Week 1: Research Design & Data

Data Analysis for Psychology in R 1

Patrick Sturt

Department of Psychology
The University of Edinburgh

Course Overview

Exploratory Data Analysis	**Research design and data**
	Describing categorical data
	Describing continuous data
	Describing relationships
	Functions
Probability	Probability theory
	Probability rules
	Random variables (discrete)
	Random variables (continuous)
	Sampling

Foundations of inference	Confidence intervals
	Hypothesis testing (p-values)
	Hypothesis testing (critical values)
	Hypothesis testing and confidence intervals
	Errors, power, effect size, assumptions
Common hypothesis tests	One sample t-test
	Independent samples t-test
	Paired samples t-test
	Chi-square tests
	Correlation

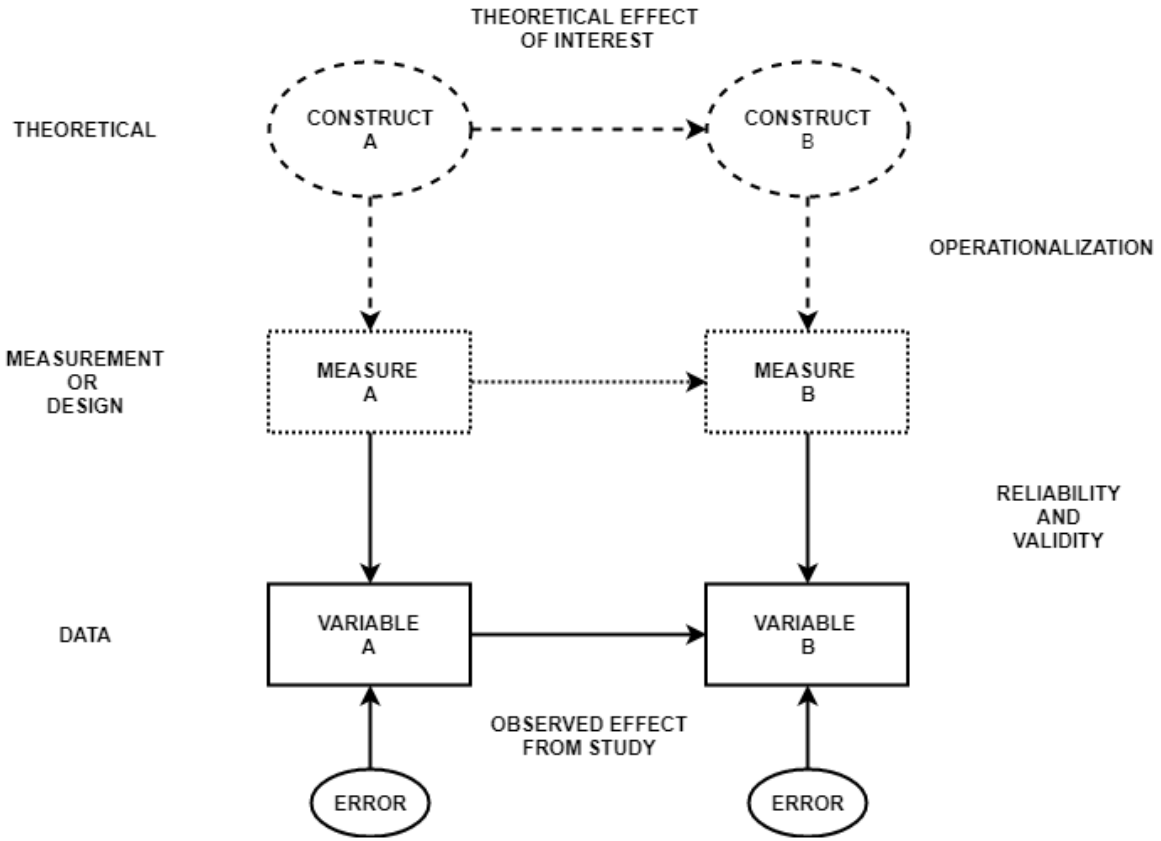
Week's Learning Objectives

1. Understand the link between study design and data.
2. Understand and define different levels of measurement.
3. Understand and define data types with psychological examples.

