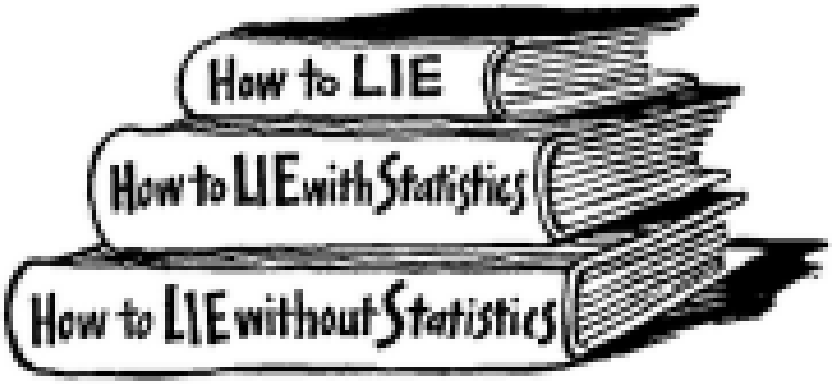
Data Analysis

Data Analysis



[Source: www.academicproofreader.com]

"Facts are stubborn things, but statistics are more pliable"

Often attributed to Mark Twain or Benjamin D'Israeli

Key ideas for this class

- What types of data are there?
- What types of variables are there?
- What is a scale?
- How do I analyze data?
- How do I assess quality of data?

Types of Data

Generally, two types of data can be distinguished:

- Qualitative Data – Data collected by means of observation of the environment in which you are interested such as text, photos, people
- Quantitative Data – Numerical results of measured/simulated variables from wind tunnel data, FEM results, surveys

The type of data you generate follows from your research strategy, research material and your defined variables.

Beware: all qualitative data is based on quantitative judgements and all qualitative data can be summarized and manipulated numerically!

Variables

For every research there are two type of variables:

- **Dependent Variables:** the outcome you measure (effect)
- **Independent Variables:** the variables you manipulate (cause)

Your dependent variables can often already be identified from your research questions

The interaction between these two types of variables will determine the answer to your research questions.

