OF YOUR DOGS (IN MILLIMETERS):

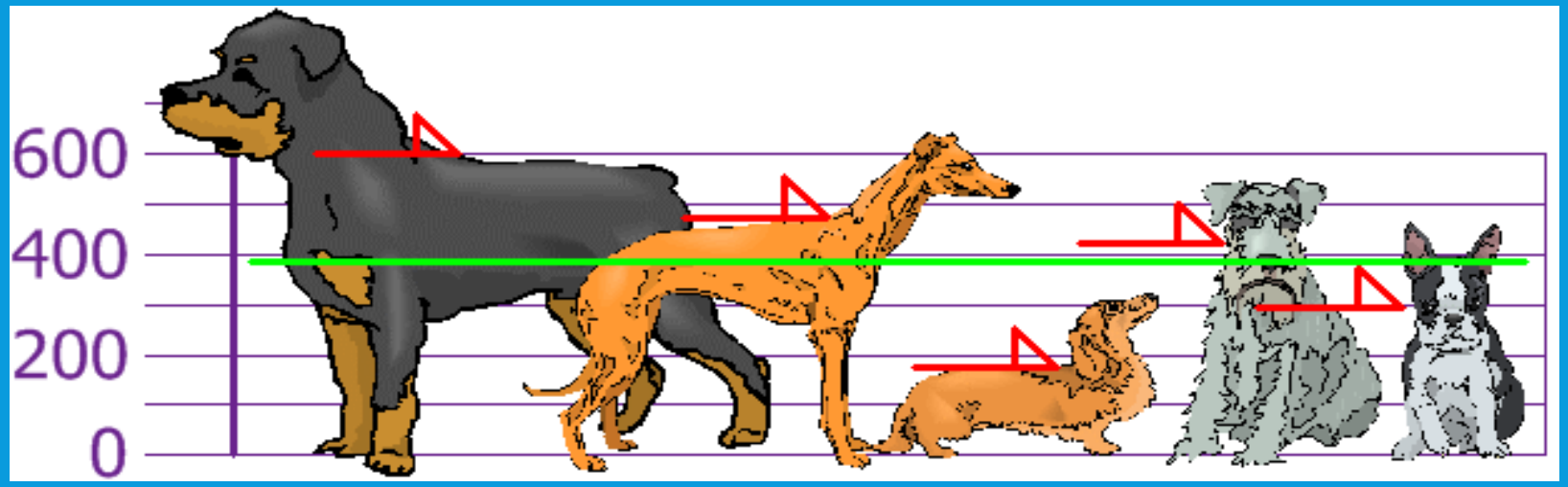


- The heights (at the shoulders) are: 600mm, 470mm, 170mm, 430mm and 300mm.
- Find out the Mean, the Variance, and the Standard Deviation.
- Your first step is to find the Mean:

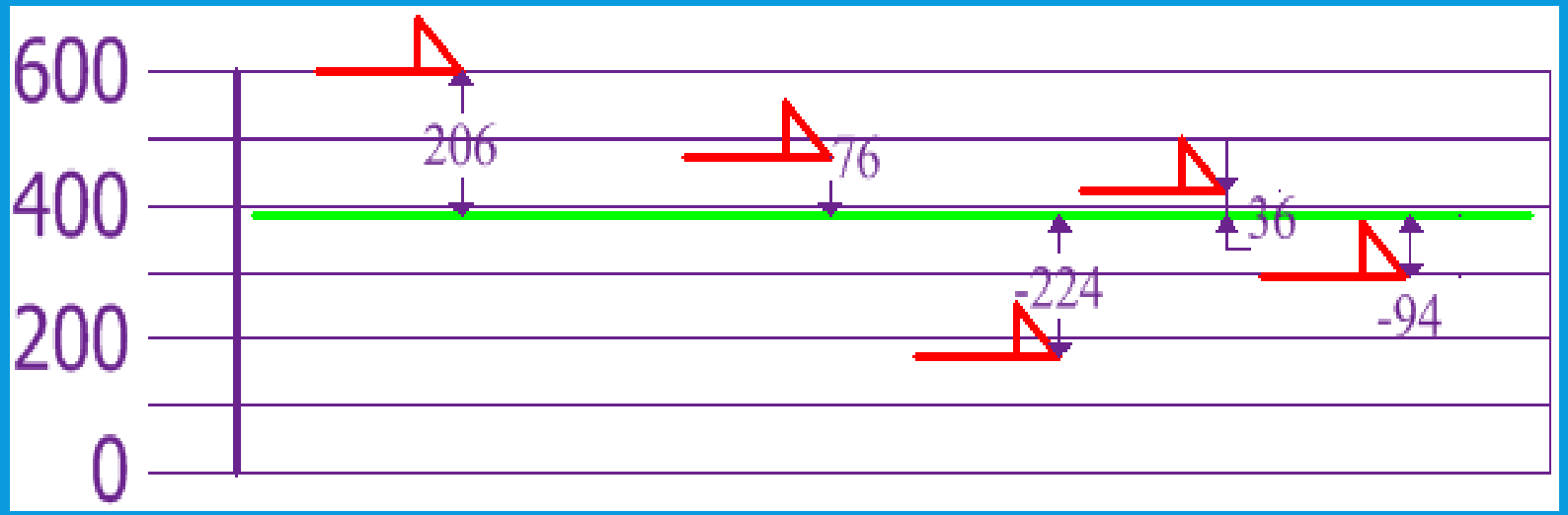
MEAN

$$\text{Mean} = \frac{600 + 470 + 170 + 430 + 300}{5} = \frac{1970}{5} = 394$$

Answer:
so the mean (average) height is 394 mm. Let's plot this on the chart:



- Now, we calculate each dogs difference from the Mean:



VARIANCE

To calculate the Variance, take each difference, square it, and then average the result:

$$\text{Variance: } \sigma^2 = \frac{206^2 + 76^2 + (-224)^2 + 36^2 + (-94)^2}{5}$$

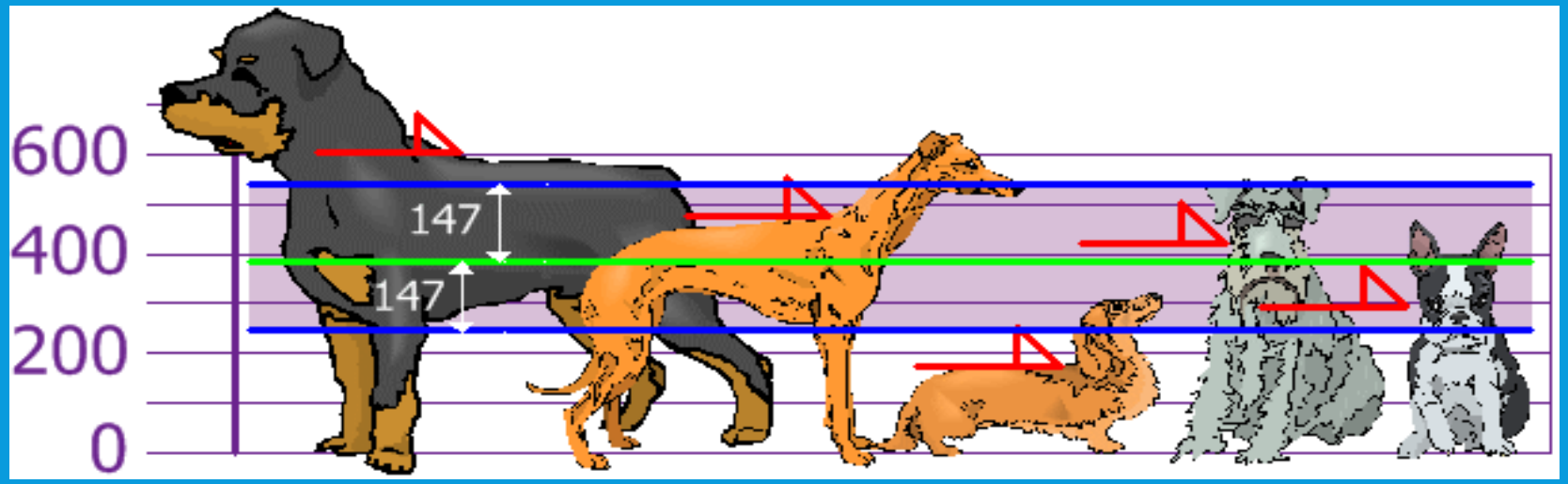
$$= \frac{42,436 + 5,776 + 50,176 + 1,296 + 8,836}{5}$$

$$= \frac{108,520}{5} = 21,704$$

So, the Variance is **21,704**.

