Statistics for Linguists

Adam J.R. Tallman

2022-10-18

Course management/grading

- Moodle
  - Slides, homeworks, textbook, other optional readings, databases for exercises
- Project
  - Not graded for homeworks, draft of final paper, final paper, project description with annotated bibliography

Textbooks

STATISTICS FOR LINGUISTS

AN INTRODUCTION USING R

BODO WINTER

ROUTLEDGE

The Seven Pillars of Statistical Wisdom

STEPHEN M. STIGLER

How to do Linguistics with R

Data exploration and statistical analysis

Natalia Levshina

John Benjamins Publishing Company

Starting

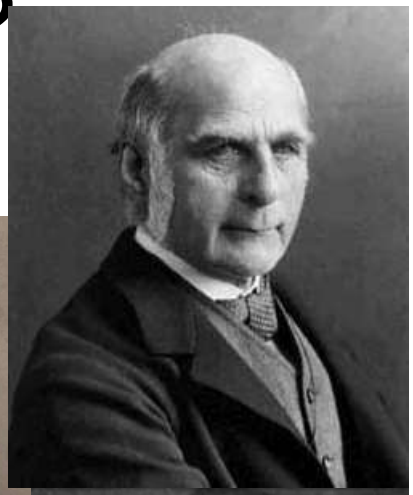
- Download R
  - http://lib.stat.cmu.edu/R/CRAN/
- Download R studio
  - https://www.rstudio.com/products/rstudio/download/

Course outline

- Using R
- Descriptive statistics
  - summarizing aggregated data, means, uncertainties etc... graphing etc.
- p-values
  - Frequentist vs. Bayesian statistics, t-tests, chi-square tests, p-values etc.

P-values are controversial?

- p-values are controversial
- They have been adopted as a standard in the reporting of scientific results
- Almost everyone seems to misinterpret them
- Even the people who invented them might have been ‘misinterpreting’ them in their own work.
- I’m not sure if I should be teaching p-values
  - they might be contributing to bad scientific practice



Course outline

- Linear models
  - when you have two quantitative variables that change together
  - called ‘regression’ for historical reasons, but maybe this is a bad term
- Multivariate linear model
  - When you have more than two quantitative variables that change together

Course outline

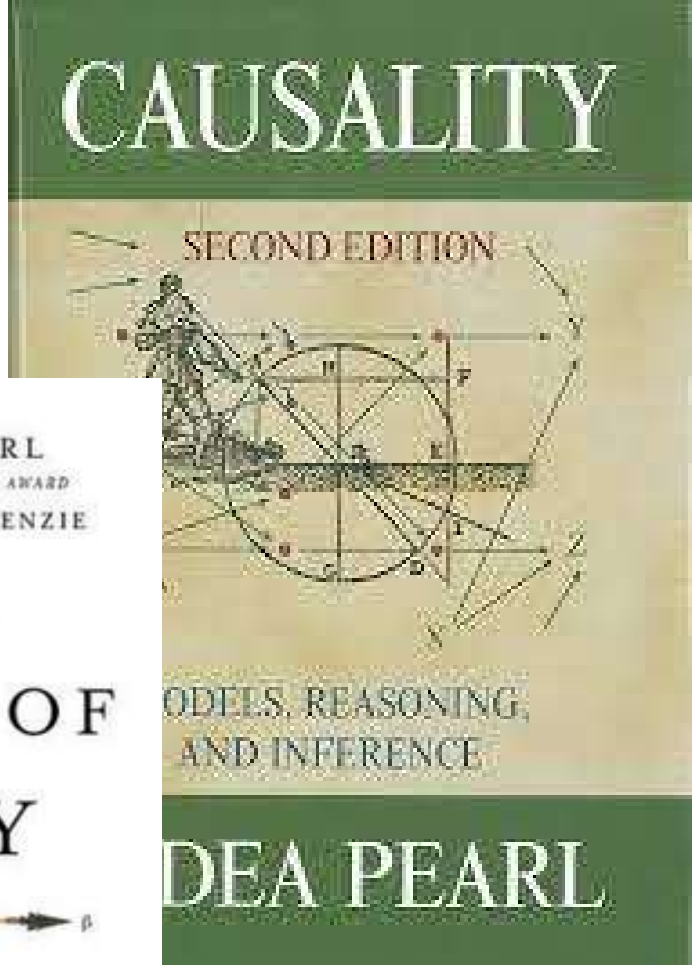
- Project descriptions
  - I want you to start reading about statistics on your own relevant for a particular topic of interest
  - That’s more like how research is typically done
  - So you are partially being assessed on how well you can teach yourself statistics

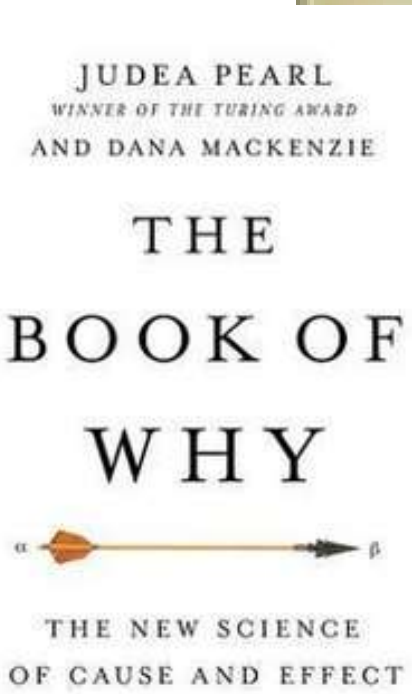
Course outline

- Binomial / logistic models
  - Models with S-shaped curves
- Causal inference
  - How do we tell causation from correlation?
- Interactions
  - What if knowing some other value influences your understanding of how your variables of interest are correlated.

Causal inference

- Causal salad means that you throw in lots of variables without thinking about causal structure – done a lot, but its bad.
- Causal inference is a set of tools for thinking about how statistical models relate to causal predictions





Course outline

- Multilevel models
  - When you are looking at relationships across different groups and you also want to see variation between the groups.
- Exploratory data analysis
  - When you are trying to figure out what your data means but you aren’t sure you have a clear hypothesis.
  - You are trying to see what you’re data might show
  - Sort of an extension of descriptive statistics

Payment on a loan

$$R = P\left(\frac{i}{1-(1+i)^{-n}}\right)$$