Module 1: Exploring Data

= information → population (all) $\neq$?
→ sample (some) ∴ ?

Unit Overview

- M1 : Exploring Data
- M2 : Modelling Data
- M3 : Sampling Data
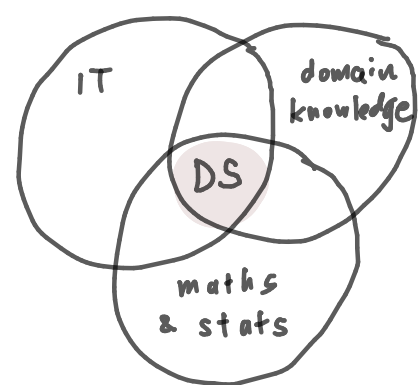- M4 : Decisions with Data

population

Sample
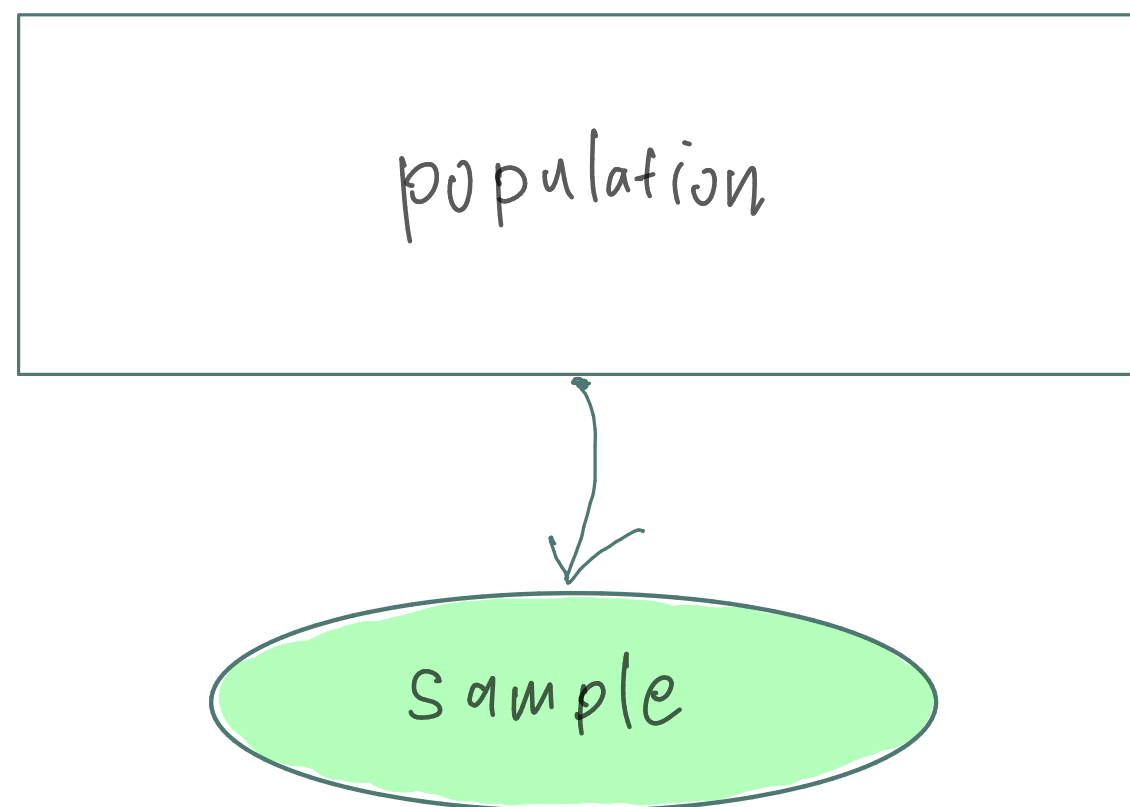
- T1 : Design of Experiments
- T2 : Data & Graphical Summaries
- T3 : Numerical Summaries

- LO1 : challenges in a data-rich world.
- LO2 : Study design behind a dataset.

Q1: Why is data science important?

