

CONTENTS

To Teachers: About This Book

To Students: What Is Statistics?

PART I: Looking at Data 1

CHAPTER 1

Looking at Data—Distributions 3

Introduction 4

Variables 4

1.1 Displaying Distributions with Graphs 7

Graphs for categorical variables 7

Data analysis in action: don't hang up on me 9

Stemplots 11

Histograms 14

Examining distributions 17

Dealing with outliers 18

Time plots 19

Beyond the basics: decomposing time series 21

Summary 23

Section 1.1 Exercises 25

1.2 Describing Distributions with Numbers 40

Measuring center: the mean 40

Measuring center: the median 42

Mean versus median 43

Measuring spread: the quartiles 44

The five-number summary and boxplots 46

The $1.5 \times IQR$ criterion for suspected outliers 47

Measuring spread: the standard deviation 49

Properties of the standard deviation 51

Choosing measures of center and spread 52

*Sections marked with asterisks are optional.

Changing the unit of measurement	53
Summary	55
Section 1.2 Exercises	56
1.3 Density Curves and Normal Distributions	64
Density curves	65
Measuring center and spread for density curves	68
Normal distributions	69
The 68–95–99.7 rule	71
Standardizing observations	73
Normal distribution calculations	74
Using the standard normal table	76
Inverse normal calculations	78
Normal quantile plots	80
Beyond the basics: density estimation	83
Summary	83
Section 1.3 Exercises	84
Chapter 1 Exercises	94
EESEE Case Studies	99

CHAPTER 2

Looking at Data—Relationships	101
Introduction	102
Examining relationships	102
2.1 Scatterplots	104
Interpreting scatterplots	105
Adding categorical variables to scatterplots	106
More examples of scatterplots	107
Beyond the basics: scatterplot smoothers	110
Categorical explanatory variables	111
Summary	112
Section 2.1 Exercises	112
2.2 Correlation	123
The correlation r	124
Properties of correlation	124
Summary	127

Section 2.2 Exercises	127
2.3 Least-Squares Regression	132
Fitting a line to data	133
Prediction	134
Least-squares regression	135
Interpreting the regression line	138
Correlation and regression	138
Understanding r^{2*}	142
Beyond the basics: transforming relationships	143
Summary	145
Section 2.3 Exercises	145
2.4 Cautions about Regression and Correlation	151
Residuals	151
Outliers and influential observations	154
Beware the lurking variable	158
Beware correlations based on averaged data	161
The restricted range problem	161
Beyond the basics: data mining	162
Summary	163
Section 2.4 Exercises	163
2.5 The Question of Causation	175
Explaining association: causation	174
Explaining association: common response	175
Explaining association: confounding	176
Establishing causation	177
Summary	178
Section 2.5 Exercises	179
Chapter 2 Exercises	181
EESEE Case Studies	188
CHAPTER 3	
Producing Data	191
Introduction	192

3.1 First Steps	192
Where to find data: the library and the Internet	194
Sampling	194
Experiments	195
Summary	196
Section 3.1 Exercises	197
3.2 Design of Experiments	198
Comparative experiments	200
Randomization	201
Randomized comparative experiments	202
How to randomize	203
Cautions about experimentation	206
Matched pairs designs	207
Block designs	208
Summary	209
Section 3.2 Exercises	210
3.3 Sampling Design	218
Simple random samples	219
Stratified samples	220
Multistage samples	221
Cautions about sample surveys	222
Summary	224
Section 3.3 Exercises	225
3.4 Toward Statistical Inference	231
Sampling variability	232
Sampling distributions	233
Bias and variability	236
Sampling from large populations	238
Why randomize?	238
Beyond the basics: capture-recapture sampling	239
Summary	240
Section 3.4 Exercises	240
Chapter 3 Exercises	246
EESEE Case Studies	249