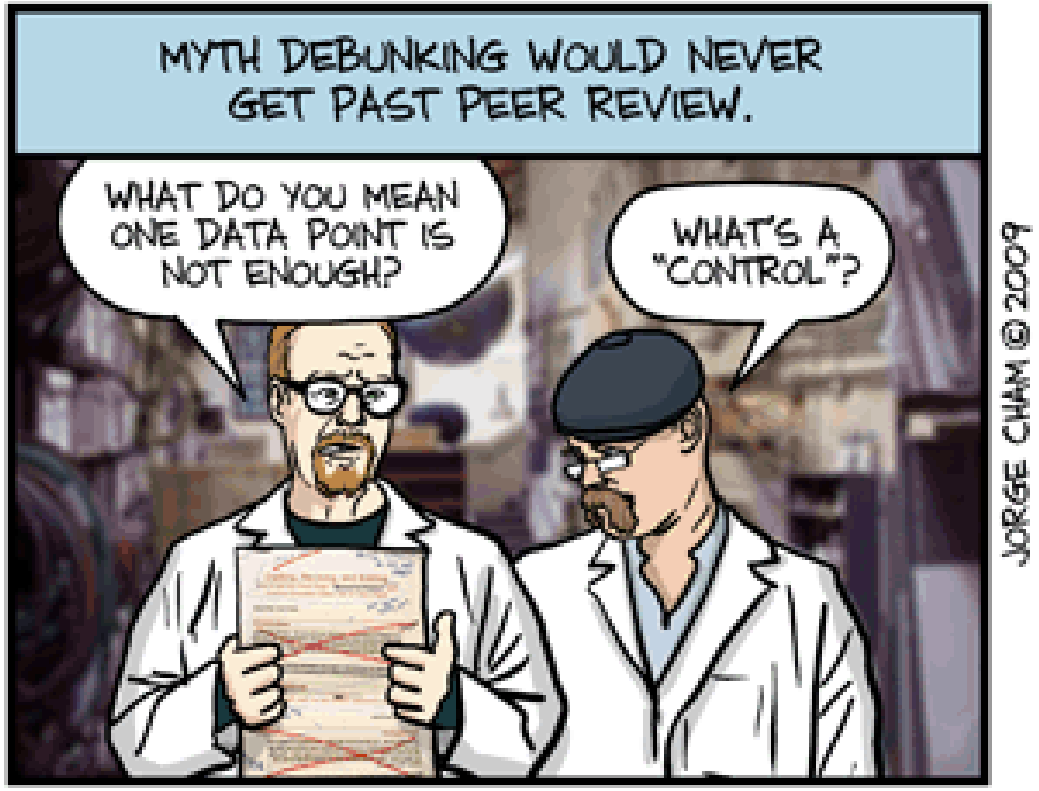
Research Material

- What is the population of your study?
  - Population is defined by Field & Hole (2006) as “entire collection of things”
- What subset of my population am I going to study & why?
  - Your chosen Subset is called a sample
- What data sources am I going to use?
  - Generating your own through testing or surveying or observing
  - Someone else’s data
  - Simulated data
  - Theoretical data

Sampling

- We can rarely study an entire population, hence we must take a representative sample.
- A **Sample** is a representative cross-section of your population that allows you to make generic statements over the entire population
- Choosing the right sample can seriously improve the quality and speed of your research
- Always explain how and why you choose your sample. Keep in mind your results must be replicable!
- Required sample sizes can be calculated and depend greatly on whether your workable population is finite or not using statistics

