Quantitative Methods

Human Sciences, 2020–21

Elias Nosrati

Lecture 9: 16 February 2021

Overview

- Syllabus review (document on Canvas).
- Today: broad overview.
- Next week: review of causal inference.

Structure of the syllabus

1. Probability.
2. Inference.
3. Data analysis.

Probability

- Describe the three things we need to define a probability:
  1. A sample space S.
  2. A class of well-defined events: A, B, A^c, $A - B$, etc.
  3. A probability function $\mathbb{P} : S \to [0, 1]$.
- Describe and justify the three axioms of probability:
  1. $\mathbb{P}(A) \geq 0$ for any event A.
  2. $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(S) = 1$.
  3. If $A_1, \ldots, A_n$ are mutually exclusive events, then

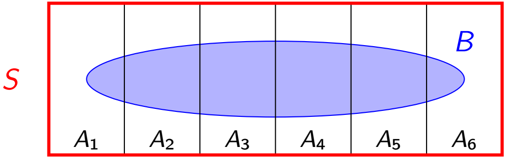
$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n).$$

- Be able to count, and explain counting identities using story proofs. (How many ways are there to form a queue of n people?)

Probability (cont.)

- Define and explain the following:
  - ⋆ Conditional probabilities: $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
  - ⋆ Bayes's Rule: $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$.
  - ⋆ The Law of Total Probability: for a partition $A_1, \ldots, A_n$ of the sample space,

$$\mathbb{P}(B) = \mathbb{P}(A_1)\mathbb{P}(B \mid A_1) + \cdots + \mathbb{P}(A_n)\mathbb{P}(B \mid A_n).$$



Probability (cont.)

- Describe and explain the following:
  - (In)dependence:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B \mid A) = \mathbb{P}(B)\mathbb{P}(A \mid B).$$

  - Conditional (in)dependence:

$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\mathbb{P}(B \mid A \cap C) = \mathbb{P}(B \mid C)\mathbb{P}(A \mid B \cap C).$$

Probability (cont.)

- Define a (discrete or continuous) random variable as a function $X : S \to \mathbb{R}$ that assigns a numerical value to each possible outcome of an experiment and identify (in)dependence structures between random variables.
- Define and describe probability distributions associated with random variables, especially:
  - Bernoulli, Binomial, Poisson, Uniform, Normal distributions.
  - Recognise the two things needed for a function f to count as a probability distribution:
    1. $f(x) \geq 0$ for all x.
    2. $\sum f(x) = 1$ (discrete) or $\int f(x)dx = 1$ (continuous).
- Describe and compute the key features of random variables with respect to their probability distributions.
- Describe the Law of Large Numbers and Central Limit Theorem and their implications.

Inference

- Understand and describe the key problem of statistical inference: from $\mathbb{P}(\text{data} \mid \text{model})$ to $\mathbb{P}(\text{model} \mid \text{data})$.
- Understand and describe the likelihood theory of inference.
- Understand and describe the Bayesian theory of inference and its relation to the likelihood theory of inference.
- Understand and describe the principles of ordinary least squares estimation.
- Understand and describe what constitutes a statistical model, its systematic and stochastic components, and its key assumptions.
- Understand, compute, and interpret regression models of the form $y_i = \alpha + x_i\beta + \epsilon_i$ and assess the distributional assumptions surrounding the error term.

Inference (cont.)

- Assess when a model parameter β might have a causal interpretation by reasoning in terms of counterfactuals (or potential outcomes).
- Describe the fundamental problem of causal inference: only one potential outcome is observed for any individual.
- Describe the key characteristics of, compare and contrast, and critically assess the strengths and weaknesses of randomised controlled trials and observational studies.
- Define and visualise using causal graphs the three main forms of systematic bias — confounding, selection, and measurement bias — that can undermine causal inferences.

Data analysis

- Be able to import data into **R**.
- Be able to tidy data in **R**.
- Be able to simulate from and compute key features of the most important probability distributions (e.g., `rnorm()`, `mean()`, etc.).
- Be able to specify linear regression models using `lm()` and interpret model outputs.
- Be able to visualise data using `ggplot()` or `plot()`.
  - A clear purpose, simple message.
  - Clearly annotated (scales, labels, captions).
  - Easy to interpret.
- Write clean and replicable code in **R**.