To the Student

Investigation 0: Random babies

CHAPTER 1: ANALYZING STATISTICAL PROCESSES

In this chapter, you will begin to analyze results from statistical studies and focus on the process of statistical inference. In particular, you will learn how to assess evidence against a particular claim about a random process.

Section 1: Analyzing a process probability

Investigation 1.1: Friend or foe – Inference for a proportion
Probability Exploration: Mathematical model
Investigation 1.2: Do names match faces – Bar graph, hypotheses, binomial test (technology)
Investigation 1.3: Heart transplant mortality – Factors affecting p-value
Investigation 1.4: Kissing the right way – Two-sided p-values
Investigation 1.5: Kissing the right way (cont.) – Interval of plausible values
Investigation 1.6: Improved batting averages – Types of error and power
Probability Exploration: Exact Binomial Power Calculations

Section 2: Normal approximations for sample proportions

Investigation 1.7: Reese's pieces – Normal model, empirical rule, CLT
Investigation 1.8: Is ESP real? – Normal probabilities, *z*-score
Investigation 1.9: Halloween treat choices – Test statistic, continuity correction
Investigation 1.10: Kissing the right way (cont.) – *z*-interval, confidence level
Investigation 1.11: Heart transplant mortality (cont.) – Adjusted Wald

Section 3: Sampling from a population

Investigation 1.12: Sampling words – Biased and random sampling
Investigation 1.13: *Literary Digest* – Issues in sampling
Investigation 1.14: Sampling words cont. – Central Limit Theorem for $\hat{p}$
Investigation 1.15: Freshmen voting pattern – Nonsampling errors, hypergeometric
Investigation 1.16: Teen hearing loss – One sample *z*-procedures
Investigation 1.17: Cat households – Practical significance
Investigation 1.18: Female senators – Cautions in inference

Example 1.1: Predicting Elections from Faces
Example 1.2: Cola Discrimination
Example 1.3: Seat Belt Usage

Appendix: Stratified random sampling

CHAPTER 2: COMPARING TWO PROPORTIONS

In this chapter, you will focus on comparing results from two groups on a categorical variable. These groups could be samples from different populations or they could have been deliberately formed during the design of the study. You will again consider different ways to analyze the statistical significance of the difference in the groups, namely simulation, exact methods, and normal approximations to answer whether the observed difference in the groups could have happened "by chance alone." You will also continue to consider issues of statistical confidence and types of errors. A key consideration to keep in mind will be the scope of conclusions that you can draw from the study based on how the data were collected.

Section 1: Comparing two population proportions
- Investigation 2.1: Teen hearing loss (cont.) – Tables, conditional props, bar graphs, *z*-procedures
- Investigation 2.2: Night Lights and Near-sightedness – Association, confounding

Section 2: Types of Studies
- Investigation 2.3: Handwriting and SAT scores – Observational studies, experiments
- Investigation 2.4: Have a nice trip – Random assignment, scope of conclusions
- Investigation 2.5: Botox for back pain – Designing experiments

Section 3: Comparing two treatment probabilities
- Investigation 2.6: Dolphin Therapy – Randomization test
- Investigation 2.7: Is yawning contagious? – Fisher's exact test
- Investigation 2.8: CPR vs. chest compressions – z-procedures

Section 4: Other Statistics
- Investigation 2.9: CPR vs. chest compressions – Relative risk
- Investigation 2.10: Smoking and lung cancer – Types of observational studies, odds ratio
- Investigation 2.11: Sleepy drivers – Odds ratio confidence interval

Example 2.1: Wording of Questions
Example 2.2 Worries about Terrorist Attacks

CHAPTER 3: COMPARISONS WITH QUANTITATIVE VARIABLES

This chapter parallels the previous one in many ways. We will continue to consider studies where the goal is to compare a response variable between two groups. The difference here is that these studies will involve a *quantitative* response variable rather than a *categorical* one. The methods that we employ to analyze these data will therefore be different, but you will find that the basic concepts and principles that you learned in Chapters 1 and 2 still apply. These include the principle of starting with numerical and graphical summaries to explore the data, the concept of statistical significance in determining whether the difference in the distribution of the response variable between the two groups is larger than we would reasonably expect from randomization alone, and the importance of considering how the data were collected in determining the scope of conclusions that can be drawn from the study.