Topics for today

- Broad aim of measurement
- Measurement, design, and data
- Data in R

Concepts in measurement



Data types & levels

- *Categorical*
  - Nominal
  - Ordinal
  - Binary (special case)
- *Numeric*
  - Interval or ratio
  - Continuous
  - Discrete (Count)

Types of data

- **Categorical:** Variables with a *discrete* number of response options.
  - These are usually coded as integers.
  - Binary data is a special case with only 2 possible values.

Types of data

ID	Hair_colour	Likert_item	Degree
ID101	Brown	Strongly Agree	No
ID102	Brown	Agree	No
ID103	Blonde	Agree	Yes
ID104	Blonde	Disagree	Yes
ID105	Black	Strongly Disagree	Yes

- Example: Hair colour, Likert Scale items, Degree or Not?

Types of data

- **Categorical:** Variables with a *discrete* number of response options.
    - Binary data is a special case with only 2 possible values.
- **Numeric:** (continuous) Variables which can take any real number value within the specified range of measurement.

Types of data

ID	ReactionTime	Height_cm	Weight_kg
ID101	1.2	191.2	88.9
ID102	0.9	180.8	76.6
ID103	3.2	165.3	52.0
ID104	55.5	177.1	81.5
ID105	2.1	201.0	105.8

- Examples: Height in cm; Weight in kg; Reaction time

Types of data

- **Categorical**: Variables with a discrete number of response options.
    - Binary data is a special case with only 2 possible values.
- **Numeric**: Variables which can take any real number value within the specified range of measurement.
- **Count**: Variables which can only take non-negative integer values (0,1,2,3 etc.).

Levels of measurement

- Terms coined by Stevens (1946), and we are still using them!
- 4 levels are general discussed (though also critiqued - see additional reading):
    - Nominal
    - Ordinal
    - Interval
    - Ratio
- With each level, the numeric values we apply hold different meanings, and we are able to do more with the values.

Nominal data

- Binary or categorical variable where numerical markers share no relationship.
- Here is no meaningful ordering.

ID	Hair_colour	Hair_values
ID101	Brown	1
ID102	Brown	1
ID103	Blonde	2
ID104	Blonde	2
ID105	Black	3

- Example: Hair colour
  - 1 = Brown, 2 = Blonde, 3 = Black

Ordinal data

- Binary or categorical variable where there exists a meaningful way to **rank-order** responses.
- Here X < Y or Y > X statements can be made, but we can not meaningfully quantify the differences.

ID	Likert_item	Likert_values
ID101	Strongly Agree	5
ID102	Agree	4
ID103	Agree	4
ID104	Disagree	2
ID105	Strongly Disagree	1

- Example: Likert scale items
  - 1 = Strongly Disagree, 2 = Disagree, 3 = Neither A/D, 4 = Agree, 5 = Strongly Agree