STANDARD DEVIATION

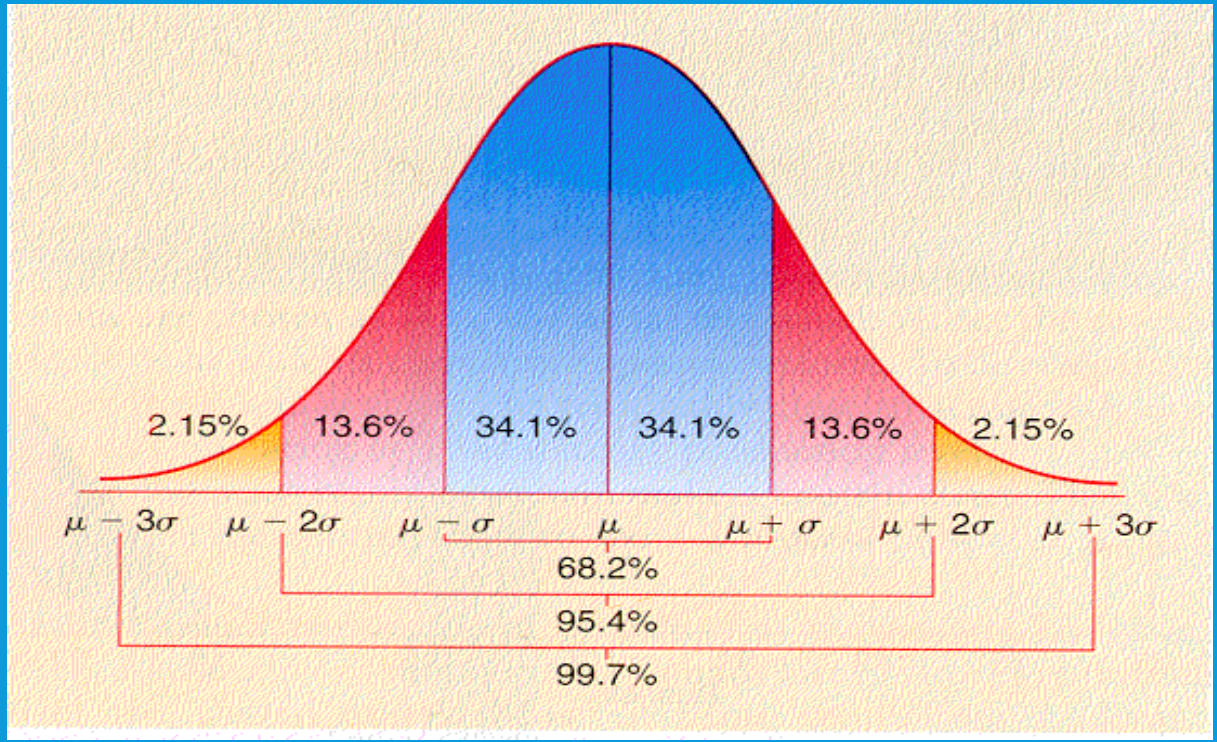
- And the Standard Deviation is just the square root of Variance, so:
- Standard Deviation: **$\sigma = \sqrt{21{,}704} = 147.32\ldots$** = **147** (to the nearest mm)
- And the good thing about the Standard Deviation is that it is useful.
- Now we can show which heights are within one Standard Deviation (147mm) of the Mean:



So, using the Standard Deviation we have a "standard" way of knowing what is normal, and what is extra large or extra small.

So – we have 3 dogs within one SD and 2 dogs within 2 SDs of the mean.

Relationship between the normal curve and the standard deviation:



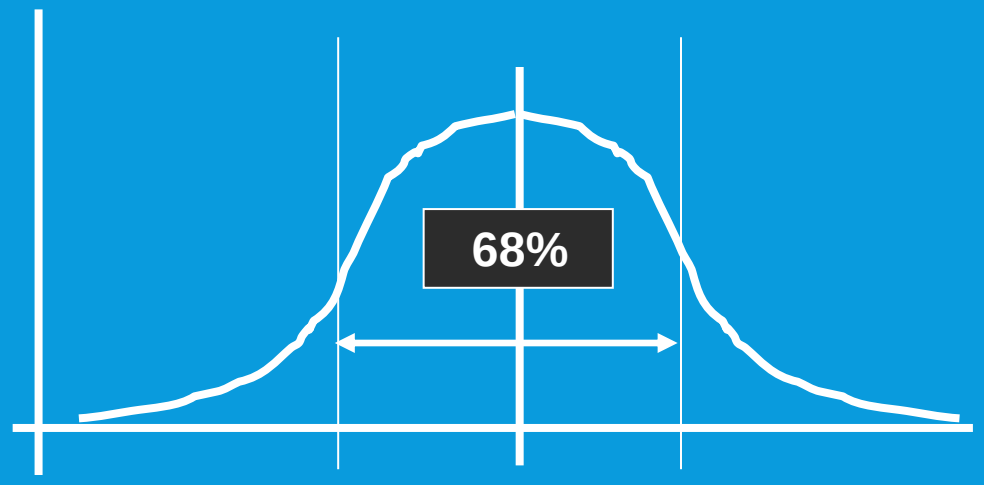
About 68% of values drawn from a **normal distribution** are within one standard deviation σ away from the mean; about 95% of the values lie within two standard deviations; and about 99.7% are within three standard deviations. This fact is known as the 68-95-99.7 (empirical) rule, or the 3-sigma rule.

e.g. IQ is normally distributed (mean = 100, SD = 15).

68% of people have IQs between 85 and 115 (100 +/- 15).

95% have IQs between 70 and 130 (100 +/- (2*15).

99.7% have IQs between 55 and 145 (100 +/- (3*15).



85 (mean - 1 SD) 115
(mean + 1 SD)

We can tell a lot about a population just from knowing the mean, SD, and that scores are normally distributed.

If we encounter someone with a particular score, we can assess how they stand in relation to the rest of their group.

e.g. someone with an IQ of 145 is quite unusual (3 SDs above the mean).

IQs of 3 SDs or above occur in only 0.15% of the population [(100-99.7) / 2].

TYPES OF NUMERICAL DATA

M5X02252 – Applied Research Skills

CATEGORICAL/NOMINAL DATA

Distinct categories, such as an Equal Opportunities form that asks about your 'ethnic' group

1. White (British) 2. White (European), 3. Asian, 4. Afro-Caribbean, 5. African etc.

Important to note that only ONE category should be used, therefore when seeking to categorise, it must be ensured that the cases are distinct and separate.

Number of participants in each category is counted

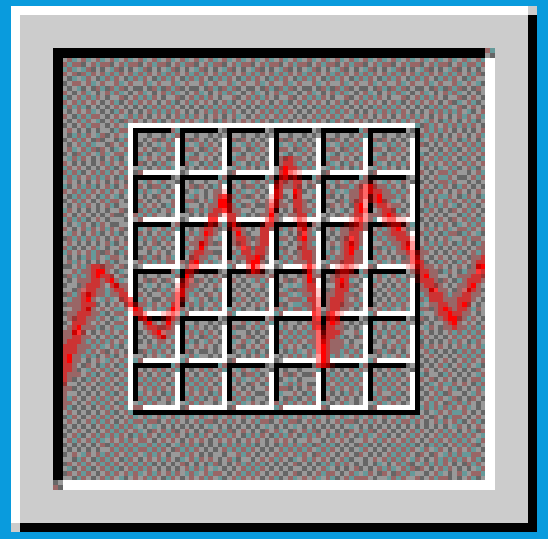
Data in form of frequencies of occurrence

CONTINUOUS DATA IS DATA THAT CAN TAKE ANY VALUE (WITHIN A RANGE)

- Examples:
  - A person's height: could be any value (within the range of human heights), not just certain fixed heights,
  - Time in a race: you could even measure it to fractions of a second,
  - A dog's weight,
  - The length of a gherkin,
  - Lots more!

"CONTINUOUS " DATA

- There are three types of continuous data
  - Ordinal
  - Interval
  - Ratio



ORDINAL SCALES

Data is ranked in some kind of order, ascending or descending

Scores reflect the position in a group - the first, third and fifth person in a race

Pay bands in an organization, as denoted by A, B, C and D

The "Likert scale" is an example of an ordinal scale:

Agree, somewhat agree, disagree, dislike a lot/ dislike a little/ don't know/ like a little/like a lot

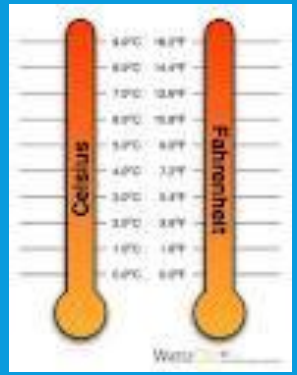
With ordinal data you cannot state with certainty whether the intervals between each value are equal. For example, on a 10 point scale, the difference between a 9 and a 10 is not necessarily the same difference as the difference between a 6 and a 7.