PART II: Probability and Inference **251**

CHAPTER 4

Probability: The Study of Randomness **253**

Introduction **254**

4.1 Randomness **254**

The language of probability	255
Thinking about randomness	256
The uses of probability	256
Summary	257
Section 4.1 Exercises	257

4.2 Probability Models **259**

Sample spaces	259
Intuitive probability	261
Probability rules	262
Assigning probabilities: finite number of outcomes	264
Assigning probabilities: equally likely outcomes	265
Independence and the multiplication rule	266
Applying the probability rules	269
Summary	270
Section 4.2 Exercises	271

4.3 Random Variables **277**

Discrete random variables	278
Continuous random variables	282
Normal distributions as probability distributions	284
Summary	286
Section 4.3 Exercises	286

4.4 Means and Variances of Random Variables **291**

The mean of a random variable	291
Statistical estimation and the law of large numbers	294
Thinking about the law of large numbers	296
Beyond the basics: more laws of large numbers	297
Rules for means	298
The variance of a random variable	300
Rules for variances	301
Summary	304
Section 4.4 Exercises	305

4.5 General Probability Rules*	311
General addition rules	312
Conditional probability	315
General multiplication rules	319
Tree diagrams	320
Bayes's rule	321
Independence again	322
Summary	323
Section 4.5 Exercises	323
Chapter 4 Exercises	328
EESEE Case Studies	331
CHAPTER 5	
Sampling Distributions	333
Introduction	334
5.1 Sampling Distributions for Counts and Proportions	335
The binomial distributions for sample counts	335
Binomial distributions in statistical sampling	336
Finding binomial probabilities: software and tables	337
Binomial mean and standard deviation	340
Sample proportions	341
Normal approximation for counts and proportions	343
The continuity correction*	347
Binomial formulas*	348
Summary	350
Section 5.1 Exercises	351
5.2 The Sampling Distribution of a Sample Mean	358
The mean and standard deviation of \bar{x}	360
The central limit theorem	362
A few more facts	365
Beyond the basics: Weibull distributions	367
Summary	368
Section 5.2 Exercises	369
Chapter 5 Exercises	375

CHAPTER 6

Introduction to Inference	381
Introduction	382
6.1 Estimating with Confidence	383
Statistical confidence	384
Confidence intervals	385
Confidence interval for a population mean	387
How confidence intervals behave	390
Choosing the sample size	391
Some cautions	393
Beyond the basics: the bootstrap	394
Summary	395
Section 6.1 Exercises	396
6.2 Tests of Significance	400
The reasoning of significance tests	400
Stating hypotheses	402
Test statistics	403
<i>P</i> -values	404
Statistical significance	406
Tests for a population mean	409
Two-sided significance tests and confidence intervals	413
<i>P</i> -values versus fixed α	415
Summary	416
Section 6.2 Exercises	416
6.3 Use and Abuse of Tests	424
Choosing a level of significance	424
What statistical significance doesn't mean	425
Don't ignore lack of significance	425
Statistical inference is not valid for all sets of data	426
Beware of searching for significance	427
Summary	427
Section 6.3 Exercises	428

6.4 Power and Inference as a Decision*	430
Power	430
Increasing the power	433
Inference as decision	435
Two types of error	435
Error probabilities	436
The common practice of testing hypotheses	438
Summary	439
Section 6.4 Exercises	439
Chapter 6 Exercises	441
EESEE Case Studies	446

CHAPTER 7

Inference for Distributions	449
Introduction	450
7.1 Inference for the Mean of a Population	450
The t distributions	450
The one-sample t confidence interval	452
The one-sample t test	454
Matched pairs t procedures	459
Robustness of the t procedures	462
The power of the t test*	464
Inference for nonnormal populations*	465
Summary	470
Section 7.1 Exercises	471
7.2 Comparing Two Means	485
The two-sample z statistic	486
The two-sample t procedures	488
The two-sample t significance test	489
The two-sample t confidence interval	492
Robustness of the two-sample procedures	493
Inference for small samples	495
Software approximation for the degrees of freedom*	498
The pooled two-sample t procedures*	499

