Understanding the basics of quantitative data analysis

Define tomorrow.

UNISA college of education

START HERE

What is quantitative research?

A definition

Paradigm

Positivist

Research design

Experimental, non-experimental, quasi-experimental

Participants

Population, type of sample, size and description

Data collection

Questionnaires, interviews (structured), meta-analysis
Valid and reliable (psychometric properties)

Data analysis

Statistical techniques / procedures

Quantitative research emphasizes **objective measurements** and the **statistical, mathematical, or numerical analysis of data** collected through **polls, questionnaires, and surveys**.

Quantitative research focuses on **gathering numerical data and generalizing** it across groups of people or to explain a particular phenomenon.

Your goal =

(More often than not) to determine the relationship between one thing [an independent variable] and another [a dependent or outcome variable] within a population.

The main characteristics are…

- The data is usually gathered using **structured research instruments**.
- The results are based on **larger sample sizes** that are representative of the population.
- The research study can usually be **replicated or repeated**, given its high reliability.
- Data are in the form of **numbers and statistics**, often arranged in tables, charts, figures, or other non-textual forms.
- Results can be used to **generalize concepts more widely, predict future results, or investigate causal relationships.**
- Researcher uses tools, such as **questionnaires or computer software**, to collect numerical data.

What is quantitative research?

A definition

Paradigm

Positivist

Research design

Experimental, non-experimental, quasi-experimental

Participants

Population, type of sample, size and description

Data collection

Questionnaires, interviews (structured), meta-analysis
Valid and reliable (psychometric properties)

Data analysis

Statistical techniques / procedures

Next…

Ontology: How do you view knowledge? Is knowledge constructed between individuals, or is knowledge law-like and stable regardless of who views it? Do you believe knowledge exists outside of the knower, or do you believe the knower constructs knowledge?

Epistemology: How will the relationship between the you (as the researcher) and the phenomena (you wish to study) be characterised?

Methodology: How you will go about investigating that which is to be known?

PARADIGM/DI-MENSION	ONTOLOGY	EPISTEMOLOGY	METHODOLOGY
Positivist (Quantitative)	• **Stable external reality** • **Law like**	• **Objective** • **Detached observer**	• **Experimental** • **Quantitative** • **Hypothesis testing**
Interpretive	• Internal reality • Subjective experience	• Empathetic • Observer • Intersubjective	• Interactional • Interpretive • Qualitative
Constructionist	• Socially constructed reality • Discourse	• Suspicious • Political • Observer constructed versions	• Deconstruction • Textual analysis • Discourse analysis

Mason, n.d.

