STATS 275: Statistical Consulting
Preliminary and Formal Analysis [1]

Babak Shahbaba

Department of Statistics, UCI

[1] Based on "Principles of Applied Statistics" by Cox and Donnelly

Data exploration

- Data screening: assessment of data quality
  - Examining measurements
  - Examining summary statistics such as mean, standard deviation, min, max, ...
  - Identifying data outliers (we don't remove outliers unless they arise from errors that cannot be fixed)
  - Detecting patterns of missing data
  - Assessing pairwise relationships to detect redundancy
- Preliminary graphical analysis
- Preliminary tabular analysis

Model formulation

- Next, we typically specify probability models representing the underlying mechanism of the observed data
- Model parameters usually capture our scientific questions
- Types of models
  - parametric, nonparametric, or semiparametric
  - substantive (specific to the underlying subject matter) or purely empirical (typical statistical models)
- Models specifications usually decompose systematic and random variations

Formal inference

- After specifying the model, we perform formal statistical inference
- Inference typically involves making decisions: hypothesis testing, estimation, or prediction
- Two general frameworks for statistical inference are frequentist (likelihood-based) and Bayesian
- Bayesian inference involves specifying priors and obtaining posterior distributions, which are used to make inference
- Frequentist inference usually involve estimation (point or interval) and significance tests (cautions should be taken when a large number of significance tests are performed simultaneously)

Post analysis

- Model checking and diagnosis
- Tabular and graphical presentation of results
- Interpretation of results in clear, plain language, and in the subject-matter context
- Publication!