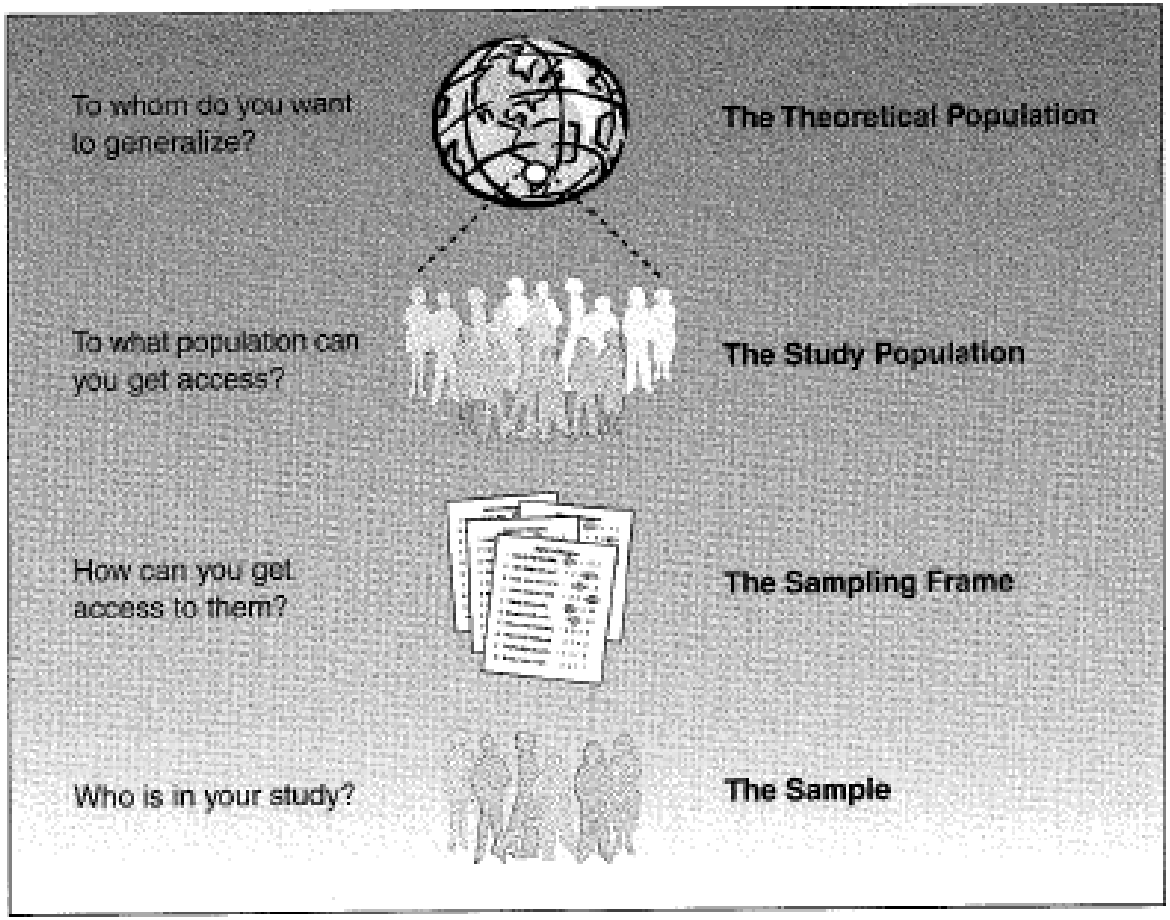
Sample Size Issues



[source: Piled Higher and Deeper" by Jorge Cham www.phdcomics.com]

Sampling Method

- Make it truly generic
- Group you want to generalise
- List how you draw population
- The actual units used in your study



Source: The research methods knowledge base. W.M.K. Trochim & J. P. Donnelly. ISBN-13: 978-1-59260-290-2. Cengage Learning, USA

Sampling Error

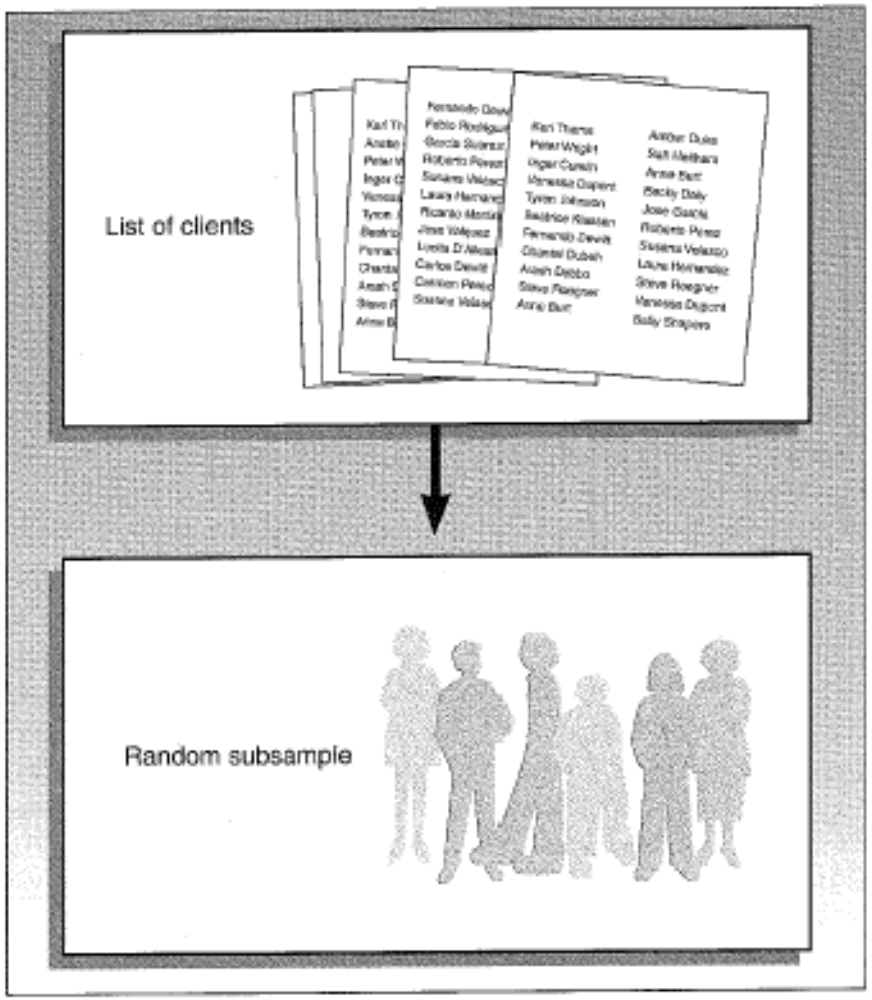
- In sampling the standard error is the sampling error – the precision of your statistical estimate. Calculated from the standard deviation:
  - greater the SD the greater the error (more variability in sampling distribution!)
  - the greater the sample size the smaller the error (nearing full population and parameter estimation!)
- Standard deviation: the spread of variability of the scores around their average in a single sample (square root of the variance). SD in same units as original measure and easy to interpret

