Study Guide

STAT 21

Swarthmore College

Review material

- Class 1 (beef demand data)
- Class 2 (body temperature data)

Introductory material

- Class 3 importing data and using R (body temperature data)
- Class 4 data manipulation and visualization in the *tidyverse*
- Class 5 Statistical terms and concepts; QQ plots
- Homework 1

Simple linear regression

- Class 6 Quantiles; SLR model
- Class 7 Normal probability plots (healthcare data)
- Class 8 SLR model and assumptions; central limit theorem (diamond and healthcare data)
- Class 9 Assumptions for estimation (healthcare data)
- Homework 2
- Class 10 Violations of estimation assumptions; R output
- Class 11 Sums of squares; Coefficient of determination
- Class 12 Assumptions for inference; Types of inference (e.g. tests)
- Homework 3 Variable transformation, QQ plots, interpreting regression model, modeling with categorical predictors (ANOVA)
- Class 13 Types of inference (e.g. confidence and prediction intervals) (restaurant data)
- Class 14 Exam I review class; Residual plots; Deterministic vs. statistical relationships (correction slide for Class 13)

ANOVA

- Class 15 Assumptions; Notation; R output (mammal data)
- Class 16 F-test
- Class 17 Chi squared test for independence (Star Trek data)
- Class 18 Interpreting ANOVA coefficients; Degrees of freedom; Power analysis
- Homework 4 prediction vs confidence intervals, ANOVA power analysis, defining research questions

Multiple linear regression

- Numerical predictors and numerical response
  - Class 19 Notation; Interpreting R output; Residual plots (SAT data)

  - Class 20 Effects of adding more predictor variables (SAT data)
  - Class 21 Visualizing MLR (SAT data)
  - Class 22 R-squared and Simpson's paradox
  - Homework 5 conceptualizing definitions and Simpson's paradox
  - Class 23 Interaction terms and collinearity
- Numerical and categorical predictors and numerical response
  - Class 24 Interaction between caegorical and numerical predictors
  - Class 25 ANOVA as a MLR model, dummy variables
  - Class 26 MLR with both numerical and categorical predictors, visualizing interaction effects
  - Homework 6 writing functions in R, interpreting interaction effects for categorical and numerical predictors
  - Class 27 distinctions between correlated variables and a linear relationship among variables, one-way ANOVA interpretations
  - Class 28 data-inherent collinearity, MLR estimation and prediction (wine data)
  - Class 29 data-inherent collinearity, MLR estimation (wine data)
  - Class 30 MLR estimation and inference (wine data)
  - Homework 7 understanding the meaning of "statistical significance", writing functions in R, complete regression analysis of airplane data