Section 1: Quantitative data
- Investigation 3.1: How faithful is Old Faithful? – Descriptive statistics
- Investigation 3.2: The *Ethan Allen* – Sampling distributions for $\bar{x}$
- Investigation 3.3: Healthy body temperatures – Student's *t*-distribution
- Investigation 3.4: Healthy body temperatures (cont.) – *t*-procedures, prediction intervals

Section 2: Comparing two population means
- Investigation 3.5: Left-handedness and life expectancy – Random sampling, 2-sample *t*-tests
- Investigation 3.6: Left-handedness and life expectancy (cont.) – Application, Power

Section 3: Comparing two treatment means
- Investigation 3.7: Lingering effects of sleep deprivation – Randomization tests
- Investigation 3.8: Lingering effects of sleep deprivation (cont.) – two-sample *t*-tests
- Investigation 3.9: Ice cream serving sizes – two-sample *t*-confidence interval
- Investigation 3.10: Cloud seeding – Strategies for non-normal data

Section 4: Matched Pairs Designs
- Investigation 3.11: Chip melting times – Independent vs. paired design, technology
- Investigation 3.12: Chip melting times (cont.) – Inference (simulation, paired *t*-test)
- Investigation 3.13: Comparison shopping – Application
- Investigation 3.14: Smoke alarms – McNemar's test (paired categorical data)

Example 3.1: Age Discrimination – Exact randomization test
Example 3.2: Speed Limit Changes – Two-sample *t*-test
Example 3.3: Distracted Driving? – Matched pairs
Example 3.4: Distracted Driving (cont.) – Sign test

CHAPTER 4: COMPARING SEVERAL POPULATIONS, EXPLORING RELATIONSHIPS

The idea of comparing two groups has been a recurring theme throughout this course. In the previous chapters, you have been limited to exploring two groups at a time. You saw that often the same analysis techniques apply whether the data have been collected as independent random samples or from a randomized experiment, although this data collection distinction strongly influences the scope of conclusions that you can draw from the study. You will see a similar pattern in this chapter as you extend your analyses to exploring two or more groups. In particular, you will study a procedure for comparing a categorical response variable across several groups and a procedure for comparing a quantitative response variable across several groups. You will also study the important notion of association between variables, first with categorical variables and then for studies in which both variables are quantitative. In this latter case, you will also learn a new set of numerical and graphical summaries for describing these relationships.

Section 1: Two Categorical Variables
- Investigation 4.1: Dr. Spock's Trial – Chi-square test for homogeneity of proportions
- Investigation 4.2: Near-sightedness and night lights (cont.) – Chi-square test for association
- Technology Exploration: Randomization test for chi-square statistic
- Investigation 4.3: Newspaper credibility decline – Comparing distributions

Section 2: Comparing Several Population Means
- Investigation 4.4: Disability discrimination – Reasoning of ANOVA
- Applet Exploration: Randomization test for ANOVA
- Investigation 4.5: Restaurant spending and music – ANOVA practice
- Applet Exploration: Exploring ANOVA

Section 3: Two Quantitative Variables
- Investigation 4.6: Cat jumping – Scatterplots
- Investigation 4.7: Drive for show, putt for dough – Correlation coefficients
- Applet Exploration: Correlation guessing game
- Investigation 4.8: Height and foot size – Least squares regression
- Applet Exploration: Behavior of regression lines – Resistance
- Excel Exploration: Minimization criteria
- Investigation 4.9: Money-making movies – Application

Section 4: Inference for Regression
- Investigation 4.10: Running out of time – Inference for regression (sampling)
- Investigation 4.11: Running out of time (cont.) – Inference for regression (shuffling)
- Investigation 4.12: Boys' heights – Regression model
- Investigation 4.13: Cat jumping (cont.) – Confidence intervals for regression
- Investigation 4.14: Housing prices – Transformations
- Technology Exploration: The regression effect

Example 4.1: Internet Use by Region
Example 4.2: Lifetimes of Notables
Example 4.3: Physical Education Class Performance
Example 4.4: Comparing Population Diets