

Table of Contents

To The Student.....	3
Investigation A: Is My Water Safe?.....	4
Investigation B: Random Babies	10
Chapter 1: Analyzing One Categorical Variable	17
Section 1: Analyzing A Process Probability	18
Section 2: Normal Approximations For Sample Proportions	56
Section 3: Sampling From A Finite Population.....	84
Chapter 2: Analyzing Quantitative Data.....	126
Section 1: Descriptive Statistics.....	127
Section 2: Inference For Population Mean	142
Section 3: Inference For Other Statistics	161
Chapter 3: Comparing Two Proportions.....	177
Section 1: Comparing Two Population Proportions	178
Section 2: Types Of Studies.....	193
Section 3: Comparing Two Treatment Probabilities	201
Section 4: Other Statistics.....	217
Chapter 4: Comparisons With Quantitative Variables.....	246
Section 1: Comparing Groups – Quantitative Reponse	247
Section 2: Comparing Two Population Means	249
Section 3: Comparing Two Treatment Means	265
Section 4: Matched Pairs Designs.....	282
Chapter 5: Comparing Several Populations, Exploring Relationships	315
Section 1: Two Categorical Variables	316
Section 2: Comparing Several Population Means	336
Section 3: Relationships Between Quantitative Variables.....	350
Section 4: Inference For Regression	378
Index	420

CHAPTER 1: ANALYZING ONE CATEGORICAL VARIABLE

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
- Investigation 1.2: Butter-side down again? – Binomial random variables
- Investigation 1.3: Are you clairvoyant? – Standardizing
- Investigation 1.4: Heart transplant mortality – Factors affecting p-value
- Investigation 1.5: Kissing the right way – Two-sided p-values
- Investigation 1.6: Kissing the right way (cont.) – Interval of plausible values
- Investigation 1.7: Improved baseball player – Types of error and power
- Probability Exploration: Exact Binomial Power Calculations

Section 2: Normal approximations for sample proportions

- Investigation 1.8: Reese's pieces – Normal model, Central Limit Theorem
- Probability Detour: Normal Random Variables
- Investigation 1.9: Halloween treat choices – One sample z-test, continuity correction
- Investigation 1.10: Kissing the right way (cont.) – z-interval, confidence level
- Investigation 1.11: Heart transplant mortality (cont.) – Plus Four/Adjusted Wald
- Probability Exploration: Normal power calculations

Section 3: Sampling from a finite 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: Counting Concussions – Nonsampling errors, hypergeometric distribution
- Probability Detour: Hypergeometric Random Variables
- Probability Exploration: Finite population correction
- 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: ANALYZING QUANTITATIVE DATA

This chapter parallels the previous one in many ways. The difference here is that these investigations will involve a *quantitative* variable rather than a *categorical* one. This requires us to learn different tools for graphing and summarizing our data, as well as for statistical inference. In the end, you will find that the basic concepts and principles that you learned in Chapters 1 still apply.

Section 1: Descriptive Statistics

Investigation 2.1: Birth weights – Normal model, Assessing model fit

Investigation 2.2: How long can you stand it? – Skewed data

Investigation 2.3: Cancer pamphlets - Application

Section 2: Inference for Mean

Investigation 2.4: The *Ethan Allen* – Sampling distributions for \bar{x}

Investigation 2.5: Healthy body temperatures – One-sample *t*-procedures

Probability Detour: Student's *t* Distribution

Investigation 2.6: Healthy body temperatures (cont.) – Prediction intervals

Section 3: Inference for Other Statistics (optional)

Investigation 2.7: Water oxygen levels – Sign test

Investigation 2.8: Turbidity – *t*-procedures with transformed data

Investigation 2.9: Heroin treatment times - Bootstrapping

Example 2.1: Pushing On – One-sample *t*-procedures

Example 2.2: Distracted Driving? – Sign test

CHAPTER 3: 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 (a *third* source of possible randomness). You will again consider multiple 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 3.1: Teen hearing loss (cont.) – Tables, conditional props, bar graphs, z-procedures