In a nutshell…

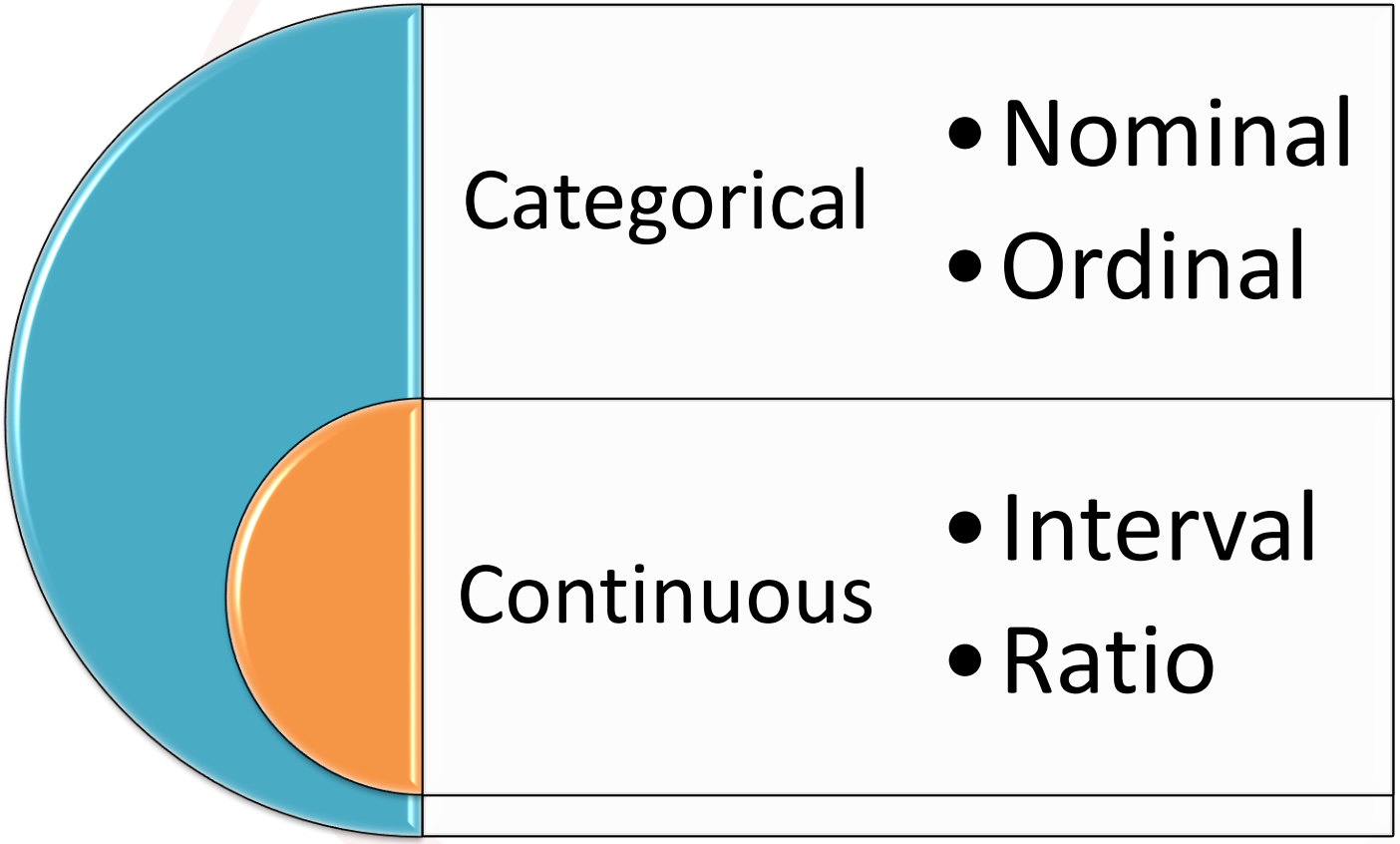
…The quantitative researcher **regards knowledge as existing objectively** (ontology) and therefore adopts the role of an **impartial and detached observer** (epistemology) who utilises **valid and reliable instruments, such as questionnaires** (methodology) to collect data (Clark Carter, 2010)…

…The researcher **collects data objectively**, assumes a **detached position** and adopts methods which allow for **generalisations** to the broader population…

Before getting into the designs, let's have a look at some common terms…

- What is a variable?
- Categorical and continuous variables (Levels of measurement)
- Independent variables and dependent variables (Predictor and outcome variables)
- Confounding, mediating and moderating variables

A ***variable*** is an entity that ***varies*** and can take on a ***variety of values***.



FRUIT

AGE CATEGORIES

TEMPERATURE

PRESSURE

Identifying variables

Independent variable (IV)

Variable(s) that affect or cause the outcome to occur

The variable that is being manipulated

Another name in a Regression study = predictor variable

Another name in a True experimental design = cause

Dependant variable (DV)

Variable that is influenced by, or acted upon, by the independent variable

The variable that is not being manipulated

Another name in a Regression study = outcome variable

Another name in a True experimental design = effect

Confounding variables

Correlated with the IV

Causally related to the DV

Check out this useful video which explains in detail

https://www.youtube.com/watch?v=WZr1jIKi_s0

Moderating variables

Influence the direction and the strength of the relationship between the IV and the DV

Mediating variables

Explains the relationship between the DV and the IV

Accounts for the relationship

Paradigm ✓

Positivist

Research design

Experimental, non-experimental, quasi-experimental

Participants

Population, type of sample, size and description

Data collection

Questionnaires, interviews

Valid and reliable (psychometric properties)

Data analysis

Statistical techniques / procedures

Two prominent {quantitative} designs...