Interval & Ratio

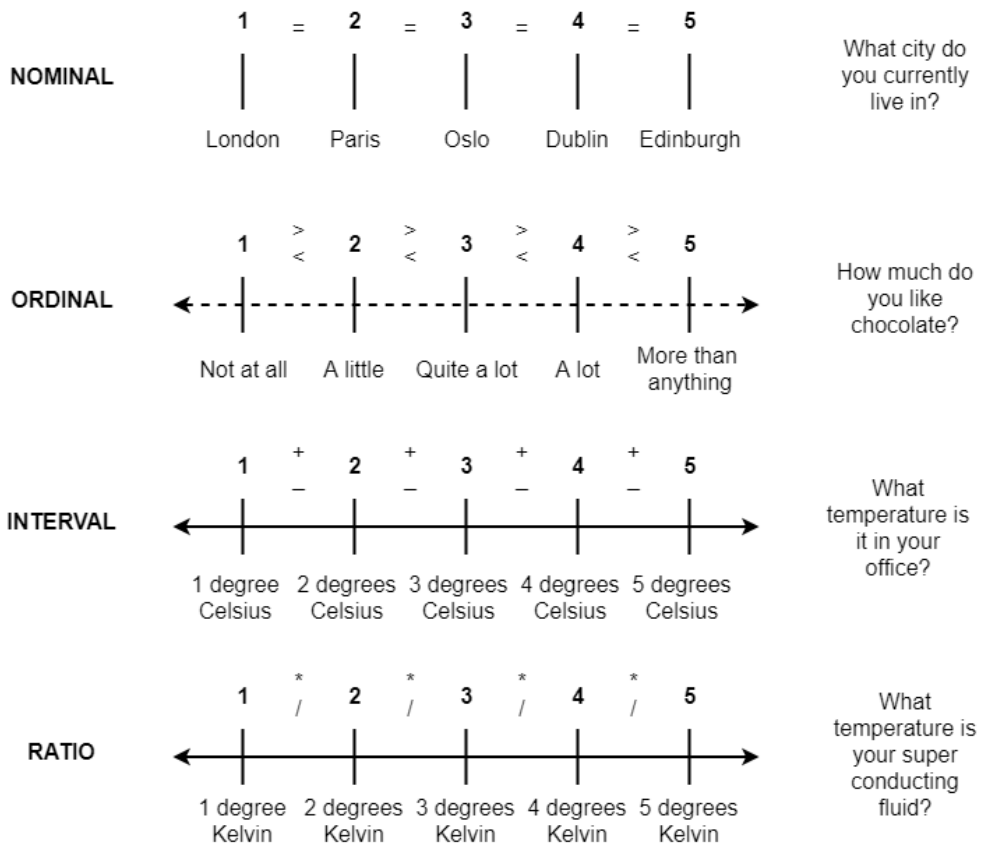
Interval data

- Variables for which numerical values have meaning.
- There is no true 0 point on an interval scale.
    - But we can consider differences.
    - And the differences have a true 0 point.
- Now it gets harder to talk about psychological examples.
    - Some would consider IQ and other test scores as interval.

Ratio data

- Variables for which numerical values have meaning.
- Variables have a true 0 point.
    - As a result, it is plausible to multiply and divide ratio variables.
    - We can legitimately talk about double X
- Some examples might be reaction time, or the firing rate of a neuron.