INTERVAL SCALES

Interval scales are those where
the values held by cases on a scale
with a distinct 'size' and this is held throughout the scale.

- Temperature, in degrees Fahrenheit.
- No zero

The difference between 29 and 30 degrees is the same magnitude as the difference between 78 and 79



RATIO SCALE

- In a ratio scale, numbers can be compared as multiples of one another. Thus one person can be twice as tall as another person.

- Important also, the number zero has meaning.

- Thus the difference between a person of 35 and a person of 38 is the same as the difference between people who are 12 and 15. A person can also have an age of zero.

- A person's weight
- The number of pizzas I can eat before throwing up

Provides:	Nominal	Ordinal	Interval	Ratio
The "order" of values is known		✔	✔	✔
"Counts," aka "Frequency of Distribution"	✔	✔	✔	✔
Mode	✔	✔	✔	✔
Median		✔	✔	✔
Mean			✔	✔
Can quantify the difference between each value			✔	✔
Can add or subtract values			✔	✔
Can multiple and divide values				✔
Has "true zero"				✔

STATISTICAL TESTING

ANALYSING THE DATA USING STATISTICS

Why use statistics?

- We are also confronted by enormous amounts of data.
  - How do changes in one variable impact other variables?
  - Is there a way we can measure that relationship?
  - What is the overall strength of that relationship and what does that mean?
- Statistics allow us to answer these kinds of questions

WHY USE STATISTICS?

- **Make Inferences Based Upon Data:**
  - By using inferential statistics, researchers can infer things about a given sample or population.
  - Psychologists use the data they have collected to test the hypothesis, or a guess about what they predict will happen.
  - Using this type of statistical analysis, researchers can determine the likelihood that a hypothesis should be either accepted or rejected.

STATISTICAL TESTING

- Can support our view of a causal relation (reject H_o) between the variables if:
  - There is a difference between conditions?
  - Could this have been purely chance/random?
  - What likelihood we would get identical results for each condition?
  - Therefore we need to find a *significant* difference between conditions

STATISTICAL TESTING

- How can we know whether a difference between conditions is significant?

  - Can we know for certain?
  - No
  - Need to test the *probability* that it is significant

For results to be 'significant' that is – the observed result is not due to chance, researchers tend to use 95% confidence intervals, indicated as $p < 0.05$.

STATISTICAL TESTING

- Probability we need to test:
  - Probability that results were due to the **null hypothesis** (chance/random fluctuation)
  - Probability of results if H_o (null hypothesis) was true
  - This is the 'probability value' or 'p value'

  - Higher p value = more probable null hypothesis true
  - Lower p value = less probable null hypothesis true

STATISTICAL TESTING

- Convention for significance testing:
  - p value lower than 0.05 means significant
  - p value higher than 0.05 means not significant
  - **Remember this!**

- P values can have maximum of 1.0
- 0.05 means there is a 5% chance that the results were due to random fluctuation (null hypothesis)
  - Lower p value = more likely that the difference was due to your manipulation of the Independent Variable (IV)
  - Lower p value = more likely that the difference is significant

QUESTIONNAIRES, SURVEYS AND SCALES

QUESTIONNAIRES AND SCALES.

- Questionnaires and Scales provide a relatively cheap, quick and efficient way of obtaining large amounts of information from a large sample of people.
- Data can be collected relatively quickly because the researcher would not need to be present when the questionnaires were completed.
- This is useful for large populations when interviews would be impractical.

ATTITUDE

- What is an attitude?
- "a psychological tendency that is expressed by evaluating a particular entity with some degree of favour or disfavour"
- (Eagly & Chaiken, 1993)

COMPONENTS OF ATTITUDE

"a psychological tendency that is expressed by evaluating a particular entity with some degree of favour or disfavour"

(Eagly & Chaiken, 1993)

Affective Component – Reflective of a person's general feelings or emotions towards an object or subject (like, dislike, love, hate)

Cognitive Component – Reflective of a person's awareness of and knowledge about an object or subject (know, believe)

Behavioural Component – Reflective of a person's intentions and behavioural expectations, and predisposition to action

ATTITUDE SCALES

- How do you feel about…
- What do you think about…
- How do you behave towards…

  - Abortion
  - The death penalty
  - Gun/knife crime
  - Social Anxiety Disorder (and many many more).

