Introduction to Statistics with Randomization and Simulation

Mine Çetinkaya-Rundel, Johanna Hardin, others…

2020-01-09

Contents

Preamble

Need to move preamble here.

Chapter 1

Introduction to data

1.1 Case Study (capable of extending to MLR or 2 by 2 table)

1.2 Taxonomy of Data

1.3 Overview of data collection principles

1.4 Observational studies and sampling strategies

1.5 Experimental design and causality

1.6 Revisit case study with new terminology we learned

Chapter 2

Exploratory Data Analysis

2.1 Cat vs. cat - segmented plots / contingency tables

- Conditional probability from contingency tables
- Bayes Theorem (law of total probability?)

2.2 Num vs. cat - side-by-side box plots / comparing distributions

- Mention univariate - center, skew, shape, spread
- Mention conditional probabilities as well

Chapter 3

Correlation and Regression

3.1 Visual summaries of data: scatterplot, side-by-side boxplots, histogram, density plot, box plot (lead out with multivariate, follow with univariate)

3.2 Describing distributions: correlation, central tendency, variability, skew, modality

3.3 Num vs. num - SLR

- correlation
- Line fitting, residuals, and correlation
- Fitting a line by least squares regression
- Types of outliers in linear regression

Chapter 4

Multiple Regression

4.1 Num vs. whatever - MLR

- Introduction to multiple regression

4.2 Parallel slopes

4.3 Hint at interaction, planes, and parallel planes but not quantify

- Visualization of higher-dimensional models (rgl demo)

4.4 Logistic regression

- Binary vs. num/whatever
- Three scales interpretation (e.g. probability, odds, log-odds)
- “parallel” logistic curves?

Chapter 5

Foundations of inference

5.1 Understanding inference through simulation

5.2 Randomization case study: gender discrimination

5.3 Randomization case study: opportunity cost

5.4 Hypothesis testing

5.5 Confidence intervals

5.6 Simulation case studies

Chapter 6

Inference for categorical data

6.1 Inference for a single proportion

- Simulation
- Exact (if we include course on probability)
- CLT and Normal approximation

6.2 Difference of two proportions

6.3 Testing for goodness of fit using chi-square (special topic, include simulation version)

6.4 Testing for independence in two-way tables (special topic)

Chapter 7

Inference for numerical data

7.1 One-sample means

- Bootstrap (for means, medians)
- t-distribution

7.2 Paired data

7.3 Difference of two means

7.4 Comparing many means with ANOVA (special topic, include simulation version)

Chapter 8

Inference for regression

8.1 Inference for linear regression

- Bootstrap for regression coefficients
- t-distribution for regression coefficients
- Model Comparison: Occam's Razor and R^2 > R^2_adj

8.2 Checking model assumptions using graphs

- L-I-N-E

8.3 Inference for multiple regression

- residuals vs. fitted instead of residuals vs. x

8.4 Inference for logistic regression

Chapter 9

Appendix: Probability

(Keep same content as before, minus the bit of probability that got moved to categorical EDA)