Experimental

Non-experimental

Experimental

- Determine causality (i.e., **cause and effect**)
- **Manipulation** of the IV
- **Random assignment** to control and experimental groups
- Often conducted in laboratory settings

Non-experimental

- Does not determine causality, rather **measures relationships** between two or more variables
- **No intervention or manipulation** of the IV

Quasi-Experimental

- Shares **similarities** with the true experimental design, but it specifically **lacks the element of random assignment** to treatment or control
- **Less rigorous** when compared to true experimental designs
- Utilised **outside of laboratory** contexts

Paradigm ✓

Positivist

Research design ✓

Experimental, non-experimental, quasi-experimental

Participants

Population, type of sample, size and description

Data collection

Questionnaires, interviews

Valid and reliable (psychometric properties)

Data analysis

Statistical techniques / procedures

A **population** is a group of individuals with some commonality (i.e., they all have something in common)

It would be impossible to collect information from all female smokers in South Africa. Therefore, the researcher would select individuals from which to collect the data. This is called **sampling.**

There are two main types of sampling: **probability and non-probability sampling**.

What's the difference? **Randomisation.**

Randomisation occurs when all members of the sampling frame have an *equal opportunity (or probability)* of being selected to participate in the study. This is a feature characteristic of **probability sampling**.

Probability sampling

SIMPLE RANDOM SAMPLING

1. Get a list or 'sampling frame'
2. Generate random numbers
3. Select one person per random number

SYSTEMATIC RANDOM SAMPLING

1. Select a random number, which will be known as *k*
2. Get a list of people
3. Select every *k*th person

STRATIFIED RANDOM SAMPLING

1. Separate your population into groups (strata)
2. Do either a simple random sample or a system random sample from there

Non-probability Sampling

CONVENIENCE SAMPLING

1. Find some people that are easy to find

SNOWBALL SAMPLING

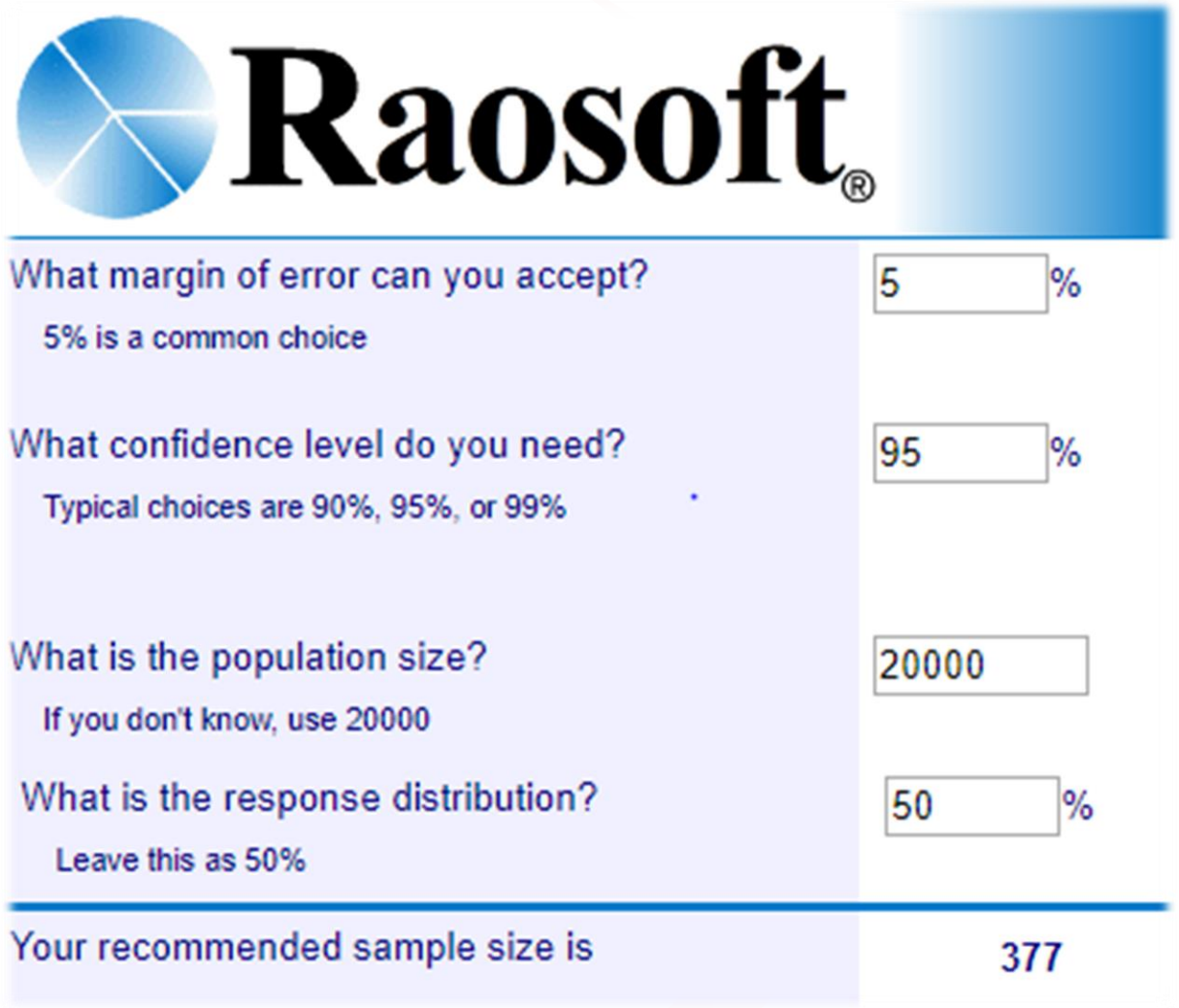
1. Find a few people that are relevant to your topic
2. Ask them to refer you to more of them