SCALES

CHARACTERISTICS DIFFERENT TYPES OF RATING SCALES

Rating Scale	Subject must:	Advantages	Disadvantages
1. Simple attitude scaling			
2.Category scale	Indicate a response category	Flexible, easy to respond	Ambiguous items, few categories, only gross distinction.
3. Likert scale	Evaluate statements on a 5- or 7-point scale	Easiest scale to construct	Hard to judge what a single score means
4. Semantic differential and numerical scales	Choose points between bipolar adjectives on relative dimensions	Easy to construct, norms exist for comparison, e.g. profile analysis	Bipolar adjectives must be found, data may be ordinal, not interval
5. Constant sum scale	Divide a construct sum among response alternatives	Scale approximates an interval measure	Difficult for respondents with low education levels
6. Stapel scale	Choose point on scale with 1 center adjective	Easier to construct than semantic differential	Endpoints are numerical, not verbal.
7. Graphic scale	Choose a point on a continuum	Visual impact, unlimited scale points	No standard answers
8. Graphic scale-picture response	Choose a visual picture	Visual impact, easy for poor readers	Hard to attach a verbal explanation to response

ASSUMPTIONS OF ATTITUDE SCALES

It is possible to express attitudes using verbal statements

A verbal statement will have the same meaning for all participants

Attitudes expressed as verbal statements can be measured or quantified

LIKERT SCALES

This scale consists of a series of statements where the respondent provides answers in the form of agreement or disagreement.

The respondent selects a numerical score for each statement to indicate the degree of agreement or otherwise.

Each such score is finally added up to measure the respondents attitude.

Eg. of Likert Scale for evaluating the attitude of customers who have not used Vacuum cleaner, but who have aware of its existence:

	Strongly Agree	Agree	Neutral	Disagree	Strongly disagree
The Product is costlier	1	2	3	4	5
I don’t find time to use it	1	2	3	4	5
Advt. is not convincing	1	2	3	4	5
Never use a VC	1	2	3	4	5
I am satisfied with the present way of cleaning	1	2	3	4	5
Its use is cumbersome	1	2	3	4	5

RELIABILITY & VALIDITY

- Reliability

  "…the degree to which a test or measure produces the same scores when applied in the same circumstances…" (Nelson 1997)

Validity

"Degree to which a test or instrument measures what it purports to measure" (Thomas & Nelson 1996)

Objectivity

"…the degree to which different observers agree on measurements…" (Atkinson & Nevill 1998)

TYPES OF VALIDITY

FACE VALIDITY

- Infers that a test is valid by definition
- It is clear that the test measures what it is supposed to

e.g.
If you want to assess reaction time, measuring how long it takes an individual to react to a given stimulus would have face validity

But what about giving someone a depression scale as a measure of well-being?

CONTENT VALIDITY

- Infers that the test measures all aspects contributing to the variable of interest

e.g.
Who is the most physically fit?

- VO_2 max test?
- Wingate test (cycle)?
- 1 RM (repitition – weights)?

CONCURRENT VALIDITY.

- Concurrent Validity – How does the test measure when compared to other tests (e.g. those that have been validated).
- Neisser et al (1996) – Strong Correlation between a variety of Intelligent tests.
- Becks Depression Inventory and 17 item Depression scale of the Kiddie-Schedule for Affective Disorders and Schizophrenia (K-SADS) (Ambrosini et al 1991).

PREDICTIVE VALIDITY.

- Predictive Validity = The relationship between the test and later scores (e.g. Knowledge, Skills or Ability).

- Studies have highlighted that some Intelligence test have a high predictive ability when it comes to success in jobs, academic qualifications etc (Maltby et al, 2010).

CONSTRUCT VALIDITY

- Infers not only that the test is measuring what it is supposed to, but also that it is capable of detecting what *should* exist, theoretically
- Therefore relates to hypothetical or intangible constructs

e.g.
Team Rivalry

Sportsmanship.

This makes assessment difficult, i.e. if what *should* exist cannot be detected, this could mean:

a) Test Invalid? b) Theory Incorrect? c) Sensitivity/Specificity Issues?