Contents

Statistics 159/259 Course Summary

This version: November 28, 2018

Concepts

Terms related to reproducibility

- reproducibility
- replicability
- repeatability
- computational reproducibility
- "preproducibility"
- the role of replication in science
- "virtual witnessing" and the role(s) of scientific publishing

Obstacles to reproducibility

- data availability
  - data
  - data format
  - data dictionary
  - data cleaning and munging
  - data pre-processing
- reliance on proprietary software
- analysis

- breadcrumbs / description
- actual code
- description and what was done are often different
- scripting is key, but not enough
- software versions, libraries, environments can matter

Obstacles to replicability

- lack of preproducibility: what was done?
- "researcher degrees of freedom"
  - what was considered but not tried, or tried and discarded?
  - choice of hypotheses, P-hacking
  - choice of data subsets
  - choice of transformations
  - choice of models
  - choice of estimators
    - if Bayesian, choice of prior
    - if frequentist, what method and why?
    - constraints?
  - choice of measures of uncertainty
    - nonparametric / model-based / parametric / asymptotic
    - local / global
    - selective inference
    - hypothesis tests: what is the full null? What does it have to do with reality?
- "file-drawer effect"
  - small n studies
- ignoring multiplicity & multiple testing
- intrinsic variability
- sensitivity to "influential" observations
- appropriate level of abstraction

Obstacles to good science and applied Statistics

- confirmation bias
- Foundational issues; misinterpretations of probability and uncertainty
  - Interpretation of probability
    - prior probabilities
  - Types of uncertainty

        - Epistemic and aleatory uncertainty
        - constraints versus priors
    - Bayesian and frequentist measures of uncertainty
    - Duality between minimax and Bayes estimation
    - models versus response schedules

- model mania
    - correlation is not causation
    - fit does not imply correctness
    - statistics and superstition

- ritualization of Statistics, cargo-cult science

- bad incentive structure in academia

Key ideas/tools from software engineering that can help improve science

- revision/version control
- documentation, documentation, documentation
- modularity and abstraction
- scripted analyses and automation
- unit tests, regression tests, coverage tests, continuous integration
- code review
- pair programming
- consistency: APIs, calling signatures, object-oriented code
- separating data, computation, presentation

Testing, statistical models, sensitivity/stability

- It's all about the null hypothesis
    - null has to let you find sampling distribution of the test
    - if the null is not appropriate, the test is not appropriate
        - examples: t-test in Karp et al.

- P-values: $\Pr\{P \le p || H_0\} \le p$.

- Multiple testing, multiplicity, multiplicity adjustments
    - Bonferroni's inequality
    - Independence
    - False discovery rate (not covered)

- The Neyman model for causal inference
  - potential outcomes
  - strong null and weak null
  - non-interference
  - responses can be distributions
  - honor the randomization!
- Permutation tests
  - nulls that imply invariance of the probability distribution under a group
  - has to match the real world
  - generating random permutations
    - comparison of PRNGs, algorithms for generating random integers, sampling algorithms
  - simulation to estimate P-values; randomized tests to find conservative P-values
  - permutation tests for regression, two-sample test, etc.
- Goodness of fit tests
  - Kolmogorov-Smirnov
  - Chi-square statistic
    - asymptotic tests versus exact tests versus conservative tests
  - other tests
- Intersection-union tests and stratified tests
  - combining information from different tests
  - combining functions, including Fisher's combining function
- Fixed-n tests versus sequential tests
  - Wald's sequential probability ratio test
- Models versus response schedules
  - Response schedules and "physics."
  - common models
    - regression
      - assumptions required to perform OLS
      - assumptions required for OLS to be unbiased
      - assumptions required to compute SE
      - assumptions required for $\hat{\beta}/SE$ to have a t-distribution
    - Linear probability models
    - Logit and probit models
    - Poisson regression
      - MLE for Poisson regression
- Sensitivity analysis and sensitivity auditing