QUOTA SAMPLING

1. Determine what the population looks like in terms of specific qualities
2. Create quotas based on those qualities
3. Select people from each quota

Now that you have a better idea of sampling techniques, you can use this handy website to determine the sample size needed for your study.

http://www.raosoft.com/samplesize.html



Paradigm ✓

Positivist

Research design ✓

Experimental, non-experimental, quasi-experimental

Participants ✓

Population, type of sample, size and description

Data collection

Questionnaires, interviews

Valid and reliable (psychometric properties)

Data analysis

Statistical techniques / procedures

Data collection methods

- Quantitative studies often rely on **random sampling** and **structured data collection instruments**
- These produce results that are easy to **summarise, compare and generalize**

Typical quantitative data gathering strategies include…

- Experiments / clinical trials;
- Observing and recording well-defined events;
- Obtaining relevant data from management information systems;
- But, the most common data gathering technique in the social sciences is the **administration of surveys with closed-ended questions** (e.g., online surveys, paper-based surveys, face-to face and telephone interviews with structured data collection instruments).

When administering surveys…

Ensure the instrument is **valid and reliable** (i.e., establish the **psychometric properties** of the instrument among your sample / determine how valid and reliable the instrument is for your sample)…

- Conduct **factor analyses** to determine **construct validity** (Is the instrument valid? I.e., Is the instrument measuring what it intends to measure?)

- Conduct **item analyses** to determine the scales' **internal consistency** (Is the instrument and it's subscales reliable?)

https://www.youtube.com/watch?v=UWP9OEoaNnE&list=PL7Tw2kQ2edvpCOKL40H_kHolyaiHRsW4Q