Summary	503
Section 7.2 Exercises	504
7.3 Optional Topics in Comparing Distributions*	515
Inference for population spread	516
The F test for equality of spread	516
Robustness of normal inference procedures	518
The power of the two-sample t test	519
Summary	521
Section 7.3 Exercises	521
Chapter 7 Exercises	524
EESEE Case Studies	533
CHAPTER 8	
Inference for Proportions	535
Introduction	536
8.1 Inference for a Single Proportion	536
Large-sample confidence interval for a single proportion	536
Plus four confidence interval for a single proportion	539
Significance test for a single proportion	540
Confidence intervals provide additional information	543
Choosing a sample size	545
Summary	548
Section 8.1 Exercises	549
8.2 Comparing Two Proportions	555
Large-sample confidence interval for a difference in proportions	556
Plus four confidence interval for a difference in proportions	558
Significance tests	561
Beyond the basics: relative risk	563
Summary	565
Section 8.2 Exercises	566
Chapter 8 Exercises	570

EESEE Case Studies	578
--------------------	-----

PART III: Topics in Inference	579
--------------------------------------	------------

CHAPTER 9	
------------------	--

Analysis of Two-Way Tables	581
-----------------------------------	------------

Introduction	582
---------------------	------------

9.1 Data Analysis for Two-Way Tables	582
---	------------

The two-way table	582
Marginal distributions	584
Describing relations in two-way tables	585
Conditional distributions	586
Simpson's paradox	588
The perils of aggregation	590
Summary	590

9.2 Inference for Two-Way Tables	591
---	------------

The hypothesis: no association	594
Expected cell counts	594
The chi-square test	595
The chi-square test and the z test	598
Beyond the basics: meta-analysis	598
Summary	600

9.3 Formulas and Models for Two-Way Tables*	600
--	------------

Computations	600
Computing conditional distributions	601
Computing expected cell counts	604
The X^2 statistic and its P -value	605
Models for two-way tables	606
Concluding remarks	608
Summary	608

9.4 Goodness of Fit*	609
-----------------------------	------------

Chapter 9 Exercises	612
----------------------------	------------

EESEE Case Studies	631
CHAPTER 10	
Inference for Regression	633
Introduction	634
10.1 Simple Linear Regression	634
Statistical model for linear regression	634
Data for simple linear regression	635
Estimating the regression parameters	639
Confidence intervals and significance tests	644
Confidence intervals for mean response	646
Prediction intervals	648
Beyond the basics: nonlinear regression	650
Summary	651
10.2 More Detail about Simple Linear Regression*	652
Analysis of variance for regression	653
The ANOVA F test	655
Calculations for regression inference	656
Inference for correlation	664
Summary	666
Chapter 10 Exercises	667
EESEE Case Studies	681
CHAPTER 11	
Multiple Regression	683
Introduction	684
11.1 Inference for Multiple Regression	684
Population multiple regression equation	684
Data for multiple regression	685
Multiple linear regression model	685
Estimation of the multiple regression parameters	685

Confidence intervals and significance tests for regression coefficients	687
ANOVA table for multiple regression	688
Squared multiple correlation R^2	690
11.2 A Case Study	690
Preliminary analysis	690
Relationships between pairs of variables	691
Regression on high school grades	693
Interpretation of results	695
Residuals	695
Refining the model	696
Regression on SAT scores	697
Regression using all variables	698
Test for a collection of regression coefficients	701
Beyond the basics: multiple logistic regression	701
Summary	702
Chapter 11 Exercises	703
EESEE Case Studies	717

CHAPTER 12

One-Way Analysis of Variance	719
Introduction	720
12.1 Inference for One-Way Analysis of Variance	720
Data for a one-way ANOVA	720
Comparing means	721
The two-sample t statistic	723
An overview of ANOVA	724
The ANOVA model	726
Estimates of population parameters	728
Testing hypotheses in one-way ANOVA	730
The ANOVA table	732
The F test	734
12.2 Comparing the Means	737
Contrasts	737
Multiple comparisons	742

