Intermediate Statistical Investigations: A Second Course

Preliminaries: Multivariable Thinking and Sources of Variation

- *Identify and apply basic terminology of statistical studies: observational units, response*
- *variable, explanatory variable, association, confounding variable*
- *Identify potential sources and measures of variation in a response variable*
- *Produce and describe some basic visualizations and numerical summaries to compare groups*
- *and explore relationships (e.g., bar graphs, dotplots/histograms/boxplots, scatterplots, means,*
- *medians, standard deviation)*
- *Explore how those comparisons and relationships can be impacted by additional variables*
- *Calculate a residual and relate it to typical prediction error*
    - Example P.A: Graduate School Admissions
    - Exploration P.A: Salary Discrimination
    - Example P.B: Predicting Birthweights
    - Exploration P.B: Housing Prices in Michigan

Chapter 1: Source of Variation

- Build simple statistical models to formally capture and summarize important sources of variation in a variable of interest.
- Section 1.1: Sources of Variation in an Experiment
    - *Distinguish experiments and observational studies*
    - *Review basic study design principles such as inclusion criteria and random assignment*
    - *Define terminology specific to an experimental study (e.g., treatments)*
    - *Produce a Sources of Variation diagram for an experiment*
    - *Apply the six-step investigative process*
        - Example 1.1: Scents and Consumer Behavior
        - Exploration 1.1: Memorizing Letters
- Section 1.2: Quantifying Sources of Variation
    - *Partitioning variation in the response variable into variation explained by the model and unexplained variation*
    - *Measuring percentage of variation explained*
    - *Understanding effect size and practical significance*
        - Example 1.2: Scents and Consumer Behavior cont
        - Exploration 1.2: Starry Navigation
- Section 1.3: Is the Variation Explained Statistically Significant?
    - Assess the statistical significance of a two-group comparison
    - Carry out and evaluate a randomization test comparing two groups on a quantitative response variable
    - Apply two-sample *t*-procedures for tests of significance and confidence intervals
        - Example 1.3: Scents and Consumer Behavior cont
        - Exploration 1.3: Starry Navigation cont
- Section 1.4: Comparing Several Groups
    - *Compare more than two treatments using randomization tests*
    - *Calculate an F-statistic and use the F distribution to find theory-based p-values*
    - *Complete an Analysis of Variance table*
    - *Assess the validity of an F-test*
        - Example 1.4: Fish consumption and Omega-3
        - Exploration 1.4: Golden Squirrels

- Section 1.5: Confidence Intervals and Prediction Intervals
    - *Post-hoc analysis after significant F test (pairwise differences)*
    - *Confidence intervals on single means*
    - *Prediction intervals on quantitative variables*
    - *Factors that impact widths of confidence and prediction intervals*
        - Example 1.5: Fish consumption and Omega-3 cont.
        - Exploration 1.5: Golden Squirrels cont
- Section 1.6: Power and Sample Size
    - *Understand statistical power and how it is impacted by sample size, variability within groups, number of groups, and significance level*
    - *Use statistical power analysis to plan the sample size of a study*
        - Example 1.6: Fish consumption and Omega-3 cont.
        - Exploration 1.6: Who is Spending More Time on Parenting?
- Chapter 1 Investigation: Tai Chi and Functional Reach

Chapter 2: Controlling Additional Sources of Variation

- *Analyze paired data appropriately*
- *Extend matched pairs analyses to repeated measures with more than two measurements*
- *Distinguish between a completely randomized design and a randomized block design*
- *Analyze a randomized block design using simulation and two-way ANOVA*
- *Analyze an observational study with a variable of interest and a nuisance variable*
- Section 2.1: Matched Pairs
    - *Use pairing to potentially reduce unexplained variation and increase the power of a study*
    - *Explain how to analyze paired data appropriately*
        - Example 2.1: Car simulator (Facebook vs. Instagram)
        - Exploration 2.1: Chip melting times
- Section 2.2: Randomized block designs
        - Example 2.2: Strawberries
        - Exploration 2.2: Finger Tapping
- Section 2.3: Observational studies with two explanatory variables
        - Example 2.3: Salary discrimination
        - Exploration 2.3: Alcohol consumption and depression
- Chapter 2 Investigation: Cholesterol Medication

Chapter 3: Multi-factor Studies and Interactions

- *How to design and analyze a multi-factor experiment*
- *Continue to explore three-way associations using multiple explanatory variables*
- *Understand the concept of an interaction*
- Section 3.1: Multi-factor Experiments
    - *Designing an experiment with more than one treatment variable*
    - *Exploring the benefits of a two-variable analysis over a one-variable analysis*
        - Example 3.1: Corporate Credibility, Endorser, and Purchase Intent
        - Exploration 3.1: Pig Growth
- Section 3.2: Statistical Interaction
    - *Understand the concept of an interaction*
    - *Calculate interaction effects*
    - *Interpret an interaction plot*
    - *Use simulation and theory-based p-values to assess the significance of an interaction*