= problem solving with data
("data story-teller" — professional
— popular)
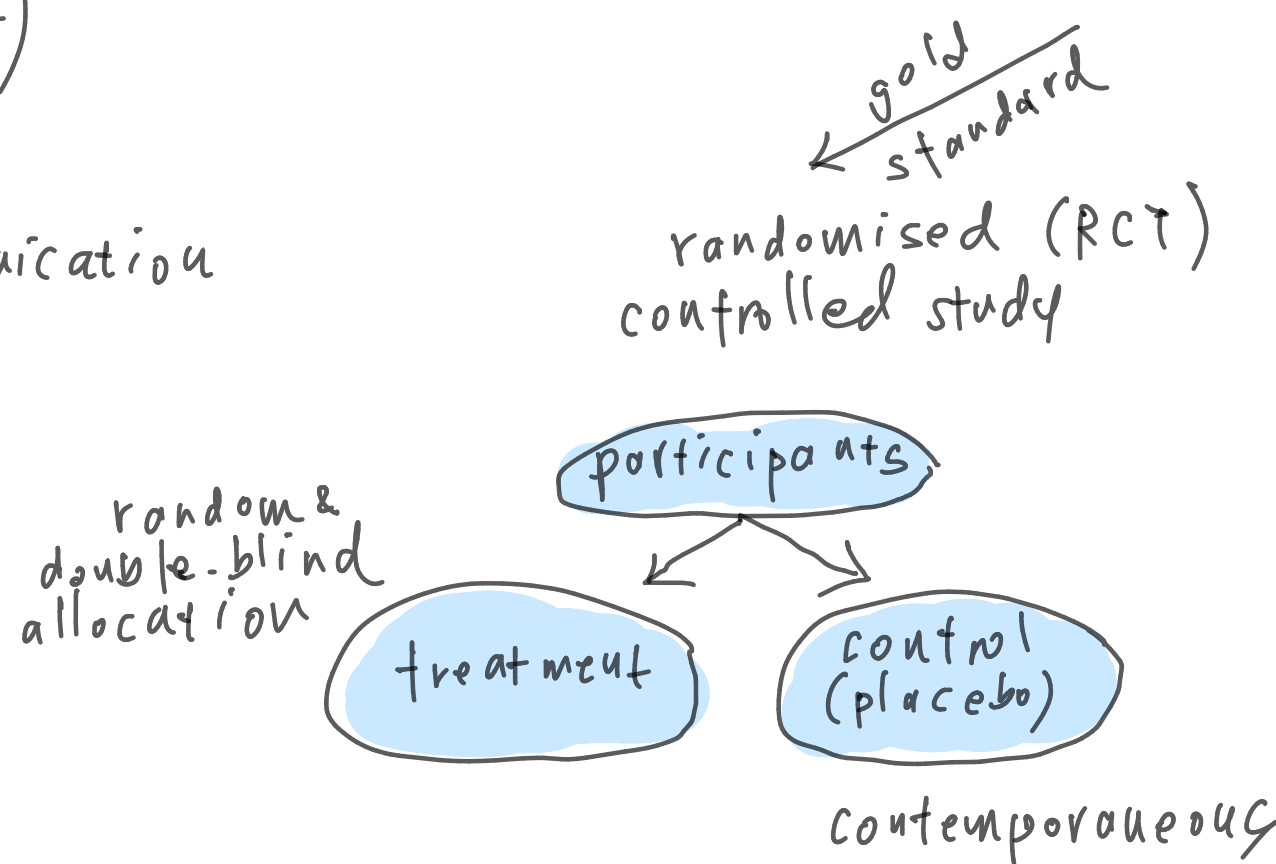

IT / domain knowledge / DS / maths & stats

curiosity; collaboration, communication

challenges: ethics, privacy

Note in Discover Video 3:

- Causation DV3/12
- Controlling for confounder DV3/21
- Simpsons Paradox DV3/23-26
- 3 uses of control DV3/33

Q2: where does data come from?

confounders: selection bias, observer bias, consent bias, survivor bias, adherer bias ...

gold standard

randomised (RCT) controlled study

observational study

participants

random & double-blind allocation

treatment

control (placebo)

contemporaneous

more challenges
- cant prove causation
- hidden confounders
- Simpsons Paradox
- historical control