Paradigm ✓

Positivist

Research design ✓

Experimental, non-experimental, quasi-experimental

Participants ✓

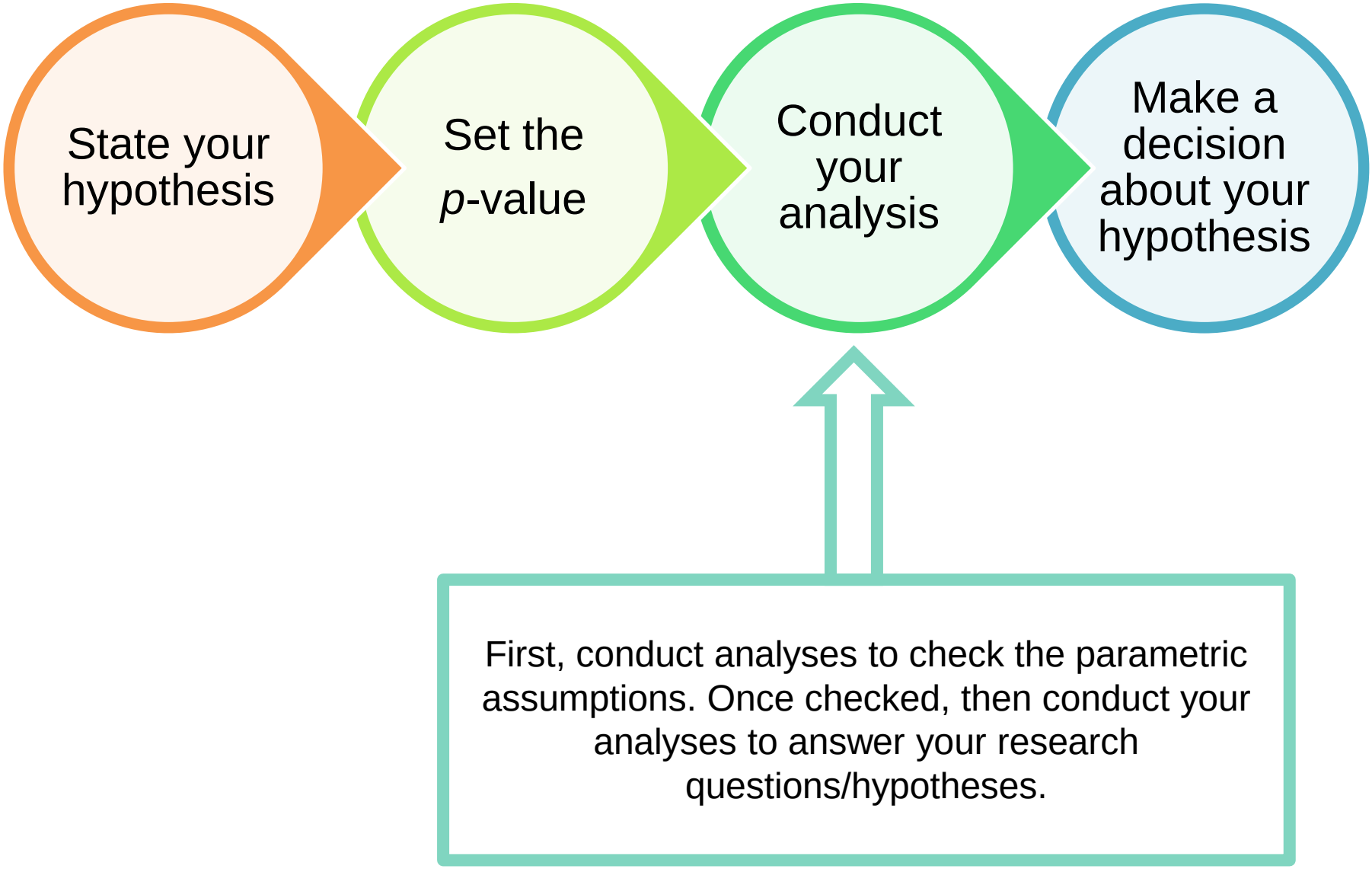
Population, type of sample, size and description

Data collection ✓

Questionnaires, interviews
Valid and reliable (psychometric properties)

Data analysis

Statistical techniques / procedures


State your hypothesis
Set the *p*-value
Conduct your analysis
Make a decision about your hypothesis
First, conduct analyses to check the parametric assumptions. Once checked, then conduct your analyses to answer your research questions/hypotheses.

What's a hypothesis?

A **hypothesis** is a tentative, testable statement

What's the difference between the null and the alternative hypothesis?

Null hypothesis:

- States that no difference/relationship exists between the variables

Alternative hypothesis:

- States that a difference/relationship exists between the variables
- Can be stated in two ways

Stating the hypotheses…

H_0 = Null hypothesis

H_0: There is no difference between IQ1 and IQ2
H_o: ų1 = ų2 or H_0: ų1-ų2 = 0

H_1 = Alternative hypothesis

Non-directional H_1 :
There is a significant difference between IQ1 and IQ2
H_1: ų1 ≠ ų2, or H_1 : p ≠ 0

Directional H_1:
There is significantly higher mean reported in IQ2 when compared to IQ1
H_1: ų1 </> ų2

Choose a (statistical) significance level…

- Called an alpha and refers to probability level (*p*-value)
- **Typically set at 0.05**, 0.01 or 0.001
- Interpretation: 0.05 = 95% chance that the relationship/effect did not occur by accident, and so forth
- But, there is a 5% chance that the relationship/effect was random

