

13586

INTRODUCTION to the **PRACTICE** of **STATISTICS**

EIGHTH EDITION

MOORE ■ McCABE ■ CRAIG



INTRODUCTION to the **PRACTICE** of **STATISTICS**

EIGHTH EDITION

David S. Moore
George P. McCabe
Bruce A. Craig
Purdue University



W. H. Freeman and Company
A Macmillan Higher Education Company

BRIEF CONTENTS

To Teachers: About This Book	xiii
To Students: What Is Statistics?	xxiii
About the Authors	xxvii
Data Table Index	xxix
Beyond the Basics Index	xxxix

PART I Looking at Data

CHAPTER 1	Looking at Data—Distributions	1
CHAPTER 2	Looking at Data—Relationships	81
CHAPTER 3	Producing Data	167

PART II Probability and Inference

CHAPTER 4	Probability: The Study of Randomness	231
CHAPTER 5	Sampling Distributions	301
CHAPTER 6	Introduction to Inference	351
CHAPTER 7	Inference for Distributions	417
CHAPTER 8	Inference for Proportions	487

PART III Topics in Inference

CHAPTER 9	Analysis of Two-Way Tables	529
CHAPTER 10	Inference for Regression	563
CHAPTER 11	Multiple Regression	611
CHAPTER 12	One-Way Analysis of Variance	643
CHAPTER 13	Two-Way Analysis of Variance	691
Companion Chapters (on the <i>IPS</i> website www.whfreeman.com/ips8e)		
CHAPTER 14	Logistic Regression	14-1
CHAPTER 15	Nonparametric Tests	15-1
CHAPTER 16	Bootstrap Methods and Permutation Tests	16-1
CHAPTER 17	Statistics for Quality: Control and Capability	17-1
	Tables	T-1
	Answers to Odd-Numbered Exercises	A-1
	Notes and Data Sources	N-1
	Photo Credits	C-1
	Index	I-1

CONTENTS

To Teachers: About This Book	xiii
To Students: What Is Statistics?	xxiii
About the Authors	xxvii
Data Table Index	xxix
Beyond the Basics Index	xxxi

PART I Looking at Data

CHAPTER 1

Looking at Data—Distributions 1

Introduction 1

1.1 Data 2

Key characteristics of a data set	4
Section 1.1 Summary	8
Section 1.1 Exercises	8

1.2 Displaying Distributions with Graphs 9

Categorical variables: bar graphs and pie charts	10
Quantitative variables: stemplots	13
Histograms	15
Data analysis in action: Don't hang up on me	18
Examining distributions	20
Dealing with outliers	21
Time plots	23
Section 1.2 Summary	25
Section 1.2 Exercises	25

1.3 Describing Distributions with Numbers 30

Measuring center: the mean	31
Measuring center: the median	33
Mean versus median	34
Measuring spread: the quartiles	35
The five-number summary and boxplots	37
The $1.5 \times IQR$ rule for suspected outliers	39
Measuring spread: the standard deviation	42
Properties of the standard deviation	44
Choosing measures of center and spread	45
Changing the unit of measurement	45
Section 1.3 Summary	47
Section 1.3 Exercises	48

1.4 Density Curves and Normal Distributions 53

Density curves	55
Measuring center and spread for density curves	56
Normal distributions	58
The 68–95–99.7 rule	59
Standardizing observations	61
Normal distribution calculations	63
Using the standard Normal table	65
Inverse Normal calculations	67
Normal quantile plots	68
Beyond the Basics: Density estimation	71
Section 1.4 Summary	72
Section 1.4 Exercises	72
Chapter 1 Exercises	77

CHAPTER 2

Looking at Data—Relationships 81

Introduction 81

2.1 Relationships 81

Examining relationships	83
Section 2.1 Summary	86
Section 2.1 Exercises	86

2.2 Scatterplots 87

Interpreting scatterplots	89
The log transformation	93
Adding categorical variables to scatterplots	94
Beyond the Basics: Scatterplot smoothers	94
Categorical explanatory variables	97
Section 2.2 Summary	99
Section 2.2 Exercises	99

2.3 Correlation 100

The correlation r	100
Properties of correlation	100
Section 2.3 Summary	100
Section 2.3 Exercises	100

2.4 Least-Squares Regression 100

Fitting a Line to Data	100
Prediction	100
Least-squares regression	100
Interpreting the regression line	100

Facts about least-squares regression	118
Correlation and regression	119
Another view of r^2	120
Section 2.4 Summary	121
Section 2.4 Exercises	122