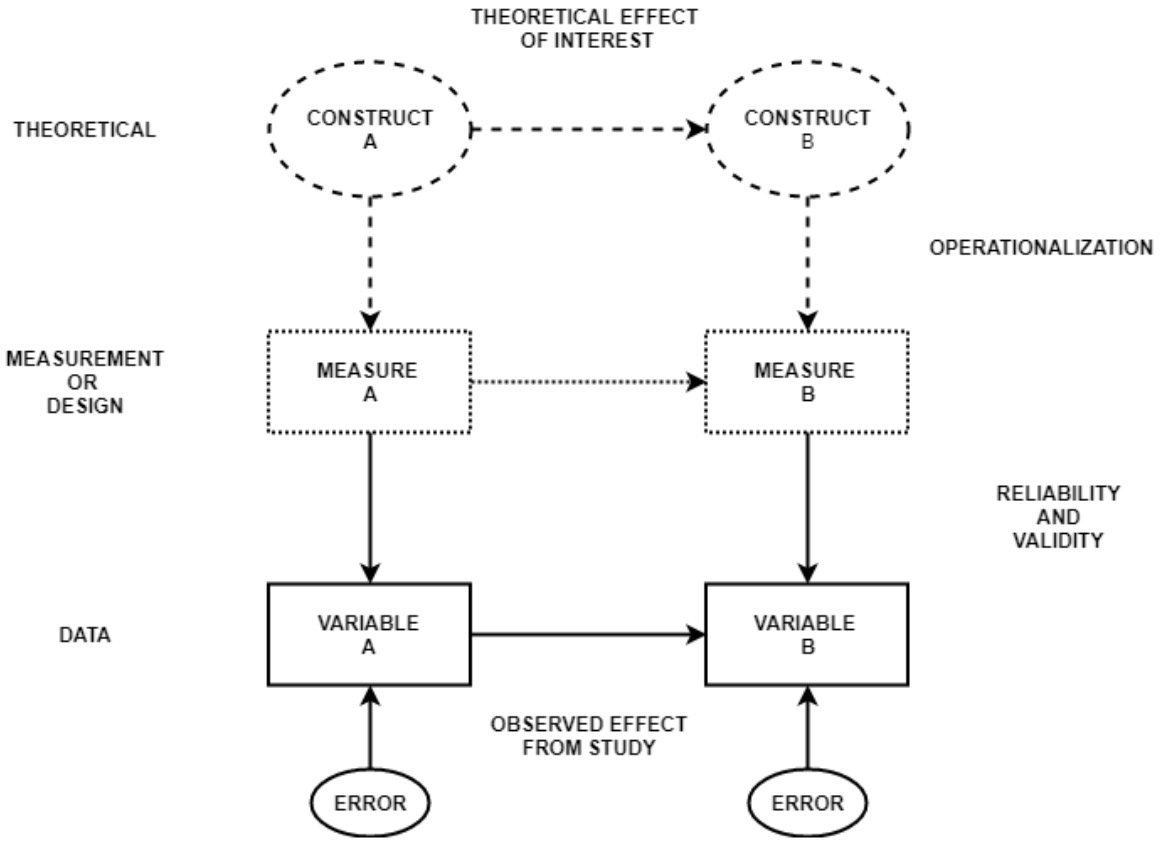
Levels of measurement



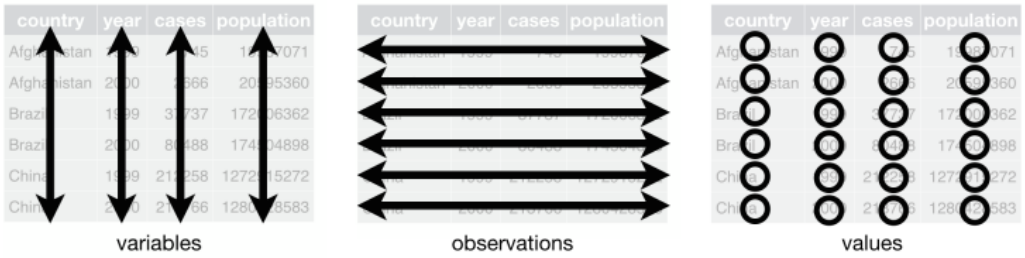
Data types and R

R Data Type	Example	Level of Measurement	Data Type
Character	ID	Nominal	(Categorical)
Numeric	Reaction Time	Interval or ratio	Continuous
Factor	Hair Colour	Nominal	Categorical
Ordered factor	Likert scale	Ordinal	Categorical

Data and data sets



Data sets


country
year
cases
population
variables
observations
values

Tidy data

1. Each variable must have its own column.
2. Each observation must have its own row.
3. Each value must have its own cell.

- This means that each individual value belongs to both a variable and an observation.

Things we need to do with data sets

- We will be constantly practising dealing with data and data sets.
- But there is a common set of things we have to do:
    - Import them into R
    - We will refer to them as data frame, data sets or tibbles
- Check each variable is of the right type
- Select columns
- Filter rows
- Recode variables
- Create variables or summaries
- Merge data sets together
- And so on...

Summary of today

- Today we have looked at the links between design and data.
- Discussed basic types of data, their properties, and the names in R.
- And briefly define what is meant by data sets and tidy data.
- All of this we will be returning to over the duration of the course.

This week

Tasks

- Attend both lectures
- Attend your lab and work together on the lab tasks
- Complete the practice quiz
    - Opens Monday at 9am
    - Unlimited attempts (unlike the assessed quizzes)

Support

- **Office hours**: for one-to-one support on course materials or assessments
  (see LEARN > Course information > Course contacts)
- **Piazza**: help each other on this peer-to-peer discussion forum
- **Student Adviser**: for general support while you are at university
  (find your student adviser on MyEd/Euclid)