        - Example 3.1: Pistachio Bleaching
        - Exploration 3.2: Optimizing Ads
- Section 3.3: Replication
    - *Define and describe advantages of a generalized block design*
    - *Define and describe advantages of within-block factorial designs*
        - Example 3.3: Optimizing Vitamin C
        - Exploration 3.3: Hurricane Names
- Section 3.4: Interactions in Observational Studies
    - *Interpret interactions with observational data*
        - Example 3.4: Salary Discrimination cont.
        - Exploration 3.4: Hopelessness and Exercise

Chapter 4: Adding Quantitative Variables

- *Review of simple linear regression*
- *Explore simulation models for assessing the strength of evidence of a linear association between two quantitative variables*
- *Utilize residual plots to explore model assumptions*
- *Use ANOVA tables to explore the contribution of different sources of variation in the response variable*
- *Build regression models with multiple quantitative and categorical explanatory variables and their interactions*
- Section 4.1: Linear Regression
    - *Describe the association between two quantitative variables numerically and graphically*
    - *Interpret least-squares regression models between two quantitative variables*
    - *Compare and contrast separate means vs. linear regression models*
        - Example 4.1: Recovering Polyphenols from Grape Seed
        - Exploration 4.1: Fatty Acids and DNA
- Section 4.2: Inference for Linear Regression
    - *Carry out simulation-based inference to assess the evidence of a linear association between the quantitative explanatory and response variables*
    - *Use a theory-based approach to assess the evidence of a linear association between the quantitative explanatory and response variables*
    - *Evaluate the validity of the theory-based approach using residual plots*
        - Example 4.2: Recovering Polyphenols from Grape Seed
        - Exploration 4.2
- Section 4.3: Quantitative and Categorical Explanatory Variables
    - *Adjusting the relationship between two quantitative variables based on a categorical variable*
    - *Create indicator variables for including binary categorical variables in the regression model*
    - *Evaluate the appropriateness of the regression model*
        - Example 4.3: Michigan Housing Prices
- Section 4.4: Two Variable Model with Interaction
    - *Include interaction between quantitative and categorical variables in a the statistical model*
    - *Interpret the nature of the interaction*
        - Example 4.3: Michigan Housing Prices cont
        - Exploration 4.3: FEV and smoking
- Section 4.5: Multi-level Categorical Variables

        - Example 4.5: Diamonds
        - Exploration 4.5: Patient Satisfaction

Chapter 5: Multiple Quantitative Explanatory Variables

- Section 5.1: Multiple Quantitative Variables
    - Example 5.1: Pistachio Bleaching
    - Exploration 5.1: Biodiesel
- Section 5.2: Observational Studies
    - Example 5.2: Car Acceleration
    - Exploration 5.2: Brain Size and IQ
- Section 5.3: Multicollinearity
    - Example 5.3: Hopelessness
    - Exploration 5.3:Brain Size and IQ
- Section 5.4: Centering and Standardizing
    - Example 5.4: Harris Bank Data
    - Exploration 5.4: Reducing Body Inflammation
- Section 5.5: Modeling Nonlinear Associations—Part 1
    - Example 5.5: Artic Sea Ice
    - Exploration 5.5: Kentucky Derby Winning Times
- Section 5.6: Modeling Nonlinear Associations—Part 2
    - Example 5.6: Salary Discrimination cont
    - Exploration 5.6: Stopping Distances
    - Exploration 5.6B: Kentucky Derby revisited

Chapter 6: Categorical Response Variable

- *Review descriptive and inferential analyses of two-way tables*
- *Consider alternative statistics for comparing two proportions*
- *Computation and interpretation of adjusted odds ratios*
- *Interpretation of logistic regression models*
- Section 6.1: Comparing Proportions
    - Example 6.1: Organ Donation
    - Exploration 6.1: Infant Attachment
- Section 6.2: Adjusted Odds Ratio
    - Example 6.2: Smoking and Survival Rates
    - Exploration 6.2: Preschool Obesity
- Section 6.3: Logistic Regression
    - Example 6.3: Smoking and Survival Rates cont.
    - Exploration 6.3: Norovirus

Chapter 7: Model building and case studies

- *Review basic model building rules discussed so far*
- *Explore techniques and strategies for messy data including handling extreme values and missing data*
- *Review and apply stepwise and non-stepwise model building techniques*
- Section 7.1: Techniques for messy data
- Section 7.2: Model building
- Case studies and projects