Types of Samples

- Probability Sampling
  - Simple random sampling – every element in population is known and has the same chance of being selected as a subject
  - Complex probability sampling – improve efficiency by no longer having every element have the same chance of being selected
- Non-Probability Sampling
  - Convenience sampling – who of the population are available to be sampled
  - Purposive sampling – who in the population will hold the desired information

Simple Random Sampling

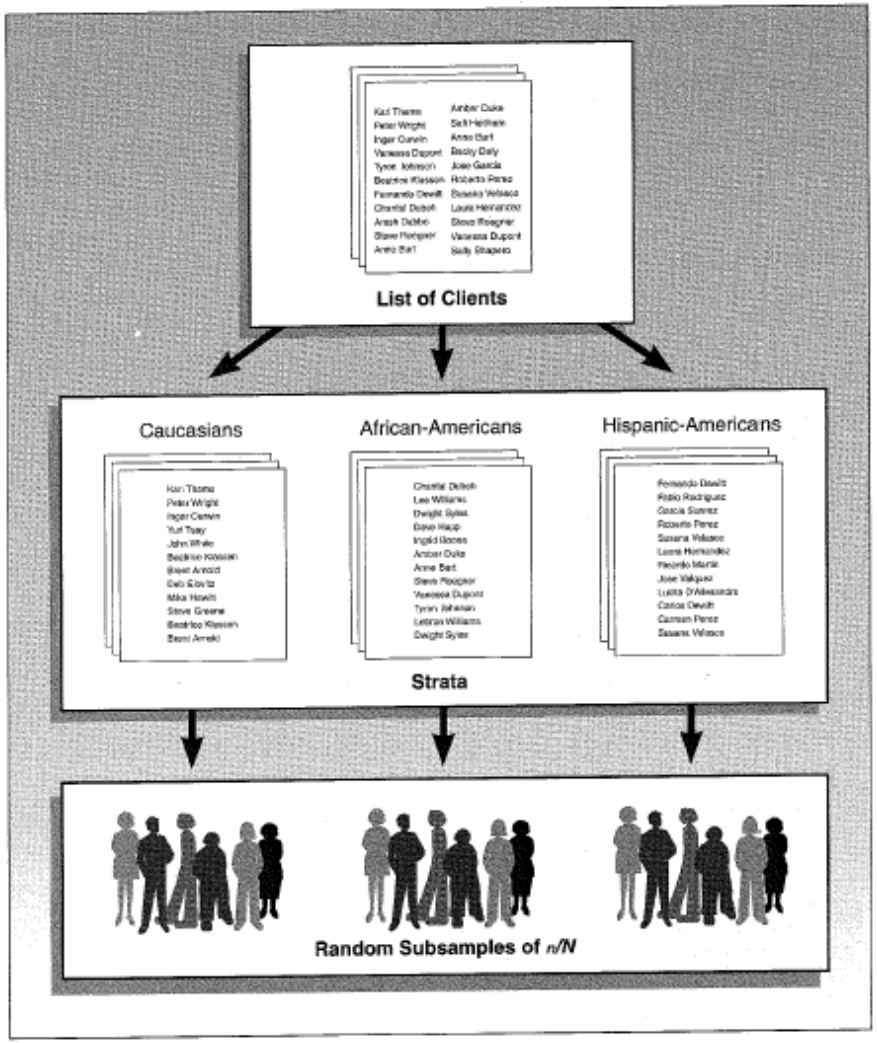
- Drawing a sample from a population so that every possible sample has an equal probability of being selected!



