INTERMEDIATE STATISTICAL INVESTIGATIONS: A SECOND COURSE

Preliminaries: Multivariable Thinking and Sources of Variation

- Example P.A: Graduate School Admissions at Berkeley
- Exploration P.A: Salary Discrimination
- Example P.B: Predicting Birthweights
- Exploration P.B: Housing Prices in Michigan

Chapter 1: Sources of Variation

- Section 1.1: Sources of Variation in an Experiment
  - Example 1.1: Scents and Consumer Behavior
  - Exploration 1.1: Memorizing Letters
- Section 1.2: Quantifying Sources of Variation
  - Example 1.2: Scents and Consumer Behavior continued
  - Exploration 1.2: Starry Navigation
- Section 1.3: Is the Variation Explained Statistically Significant?
  - Example 1.3: Scents and Consumer Behavior continued
  - Exploration 1.3: Starry Navigation continued
- Section 1.4: Comparing Several Groups
  - Example 1.4: Fish consumption and Omega-3
  - Exploration 1.4: Golden Squirrels
- Section 1.5: Confidence Intervals and Prediction Intervals
  - Example 1.5: Fish consumption and Omega-3 continued
  - Exploration 1.5: Golden Squirrels continued
- Section 1.6: Power and Sample Size
  - Example 1.6: Fish consumption and Omega-3 continued
  - Exploration 1.6: Who is Spending More Time on Parenting?

Chapter 2: Controlling Additional Sources of Variation

- Section 2.1: Matched Pairs
  - Example 2.1: Texts vs. Visual Distractions (Facebook vs. Instagram)
  - Exploration 2.1: Chip melting times
- Section 2.2: Randomized Complete 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 continued
  - Exploration 2.3: Car acceleration

Chapter 3: Multi-factor Studies and Interactions

- Section 3.1: Multi-factor Experiments
  - Example 3.1: Corporate Credibility, Endorser, and Purchase Intent
  - Exploration 3.1: Pig Growth
- Section 3.2: Statistical Interactions
  - Example 3.2: Pistachio Bleaching
  - Exploration 3.2: Optimizing Ads
- Section 3.3: Replication
  - Example 3.3: Optimizing Vitamin C
  - Exploration 3.3: Hurricane Names
- Section 3.4: Interactions in Observational Studies

- Example 3.4: Salary Discrimination continued
- Exploration 3.4: Hopelessness and Exercise

Chapter 4: Including a Quantitative Variable

- Section 4.1: Quantitative Explanatory Variables
  - Example 4.1: Recovering Polyphenols from Grape Seed
  - Exploration 4.1: Fatty Acids and DNA
- Section 4.2: Inference for Simple Linear Regression
  - Example 4.2: Recovering Polyphenols from Grape Seed continued
  - Exploration 4.2: Fatty Acids and DNA continued
- Section 4.3: Quantitative and Categorical Explanatory Variables
  - Example 4.3: Michigan Housing Prices
  - Exploration 4.3: Predicting Height
- Section 4.4: T Quantitative/Categorical Interactions
  - Example 4.3: Michigan Housing Prices continued
  - 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: Experiments with Multiple Quantitative Explanatory Variables
  - Example 5.1: Pistachio Bleaching revisited
  - Exploration 5.1: Biodiesel
- Section 5.2: Observational Studies with Multiple Quantitative Explanatory Variables
  - Example 5.2: Brain Size and IQ
  - Exploration 5.2: SLO Real Estate Data
- Section 5.3: Modeling Nonlinear Associations part I – Polynomial models
  - Example 5.3: Artic Sea Ice
  - Exploration 5.3: Kentucky Derby Winning Times
- Section 5.4: Modeling Nonlinear Associations part II – Transformations
  - Example 5.4: Salary Discrimination continued
  - Exploration 5.4A: Stopping Distances
  - Exploration 5.4B: Kentucky Derby revisited

Chapter 6: Categorical Response Variable

- Section 6.1: Comparing Proportions
  - Example 6.1: Organ Donation
  - Exploration 6.1: Infant Attachment
- Section 6.2: Introduction to Logistic Regression
  - Example 6.2: Smoking and Survival Rates
  - Exploration 6.2: Alcohol Abuse in Ukraine
- Section 6.3: Multiple Logistic Regression Models
  - Example 6.3: Smoking and Survival Rates continued
  - Exploration 6.3: Alcohol Abuse in the Ukraine continued

Chapter 7: Practical Issues

- Section 7.1: Dealing with the Messes Created by Messy Data
  - Example 7.1: Public Health Screening for Omega-3 Index

INTERMEDIATE STATISTICAL INVESTIGATIONS: A SECOND COURSE

Preliminaries: Multivariable Thinking and Sources of Variation

- Chapter Level Learning Outcomes
    - *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*

Chapter 1: Sources of Variation

- Chapter Level Learning Outcome
    - *Build simple statistical models to formally capture and summarize important sources of variation in a variable of interest.*
- Section 1.1 Learning Outcomes
    - *Distinguish experiments and observational studies*
    - *Describe 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*
- Section 1.2 Learning Outcomes
    - *Explain partitioning variation in the response variable into variation explained by the model and unexplained variation*
    - *Measuring and Interpreting percentage of variation explained*
    - *Understanding effect size and practical significance*
- Section 1.3 Learning Outcomes
    - *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*
- Section 1.4 Learning Outcomes
    - *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*
- Section 1.5 Learning Outcomes
    - *Carry out post-hoc analysis after significant F test (pairwise differences)*
    - *Produce and interpret confidence intervals on single means*
    - *Produce and interpret prediction intervals on quantitative variables*
    - *Identify factors that impact widths of confidence and prediction intervals*
- Section 1.6 Learning Outcomes
    - *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*