2.5 Cautions about Correlation and Regression

Residuals	126
Outliers and influential observations	130
Beware of the lurking variable	132
Beware of correlations based on averaged data	134
Beware of restricted ranges	134
Beyond the Basics: Data mining	135
Section 2.5 Summary	136
Section 2.5 Exercises	136

2.6 Data Analysis for Two-Way Tables

The two-way table	139
Joint distribution	141
Marginal distributions	142
Describing relations in two-way tables	143
Conditional distributions	144
Simpson's paradox	146
Section 2.6 Summary	148
Section 2.6 Exercises	148

2.7 The Question of Causation

Explaining association	152
Establishing causation	154
Section 2.7 Summary	156
Section 2.7 Exercises	156
Chapter 2 Exercises	157

How to randomize

Cautions about experimentation

Matched pairs designs

Block designs

Section 3.2 Summary

Section 3.2 Exercises

3.3 Sampling Design

Simple random samples	192
Stratified random samples	194
Multistage random samples	196
Cautions about sample surveys	197
Section 3.3 Summary	198
Section 3.3 Exercises	201

3.4 Toward Statistical Inference

Sampling variability	205
Sampling distributions	207
Bias and variability	208
Sampling from large populations	210
Why randomize?	212
Beyond the Basics: Capture-recapture sampling	213
Section 3.4 Summary	214
Section 3.4 Exercises	215

3.5 Ethics

Institutional review boards	217
Informed consent	219
Confidentiality	220
Clinical trials	220
Behavioral and social science experiments	222
Section 3.5 Summary	224
Section 3.5 Exercises	226
Chapter 3 Exercises	226

CHAPTER 3

Producing Data 167

Introduction 167

3.1 Sources of Data 168

Anecdotal data	168
Available data	169
Sample surveys and experiments	171
Section 3.1 Summary	174
Section 3.1 Exercises	174

3.2 Design of Experiments 175

Comparative experiments	178
Randomization	179
Randomized comparative experiments	181

PART II Probability and Inference

CHAPTER 4

Probability: The Study of Randomness 231

Introduction 231

4.1 Randomness 231

The language of probability	233
Thinking about randomness	234
The uses of probability	235
Section 4.1 Summary	235
Section 4.1 Exercises	236

4.2 Probability Models	236	The central limit theorem	307
Sample spaces	237	A few more facts	313
Probability rules	240	Beyond the Basics: Weibull distributions	315
Assigning probabilities: finite number of outcomes	242	Section 5.1 Summary	316
Assigning probabilities: equally likely outcomes	243	Section 5.1 Exercises	316
Independence and the multiplication rule	244		
Applying the probability rules	248	5.2 Sampling Distributions for Counts and Proportions	320
Section 4.2 Summary	248	The binomial distributions for sample counts	322
Section 4.2 Exercises	249	Binomial distributions in statistical sampling	324
4.3 Random Variables	252	Finding binomial probabilities	325
Discrete random variables	253	Binomial mean and standard deviation	328
Continuous random variables	256	Sample proportions	329
Normal distributions as probability distributions	259	Normal approximation for counts and proportions	331
Section 4.3 Summary	260	The continuity correction	335
Section 4.3 Exercises	261	Binomial formula	336
		The Poisson distributions	339
4.4 Means and Variances of Random Variables	263	Section 5.2 Summary	343
The mean of a random variable	264	Section 5.2 Exercises	344
Statistical estimation and the law of large numbers	267	Chapter 5 Exercises	349
Thinking about the law of large numbers	269		
Beyond the Basics: More laws of large numbers	270	CHAPTER 6	
Rules for means	271	Introduction to Inference	351
The variance of a random variable	273	Introduction	351
Rules for variances and standard deviations	275	Overview of inference	352
Section 4.4 Summary	279	6.1 Estimating with Confidence	353
Section 4.4 Exercises	280	Statistical confidence	354
4.5 General Probability Rules	282	Confidence intervals	356
General addition rules	283	Confidence interval for a population mean	358
Conditional probability	286	How confidence intervals behave	362
General multiplication rules	289	Choosing the sample size	363
Tree diagrams	290	Some cautions	365
Bayes's rule	292	Beyond the Basics: The bootstrap	367
Independence again	293	Section 6.1 Summary	368
Section 4.5 Summary	294	Section 6.1 Exercises	368
Section 4.5 Exercises	294	6.2 Tests of Significance	372
Chapter 4 Exercises	297	The reasoning of significance tests	372
		Stating hypotheses	374
CHAPTER 5		Test statistics	375
Sampling Distributions	301	P-values	377
Introduction	301	Statistical significance	378
5.1 The Sampling Distribution of a Sample Mean	303	Tests for a population mean	382
The mean and standard deviation of $\bar{x}$	305	Two-sided significance tests and confidence intervals	386
		The P-value versus a statement of significance	387
		Section 6.2 Summary	390
		Section 6.2 Exercises	390