Investigation 3.2: Nightlights and near-sightedness – Association, confounding

Section 2: Types of Studies

Investigation 3.3: Handwriting and SAT scores – Observational studies, experiments

Investigation 3.4: Botox for back pain – Designing experiments

Section 3: Comparing two treatment probabilities

Investigation 3.5: Dolphin therapy – Randomization test

Investigation 3.6: Is yawning contagious? – Fisher’s exact test

Investigation 3.7: CPR vs. chest compressions – z-procedures

Section 4: Other Statistics

Investigation 3.8: Peanut allergies – Relative risk

Investigation 3.9: Smoking and lung cancer – Types of observational studies, odds ratio

Investigation 3.10: Sleepy drivers – Application

Example 3.1: Wording of Questions

Example 3.2: Worries about Terrorist Attacks

CHAPTER 4: 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 1–3 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 randomness 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: Comparing groups – Quantitative response

Investigation 4.1: Employment discrimination?

Section 2: Comparing two population means

Investigation 4.2: The elephants in the room – Comparing groups

Probability Exploration: Simulation of Difference in Sample Means

Investigation 4.2 cont: The elephants in the room cont. – t procedures

Investigation 4.3: Left-handedness and life expectancy – Factors influencing significance

Applet Exploration: Guess the p-value

Section 3: Comparing for two treatment means

Investigation 4.4: Lingering effects of sleep deprivation – Randomization tests

Investigation 4.5: Lingering effects of sleep deprivation (cont.) – Two-sample t -tests

Investigation 4.6: Ice cream serving sizes – Two-sample t -confidence interval

Investigation 4.7: Cloud seeding – Strategies for non-normal data

Section 4: Matched Pairs Designs

Investigation 4.8: Swimming in Syrup– Independent vs. paired design, technology

Investigation 4.9: Swimming in Syrup (cont.) – Inference (simulation, paired t -test)

Investigation 4.10: Comparison shopping – Application

Investigation 4.11: Smoke alarms – McNemar’s test (paired categorical data)

Example 4.1: Age Discrimination? – Randomization test

Example 4.2: Speed Limit Changes – Two-sample t -procedures

Example 4.3: Distracted Driving? (cont.) – Paired t -procedures

Example 4.4: Comparison Shopping in the Future – Power for paired vs. independent samples

CHAPTER 5: 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 5.1: Dr. Spock's trial – Chi-square test for homogeneity of proportions

Investigation 5.1A: Newspaper credibility decline – Comparing distributions

Investigation 5.2: A moral tale – Randomized experiment

Investigation 5.3: Nightlights and near-sightedness (cont.) – Chi-square test for association

Section 2: Comparing Several Population Means

Investigation 5.4: Disability discrimination – Reasoning of ANOVA

Investigation 5.5: Restaurant spending and music – ANOVA practice

Applet Exploration: Exploring ANOVA

Section 3: Two Quantitative Variables

Investigation 5.6: Cat jumping – Scatterplots

Investigation 5.7: Drive for show, putt for dough – Correlation coefficients

Applet Exploration: Correlation guessing game

Investigation 5.8: Height and foot size – Least squares regression

Applet Exploration: Behavior of regression lines – Resistance

Excel Exploration: Minimization criteria

Investigation 5.9: Money-making movies – Application

Section 4: Inference for Regression

Investigation 5.10: Running out of time – Inference for regression (sampling)

Investigation 5.11: Running out of time (cont.) – Inference for regression (shuffling)

Investigation 5.12: Boys' heights – Regression model

Investigation 5.13: Cat jumping (cont.) – Confidence intervals for regression

Investigation 5.14: Housing prices – Transformations

Technology Exploration: The regression effect

Example 5.1: Internet Use by Region

Example 5.2: Lifetimes of Notables

Example 5.3: Physical Education Class Performance

Example 5.4: Comparing Popular Diets