Chapter 2: Controlling Additional Sources of Variation

- Chapter Level Learning Outcomes

    - *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*
    - Adjust a comparison of interest for another variable in an observational
- Section 2.1 Learning Outcomes
    - *Use pairing to potentially reduce unexplained variation and increase the power of a study*
    - *Explain how to analyze paired data appropriately*
- Section 2.2 Learning Outcomes
    - *Analyze a randomized complete block design using simulation-based methods*
    - *Analyze a randomized complete block design using theory-based methods*
- Section 2.3 Learning Outcomes
    - *Analyze an observational study with two sources of explained variation*
    - *Understand what it means to adjust for another variable in the analysis*
    - *Explore effects of covariation in explanatory variables on the analysis*

Chapter 3: Multi-factor Studies and Interactions

- Chapter Level Learning Outcomes
    - *Design and analyze a multi-factor experiment*
    - *Describe three-way associations using multiple explanatory variables*
    - *Explain the concept of an interaction*
    - *Understand study designs that do and do not allow testing of an interaction*
    - *Understand the differences between testing and interpreting interactions in observational studies versus experiments*
- Section 3.1 Learning Outcomes
    - *Design an experiment with more than one treatment variable*
    - *Compare the benefits of a two-variable study where the levels of both variables are assigned by the researcher*
- Section 3.2: Statistical Interactions
    - *Understand the concept of a statistical interaction*
    - *Interpret an interaction plot*
    - *Calculate interaction effects*
    - *Use simulation and theory-based p-values to assess the significance of an interaction*
- Section 3.3 Learning Outcomes
    - *Explain the benefits and challenges of replication*
    - *Define and describe advantages of a generalized block design*
    - *Define and describe advantages of within-block factorial designs*
- Section 3.4 Learning Outcomes
    - *Interpret interactions with observational data*
    - *Be able to sketch an interaction plot and interpret it*
    - *Run a two-variable ANOVA including an interaction term for an observational study with 2 two-level variables*

Chapter 4: Including a Quantitative Variable

- Chapter Leven Learning Outcomes
    - *Review of simple linear regression*
    - *Use 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 Learning Outcomes
    - *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*
- Section 4.2 Learning Outcomes
    - *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 strength of evidence of a linear association between the quantitative explanatory and response variables*
    - *Evaluate the validity of the theory-based approach using residual plots*
- Section 4.3 Learning Outcomes
    - *Adjust the relationship between two quantitative variables based on a categorical variable*
    - *Create indicator variables in order to include binary categorical variables in the regression model*
    - *Evaluate the validity of the regression model*
- Section 4.4 Learning Outcomes
    - *Include interaction between quantitative and categorical variables in a statistical model*
    - *Interpret the nature of the interaction*
- Section 4.5 Learning Outcomes
    - *Include categorical variables with more than two categories in a linear model*
    - *Interpret an interaction between a quantitative variable and a multi-level categorical variable*

Chapter 5: Multiple Quantitative Explanatory Variables

- Chapter Level Learning Outcomes
    - *Identify and discuss potential issues related to adding multiple quantitative variables to a model including multidimensional graphs, adjusted vs. unadjusted associations, associations among explanatory variables, and interactions between quantitative variables*
    - *Explore benefits of standardizing quantitative variables*
    - *Use of polynomial models and transformations to model nonlinear associations*
- Section 5.1 Learning Outcomes
    - *Consider design issues with quantitative explanatory variables*
    - *Visualize relationships among three or more quantitative variables*
    - *Interpret a "response surface" with quantitative explanatory variables*
    - *Describe interactions between quantitative variables*
- Section 5.2 Learning Outcomes
    - *Visualize adjusted associations when adjusting for a quantitative variable*
    - *Create and interpret added variable plots*
    - *Interpret model coefficients*
    - *Interpret adjusted sums of squares*
    - *Identify potential problems when using explanatory variables that are linearly related*
- Section 5.3 Learning Outcomes
    - *Fit a polynomial model*
    - *Assess when a polynomial model is appropriate*
- Section 5.4 Learning Outcomes
    - *Transform the response variable to meet model conditions*

- *Assess different model transformations*

Chapter 6: Categorical Response Variable

- Chapter Level Learning Outcomes
  - *Review descriptive and inferential analyses of two-way tables*
  - *Consider alternative statistics for comparing two proportions*
  - *Compute and interpret of adjusted odds ratios*
  - *Interpret logistic regression models*
- Section 6.1 Learning Outcomes
  - *Review descriptive and inferential methods for comparing groups with a categorical response variable*
  - *Compare and contrast different statistics for evaluating group differences on a binary response variable*
- Section 6.2 Learning Outcomes
  - *Explain the motivation and need for logistic regression*
  - *Utilize a logistic regression model using categorical or quantitative explanatory variables*
- Section 6.3 Learning Outcomes
  - *Utilize a logistic regression model using multiple categorical and/or quantitative explanatory variables*

Chapter 7: Practical Issues

- Chapter Level Learning Outcomes
  - *Understand basic techniques for evaluating and handling missing data and outliers in statistical analyses*
  - *Understand basic techniques for building robust multivariable statistical models*
- Section 7.1 Learning Outcomes
  - *Understand how missing data and outliers can impact statistical analyses*
  - *Introduce techniques for exploring and handling missing data and outliers*
- Section 7.2 Learning Outcomes
  - *Understand the impact of relationships among explanatory variables in multiple regression models*
  - *Understand best practices for multiple regression model building*