Source: The research methods knowledge base. W.M.K. Trochim & J. P. Donnelly. ISBN-13: 978-1-59260-290-2. Cengage Learning, USA

Stratified Random Sampling

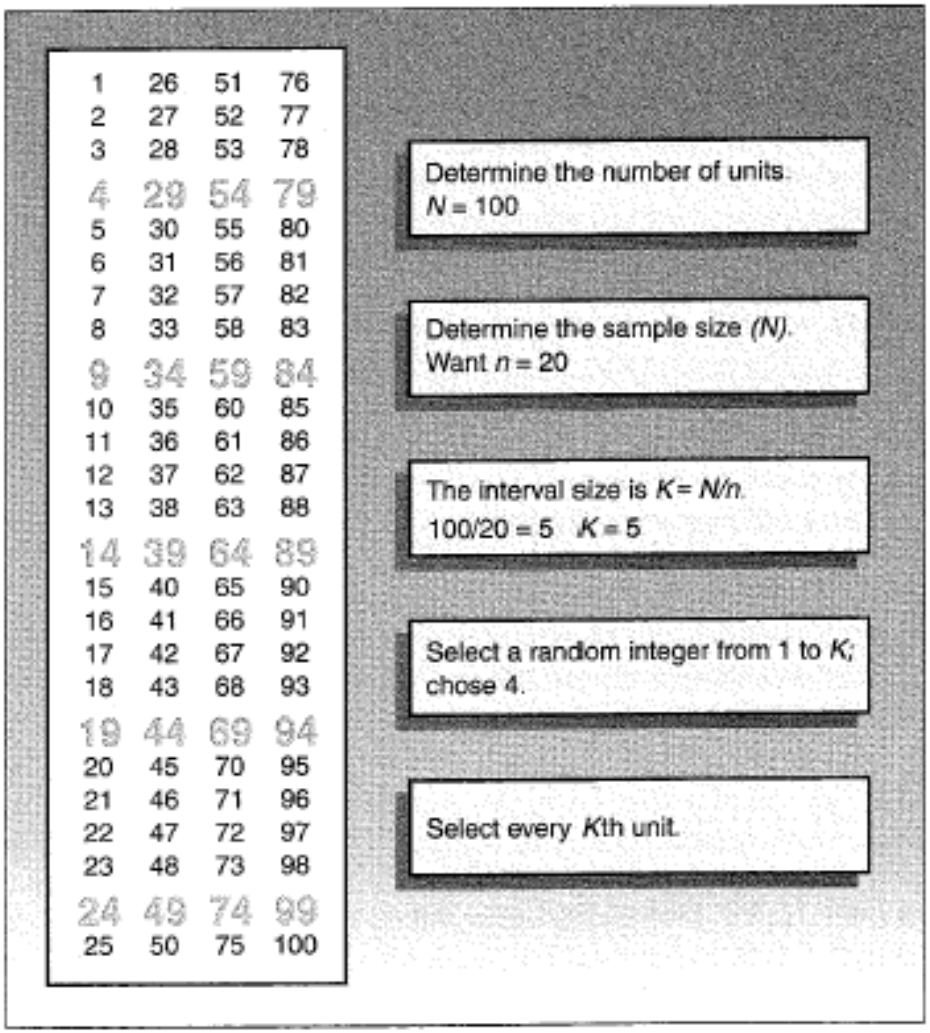
- Dividing your population into homogenous subgroups and then taking a simple random sample in each sub-group!



Source: The research methods knowledge base. W.M.K. Trochim & J. P. Donnelly. ISBN-13: 978-1-59260-290-2. Cengage Learning, USA