Check the assumptions associated with your planned statistical technique…

- First and foremost – Normally distributed data (https://www.youtube.com/watch?v=2GRZ_d4ftoo)
- Equal variances (https://www.youtube.com/watch?v=O6taUlWejB0)
- Independence
- Data are measured at the interval or ratio scale

Assumptions met

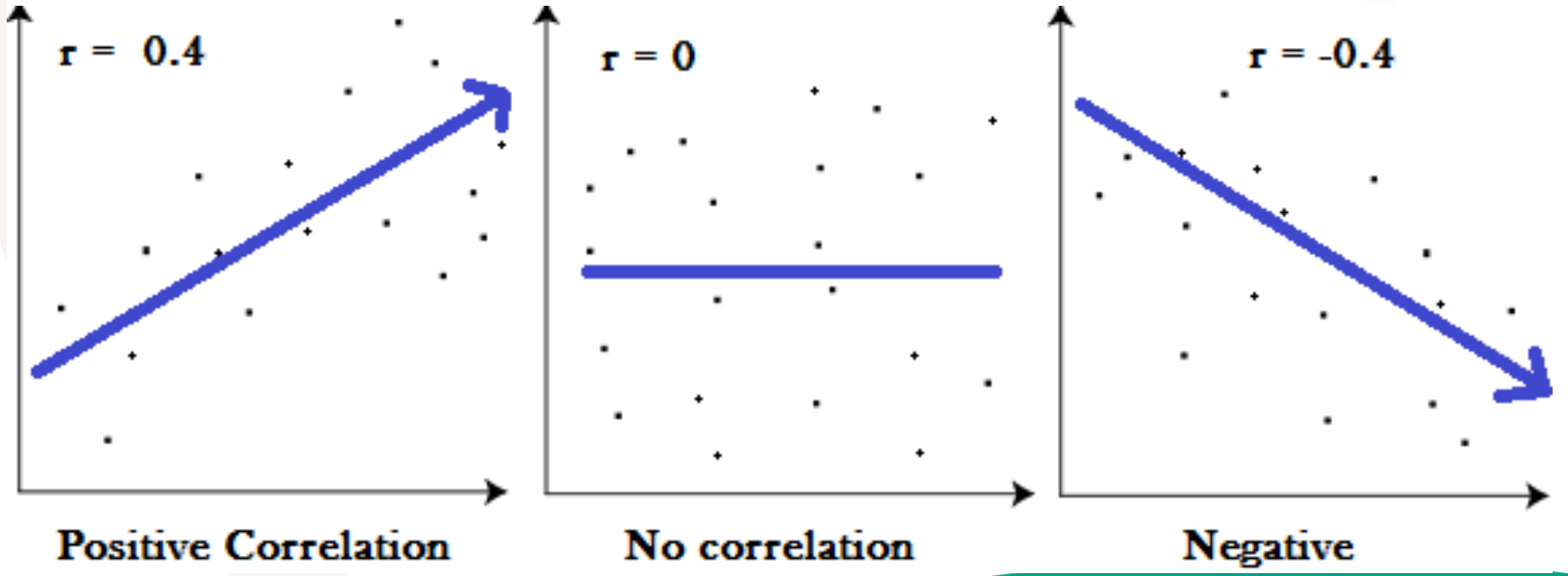
- Pearson's correlation
- Independent t-test (2 groups)
- Independent t-test (>2 groups)
- Dependent t-test (2 conditions)
- Dependent t-test (>2 conditions)

Assumptions not met

- Spearman's rho or Kendall tau's correlation
- Mann-Whitney test
- Kruskal Wallis test
- Wilcoxon test
- Friedman's ANOVA

Correlation

- Statistical techniques used to describe the relationship/association between two variables – bivariate correlation
  - Positive, negative or no correlation (ranging from -1 to +1)
    - Zero indicates no relationship



Regression

Taking correlations one step further … Predicting one variable from another

Simple regression: ONE predictor variable; CONTINUOUS outcome variable

Multiple regression: SEVERAL predictor variables; CONTINUOUS outcome variable

Binary logistic regression: ONE OR MORE predictor variables; CATEGORICAL outcome variable (Only two outcomes)

Multinomial logistic regression: ONE OR MORE predictor variables; CATEGORICAL outcome variable (More than two outcomes)

Independent measures (independent *t*-test, Mann-Whitney)

Used to compare the mean scores of **two different (completely unrelated) groups**.

