

Contents

Acknowledgments

1 Introduction

2 Common Pitfalls

2.1 Examples of Common Mistakes

2.2 Defining Your Question

2.3 Working with and Talking to a Statistician

2.4 Exploratory versus Inferential Statistics

2.5 Different Sources of Variation

2.6 The Importance of Checking Assumptions and the Ramifications of Ignoring the Obvious

2.7 Statistical Software Packages

2.8 Installing and Using R and R Commander

2.8.1 Loading Data

2.8.2 Variable Types

2.8.3 Handling Graphics

2.8.4 Saving Your Work

2.8.5 Getting Help

3 Descriptive Statistics

3.1 Definitions

3.2 Numerical Ways to Describe Data

3.2.1 Categorical Data

3.2.2 Quantitative Data

3.2.3 Determining Outliers

3.2.4 How to Choose a Descriptive Measure

3.3 Graphical Methods to Display Data

3.3.1 How to Choose the Appropriate Graphical Display for Your Data

3.4 Probability Distributions

3.4.1 The Binomial Distribution

3.4.2 The Normal Distribution

3.4.3 Assessing Normality in Your Data

3.4.4 Data Transformations

3.5 The Central Limit Theorem

3.5.1 The Central Limit Theorem for Sample Proportions

3.5.2 The Central Limit Theorem for Sample Means

3.6 Standard Deviation versus Standard Error

3.7 Error Bars

3.8 Correlation

3.8.1 Correlation and Causation

4 Design of Experiments

4.1 Mathematical and Statistical Models

4.1.1 Biological Models

4.2 Describing Relationships between Variables

4.3 Choosing a Sample

4.3.1 Problems in Sampling: Bias

4.3.2 Problems in Sampling: Accuracy and Precision

4.4 Choosing a Model

4.5 Sample Size

4.6 Resampling and Replication

5 Confidence Intervals

5.1 Interpretation of Confidence Intervals

5.1.1 Confidence Levels

5.1.2 Precision

5.2 Computing Confidence Intervals

5.2.1 Confidence Intervals for Large Sample Mean

5.2.2 Confidence Interval for Small Sample Mean

5.2.3 Confidence Interval for Population Proportion

5.3 Sample Size Calculations

6 Hypothesis Testing

6.1 The Basic Principle

6.1.1 p-values

6.1.2 Errors in Hypothesis Testing

6.1.3 Power of a Test

6.1.4 Interpreting Statistical Significance

6.2 Common Hypothesis Tests

6.2.1 t-test

6.2.2 z-test

6.2.3 F-test

6.2.4 Tukey's Test and Scheffé's Test

6.2.5 χ^2 -test: Goodness-of-Fit or Test of Independence

6.2.6 Likelihood Ratio Test

6.3 Non-parametric Tests

6.3.1 Wilcoxon-Mann-Whitney Rank Sum Test

6.3.2 Fisher's Exact Test

6.3.3 Permutation Tests

6.4 E-values

7 Regression and ANOVA

7.1 Regression

- 7.1.1 Correlation and Regression**
- 7.1.2 Parameter Estimation**
- 7.1.3 Hypothesis Testing**
- 7.1.4 Logistic Regression**
- 7.1.5 Multiple Linear Regression**
- 7.1.6 Model Building in Regression: Which Variables to Use?**
- 7.1.7 Verification of Assumptions**
- 7.1.8 Outliers in Regression**
- 7.1.9 A Case Study**
- 7.2 ANOVA**
 - 7.2.1 One-Way ANOVA Model**
 - 7.2.2 Two-Way ANOVA Model**
 - 7.2.3 ANOVA Assumptions**
 - 7.2.4 ANOVA Model for Microarray Data**
- 7.3 What ANOVA and Regression Models Have in Common**
- 8 Special Topics**
 - 8.1 Classification**
 - 8.2 Clustering**
 - 8.2.1 Hierarchical Clustering**
 - 8.2.2 Partitional Clustering**
 - 8.3 Principal Component Analysis**
 - 8.4 Microarray Data Analysis**
 - 8.4.1 The Data**
 - 8.4.2 Normalization**
 - 8.4.3 Statistical Analysis**
 - 8.4.4 The ANOVA Model**
 - 8.4.5 Variance Assumptions**
 - 8.4.6 Multiple Testing Issues**
 - 8.5 Next-Generation Sequencing Analysis**
 - 8.5.1 Experimental Overview**
 - 8.5.2 Statistical Issues in Next-Generation Sequencing Experiments**
 - 8.6 Maximum Likelihood**
 - 8.7 Frequentist and Bayesian Statistics**
- References**
- Index**
 - Index of Worked Out Examples**
 - Index of R Commander Commands**