Systematic Random Sampling

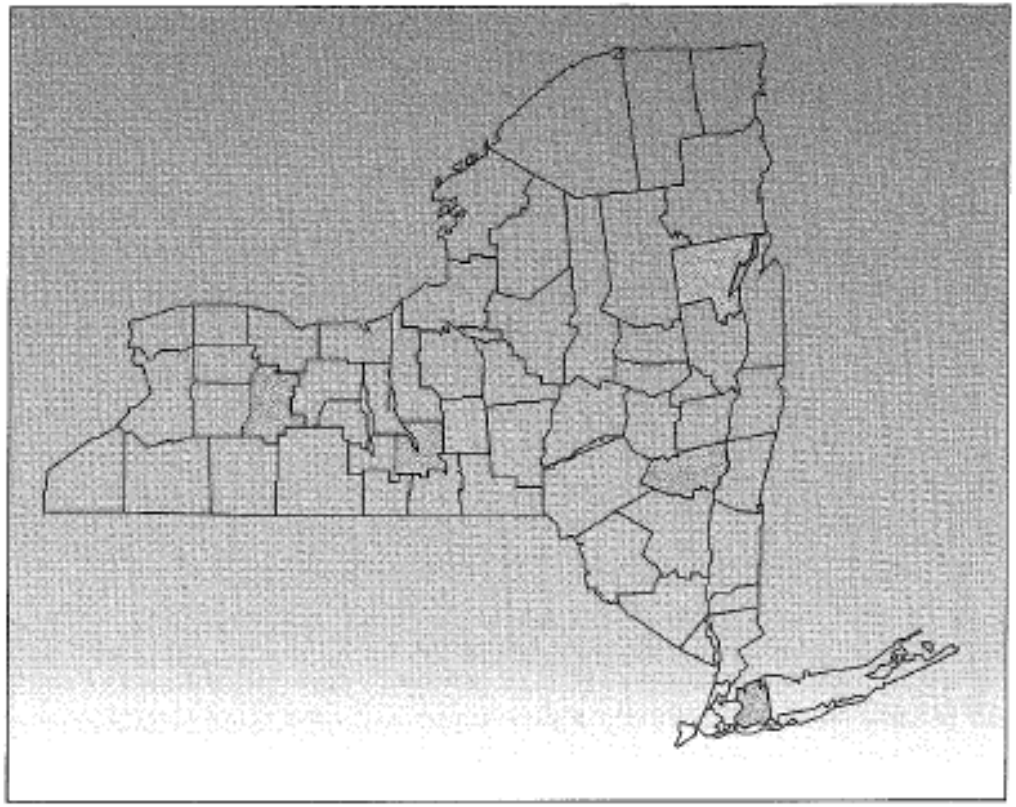
- Determine randomly where you want to start selecting and then select every n^{th} element in a randomly ordered list!



Source: The research methods knowledge base. W.M.K. Trochim & J. P. Donnelly. ISBN-13: 978-1-59260-290-2. Cengage Learning, USA

Cluster (area) Random Sampling

- Dividing population into groups, randomly selecting these clusters and then sampling each element!



Source: The research methods knowledge base. W.M.K. Trochim & J. P. Donnelly. ISBN-13: 978-1-59260-290-2. Cengage Learning, USA

Multistage Sampling

- Similar to single stage sampling but reiterating the sampling method within the sample.

- Example: Select 5 countries within EU to represent the EU, within that select 5 towns, 5 suburbs and 5 villages and within those 2 random ATM machines in order to determine the bacteria level on EURO notes

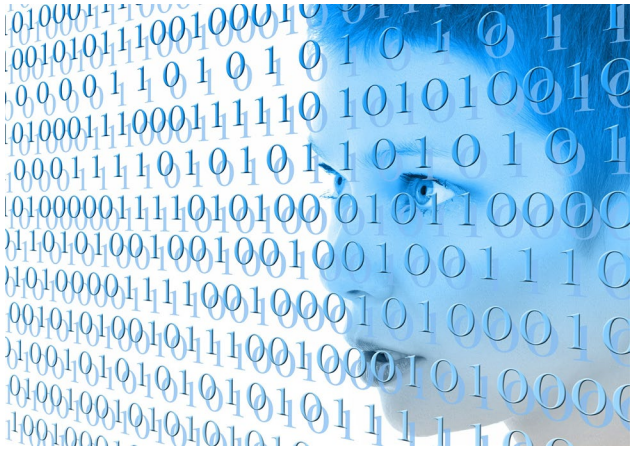
Non-Probability Sampling

- Accidental, haphazard, or convenience sampling
- Purposive sampling
  - Modal instance (most likely)
  - Expert (ask the expert!)
  - Quota (satisfying identified quotas of sub-groups)
  - Heterogeneity (all and sundry - diversity!)
  - Snowball (respondent driven sampling or pyramiding)

Response

- Response is defined as the number of elements in the sample that were successfully measured
- If the response is low your sample may no longer be representative
- The minimum required size of your response can be calculated (using sampling size formulas) or deduced from literature
- You may need to do a Non-response analysis to find out reasons why and save your research!