6.3 Use and Abuse of Tests	394	7.3 Other Topics in Comparing Distributions	473
Choosing a level of significance	395	Inference for population spread	474
What statistical significance does not mean	396	The <i>F</i> test for equality of spread	474
Don't ignore lack of significance	397	Robustness of Normal inference procedures	477
Statistical inference is not valid for all sets of data	398	The power of the two-sample <i>t</i> test	477
Beware of searching for significance	399	Section 7.3 Summary	479
Section 6.3 Summary	400	Section 7.3 Exercises	480
Section 6.3 Exercises	400	Chapter 7 Exercises	481
6.4 Power and Inference as a Decision	402	CHAPTER 8	
Power	402	Inference for Proportions	487
Increasing the power	405	Introduction	487
Inference as decision	406	8.1 Inference for a Single Proportion	488
Two types of error	407	Large-sample confidence interval for a single proportion	489
Error probabilities	408	Beyond the Basics: The plus four confidence interval for a single proportion	493
The common practice of testing hypotheses	410	Significance test for a single proportion	495
Section 6.4 Summary	411	Choosing a sample size	500
Section 6.4 Exercises	411	Section 8.1 Summary	503
Chapter 6 Exercises	413	Section 8.1 Exercises	504
CHAPTER 7		8.2 Comparing Two Proportions	508
Inference for Distributions	417	Large-sample confidence interval for a difference in proportions	509
Introduction	417	Beyond the Basics: Plus four confidence interval for a difference in proportions	514
7.1 Inference for the Mean of a Population	418	Significance test for a difference in proportions	516
The <i>t</i> distributions	418	Beyond the Basics: Relative risk	520
The one-sample <i>t</i> confidence interval	420	Section 8.2 Summary	521
The one-sample <i>t</i> test	422	Section 8.2 Exercises	522
Matched pairs <i>t</i> procedures	429	Chapter 8 Exercises	525
Robustness of the <i>t</i> procedures	432		
The power of the <i>t</i> test	434		
Inference for non-Normal populations	436		
Section 7.1 Summary	440		
Section 7.1 Exercises	441		
7.2 Comparing Two Means	447	PART III Topics in Inference	
The two-sample <i>z</i> statistic	448	CHAPTER 9	
The two-sample <i>t</i> procedures	450	Analysis of Two-Way Tables	529
The two-sample <i>t</i> confidence interval	451	Introduction	529
The two-sample <i>t</i> significance test	454	9.1 Inference for Two-Way Tables	530
Robustness of the two-sample procedures	455	The hypothesis: no association	536
Inference for small samples	457	Expected cell counts	537
Software approximation for the degrees of freedom	460	The chi-square test	537
The pooled two-sample <i>t</i> procedures	461	Computations	540
Section 7.2 Summary	466		
Section 7.2 Exercises	467		