Software	746
Power*	748
Summary	752
Chapter 12 Exercises	751
EESEE Case Studies	769
CHAPTER 13	
Two-Way Analysis of Variance	771
Introduction	772
13.1 The Two-Way ANOVA Model	772
Advantages of two-way ANOVA	772
The two-way ANOVA model	776
Main effects and interactions	777
13.2 Inference for Two-Way ANOVA	783
The ANOVA table for two-way ANOVA	783
Summary	787
Chapter 13 Exercises	788
Data Appendix	
Tables	
Solutions to Selected Exercises	
Notes	
Index	

Additional chapters are on CD-ROM and available in separate print supplements.

CHAPTER 14

Bootstrap Methods and Permutation Tests

Introduction

Software

14.1 The Bootstrap Idea

The big idea: resampling and the bootstrap distribution

Thinking about the bootstrap idea

Using software

Section 14.1 Summary

Section 14.1 Exercises

14.2 First Steps in Using the Bootstrap

Bootstrap t confidence intervals

Bootstrapping to compare two groups

Beyond the basics: the bootstrap for a scatterplot smoother

Section 14.2 Summary

Section 14.2 Exercises

14.3 How Accurate Is a Bootstrap Distribution?*

Bootstrapping small samples

Bootstrapping a sample median

Section 14.3 Summary

Section 14.3 Exercises

14.4 Bootstrap Confidence Intervals

Bootstrap percentile confidence intervals

Confidence intervals for the correlation

More accurate bootstrap confidence intervals: BCa and tilting

Section 14.4 Summary

Section 14.4 Exercises

14.5 Significance Testing Using Permutation Tests

Using software

Permutation tests in practice

Permutation tests in other settings

Section 14.5 Summary

Section 14.5 Exercises

Chapter 14 Exercises

Notes

CHAPTER 15

Nonparametric Tests

Introduction

15.1 The Wilcoxon Rank Sum Test

The rank transformation

The Wilcoxon rank sum test

The normal approximation

What hypotheses does Wilcoxon test?

Ties

Rank, t , and permutation tests

Section 15.1 Summary

Section 15.1 Exercises

15.2 The Wilcoxon Signed Rank Test

The normal approximation

Ties

Section 15.2 Summary

Section 15.2 Exercises

15.3 The Kruskal-Wallis Test

Hypotheses and assumptions

The Kruskal-Wallis test

Section 15.3 Summary

Section 15.3 Exercises

Chapter 15 Exercises

Notes

CHAPTER 16

Logistic Regression

Introduction

16.1 The Logistic Regression Model

Binomial distributions and odds	
Model for logistic regression	
Fitting and interpreting the logistic regression model	

16.2 Inference for Logistic Regression

Confidence intervals and significance tests	
Multiple logistic regression	
Summary	

Chapter 16 Exercises

CHAPTER 17

Statistics for Quality: Control and Capability

Introduction

17.1 Processes and Statistical Process Control

Describing processes	
Statistical process control	
\bar{x} charts for process monitoring	
s charts for process monitoring	
Section 17.1 Summary	
Section 17.1 Exercises	

17.2 Using Control Charts

\bar{x} and R charts	
Additional out-of-control signals	
Setting up control charts	
Comments on statistical control	
Don't confuse control with capability!	
Section 17.2 Summary	
Section 17.2 Exercises	

17.3 Process Capability Indexes*

The capability indexes C_p and C_{pk}	
Cautions about capability indexes	
Section 17.3 Summary	
Section 17.3 Exercises	

17.4 Control Charts for Sample Proportions

Control limits for p charts	
Section 17.4 Summary	
Section 17.4 Exercises	

Chapter 17 Exercises

Notes