Data



- Generally collected from surveys, interviews, experiments, simulations etc.
- Data can be numbers, text, graphics etc.
- Hence data can be divided into two types:
  - Qualitative – observations of environment you are interested in
  - Quantitative – numerical data
- Data is generally collected in scales or can be scaled

Scales

- A Scale is defined as: The construction of an instrument that can associate qualitative constructs with quantitative metric units

- Scales can be subdivided into three types:
  - Scale – numeric values on an interval or ratio scale – i.e. age or income
  - Nominal – data values with no intrinsic order – i.e. do distinguish groups, such as male & female
  - Ordinal - data values with some intrinsic order (low, medium, high)

Data Quality

Before starting to analyze you must ensure data is of sufficient quality:

- Check for typing errors
- Check for outliers
- Check for errors in coding your data
- Check if data meets expectations

To save yourself a lot of hassle:

- Keep a good logbook of what you do
- Save your data file under a different version after changes

Data Analysis

- The purpose of data analysis is to find the answers to your research questions posed.
- It is important that you pick the right analysis tool that is Reliable and Valid for the research method you have designed and carried out.
- Every analysis tool has its pros and cons. Always carry out a trade off which tool is suitable for your type of data.

Rules of thumb on Data Analysis

- Check in literature what type of data analysis can be carried out on the data you will generate
- Check if your data meets the requirements the data analysis method has, - e.g. normally distributed data
- Run a trial with dummy data (from a test run or a known example) to test if the analysis tool produces correct data – especially if you programmed it yourself!!
- Keep a log book in which you keep track of all the data and analyses you carry out
- If necessary create a code book to order your data

Log Book

List

- Choices you made in your research design
- Assumptions you made
- How you carried out your experiments
- Any incidents during your experiments
- Where you have stored pictures & footage & relevant files
- File names and changes per file
- People you talked to (including your meetings with your supervisors)

A good log book makes writing up a walk in the park!

Code Book

- List names of variables you have given
- Serves as dictionary for the interpretation of your results
- Will also form basis for your list of symbols

Code Book

Variable Name	SPSS variable name	Description of Variable options
identification number	ID	Handwritten number on top of questtionaire
Why Aerospace engineering 1	WHYAE1	1 = Really into Aviation and/or space 2 = Most challenging/Difficult degree 3 = Wanted to be a pilot/astronaut 4 = Good reputation of degree/Delft 5 = other
other specify	OTHER	openanswer
Track Profile	MSCTRACK	1= Aerodynamics 2 = Wind Energy 3 = Control & Simulation 4 = Air Traffic Management & Airports 5 = Air Transport & Aerospace Operations 6 = Space Engineering 7 = Space Applications 8 = Design & Production of Composite Structures 9 = Novel Aerospace Materials 10 = Structural Integrity 11 = Aerospace Structures 12 = Systems Engineering and Aircraft Design 13 = Other
Career employment type Aerospace	CETAE	1 = strongly disagree 2 = disagree 3 = neutral 4 = agree 5 = strongly agree
Career employment type Management	CETMAN	1 = strongly disagree 2 = disagree 3 = neutral 4 = agree 5 = strongly agree

Interpreting Outcomes

The saying: “Garbage in leads to Garbage out” also holds for data analysis, therefore:

- Carefully look at outcomes!
  - Do they meet expectations?
  - Check if you may use the result for interpretations
  - Be very careful as to how you generalize your outcomes w.r.t. your research questions

Highly Recommended Literature

For real research practice:

- Getting it Right – R&D methods for science & engineering by P. Bock

For the practicalities of SPSS:

- The SPSS survival manual – Julie Pallant

For the scientific & statistical basis of SPSS:

- Discovering Statistics using SPSS – Andy Field