Is the mean score on a happiness scale lower for people who have children versus people who don't have children?

IV: People have children vs No children (**two groups**)

DV: Scores on happiness scale (**one DV**)

Repeated measures (dependent *t*-test, Wilcoxon)

Used to compare the mean scores of the **same group**

Is the mean score on a happiness scale higher following the completion of an online course on happiness?

IV: Online course on happiness (**pre-and post-test groups**)

DV: Scores on happiness scale (**one DV**)

One-way (independent-measures) ANOVA or Kruskal-Wallis…

Similar to *t*-test, but used with **more than two (independent) groups**.

One-way (repeated-measures) ANOVA or Friedman's ANOVA…

Similar to *t*-test, but used with **more than two (dependent) groups**.

Make a decision about the hypotheses

So, we've conducted statistical analyses on our data. Now what?

Well, each statistical procedure provides a test statistic and its associated *p*-value. You need to use these statistics to make a decision about your hypotheses.

The *p*-value indicates the statistical significance of the result – did it occur by chance or not?

If you set your *p*-value at 0.05 and your results reveal a *p*-value of:

$p \leq .05$ – statistically significant and reject the H_o

$p > 0.05$ – not statistically significant and accept the H_o

Researchers could make errors at this stage…

Type I error "Rejecting the H_o if it was in fact true"

Type II error "Accepting the H_o if it was false"

Type I error

- Rejecting the H_o if it was in fact true
- Thus, there was no difference between two groups, but the results and *p*-value erroneously suggested that there was a difference

Type II error

- Accepting (fail to reject) the H_o if it was false
- There were in fact differences between the groups, but the results and *p*-value indicated that there were no differences.
- There is always a chance that the researcher could reach incorrect conclusions

Some key tips to keep in mind when reporting your results:

1. **Explain the data collected as well as the results** in relation to the research problem you are investigating. **Interpretation of results is not appropriate in this section** < Leave that to the discussion chapter/section
2. **Report unanticipated events** that occurred during your data collection. Explain how the actual analysis differs from the planned analysis. Explain your handling of missing data and why any missing data does not undermine the validity of your analysis.
3. **Explain the techniques** you used to clean your dataset.

4. **Describe the assumptions** for each statistical test and the steps you took to ensure that they were not violated.
5. **Choose an appropriate statistical procedure**; provide a rationale for its use and a reference for it. Also, specify any computer programs used, such as SPSS or MS Excel.
6. **When using inferential statistics**, provide the descriptive statistics, confidence intervals, and sample sizes for each variable as well as the value of the test statistic, its direction, the degrees of freedom, and the significance level [the *p* value].
7. **Avoid inferring causality** (unless, of course, you adopted an experimental design, which proves causality)

write

Thank you

Dr Kelly Anne Young: youngka@unisa.ac.za

Bibliography

- Clark-Carter, D. (2010). Quantitative psychological research: The complete student's companion (3rd Ed.). New York, NY: Psychology Press.
- Field, A. (2013). Factor analysis and PCA. Retrieved from https://www.youtube.com/watch?v=UWP9OEoaNnE&list=PL7Tw2kQ2edvpCOKL40H_kHolyaiHRsW4Q
- Gaskin, J. (2018). Mediator or Moderator? Retrieved from https://www.youtube.com/watch?v=WZr1jlKi_s0
- Grande, T. (2015). Normality Tests in SPSS. Retrieved from https://www.youtube.com/watch?v=2GRZ_d4ftoo
- Löfgren, K. (2013). Test for equality of variances: Parametric and nonparametric Levene's test in SPSS. Retrieved from https://www.youtube.com/watch?v=O6taUlWejB0
- Mason, H. (nd). Research methodology. Pretoria, The South African College of Applied Psychology (SACAP)
- Parametric and non-parametric tests (Image on Slide 31). Retrieved from http://changingminds.org/explanations/research/analysis/parametric_non-parametric.htm
- Raosoft sample size calculator. Retrieved from http://www.raosoft.com/samplesize.html