Computing conditional distributions	541	11.2 A Case Study	619
The chi-square test and the z test	544	Preliminary analysis	619
Models for two-way tables	545	Relationships between pairs of variables	621
Beyond the Basics: Meta-analysis	548	Regression on high school grades	623
Section 9.1 Summary	550	Interpretation of results	624
9.2 Goodness of Fit	551	Residuals	625
Section 9.2 Summary	556	Refining the model	625
Chapter 9 Exercises	557	Regression on SAT scores	627
		Regression using all variables	628
		Test for a collection of regression coefficients	630
CHAPTER 10		Beyond the Basics: Multiple logistic regression	631
Inference for Regression	563	Chapter 11 Summary	633
Introduction	563	Chapter 11 Exercises	634
10.1 Simple Linear Regression	564	CHAPTER 12	
Statistical model for linear regression	564	One-Way Analysis of Variance	643
Data for simple linear regression	566	Introduction	643
Estimating the regression parameters	568	12.1 Inference for One-Way Analysis of Variance	644
Confidence intervals and significance tests	574	Data for one-way ANOVA	644
Confidence intervals for mean response	576	Comparing means	645
Prediction intervals	578	The two-sample t statistic	647
Transforming variables	580	An overview of ANOVA	647
Beyond the Basics: Nonlinear regression	582	The ANOVA model	651
Section 10.1 Summary	584	Estimates of population parameters	653
10.2 More Detail about Simple Linear Regression	585	Testing hypotheses in one-way ANOVA	655
Analysis of variance for regression	586	The ANOVA table	657
The ANOVA F test	588	The F test	660
Calculations for regression inference	590	12.2 Comparing the Means	663
Inference for correlation	597	Contrasts	663
Section 10.2 Summary	599	Multiple comparisons	668
Chapter 10 Exercises	600	Software	673
		Power	675
		Chapter 12 Summary	677
		Chapter 12 Exercises	678
CHAPTER 11		CHAPTER 13	
Multiple Regression	611	Two-Way Analysis of Variance	691
Introduction	611	Introduction	691
11.1 Inference for Multiple Regression	612	13.1 The Two-Way ANOVA Model	692
Population multiple regression equation	612	Advantages of two-way ANOVA	692
Data for multiple regression	613	The two-way ANOVA model	696
Multiple linear regression model	614	Main effects and interactions	697
Estimation of the multiple regression parameters	615		
Confidence intervals and significance tests for regression coefficients	616		
ANOVA table for multiple regression	617		
Squared multiple correlation R^2	618		

13.2 Inference for Two-Way ANOVA

The ANOVA table for two-way ANOVA
Chapter 13 Summary
Chapter 13 Exercises

702

702

708

708

15.3 The Kruskal-Wallis Test

Hypotheses and assumptions
The Kruskal-Wallis test
Section 15.3 Summary
Section 15.3 Exercises
Chapter 15 Exercises
Chapter 15 Notes and Data Sources

15-28
15-29
15-29
15-33
15-33
15-35
15-35
15-36

Companion Chapters

(on the IPS website www.whfreeman.com/ips8e)

CHAPTER 14

Logistic Regression

14-1

Introduction

14-1

14.1 The Logistic Regression Model

14-2

Binomial distributions and odds

14-2

Odds for two groups

14-3

Model for logistic regression

14-5

Fitting and interpreting the logistic regression model

14-6

14.2 Inference for Logistic Regression

14-9

Confidence intervals and significance tests

14-10

Multiple logistic regression

14-16

Chapter 14 Summary

14-19

Chapter 14 Exercises

14-20

Chapter 14 Notes and Data Sources

14-27

CHAPTER 15

Nonparametric Tests

15-1

Introduction

15-1

15.1 The Wilcoxon Rank Sum Test

15-3

The rank transformation

15-4

The Wilcoxon rank sum test

15-5

The Normal approximation

15-7

What hypotheses does Wilcoxon test?

15-9

Ties

15-10

Rank, t , and permutation tests

15-13

Section 15.1 Summary

15-15

Section 15.1 Exercises

15-15

15.2 The Wilcoxon Signed Rank Test

15-18

The Normal approximation

15-22

Ties

15-23

Testing a hypothesis about the median of a distribution

15-25

Section 15.2 Summary

15-25

Section 15.2 Exercises

15-25

CHAPTER 16

Bootstrap Methods and Permutation Tests

16-1

Introduction

16-1

Software

16-2

16.1 The Bootstrap Idea

16-3

The big idea: resampling and the bootstrap distribution

16-4

Thinking about the bootstrap idea

16-8

Using software

16-9

Section 16.1 Summary

16-10

Section 16.1 Exercises

16-11

16.2 First Steps in Using the Bootstrap

16-13

Bootstrap t confidence intervals

16-14

Bootstrapping to compare two groups

16-17

Beyond the Basics: The bootstrap for a scatterplot smoother

16-20

Section 16.2 Summary

16-22

Section 16.2 Exercises

16-22

16.3 How Accurate Is a Bootstrap Distribution?

16-24

Bootstrapping small samples

16-27

Bootstrapping a sample median

16-29

Section 16.3 Summary

16-30

Section 16.3 Exercises

16-31

16.4 Bootstrap Confidence Intervals

16-32

Bootstrap percentile confidence intervals

16-32

A more accurate bootstrap confidence interval: BCa

16-34

Confidence intervals for the correlation

16-36

Section 16.4 Summary

16-38

Section 16.4 Exercises

16-38

16.5 Significance Testing Using Permutation Tests

16-42

Using software

16-46

Permutation tests in practice	16-46	Setting up control charts	17-26
Permutation tests in other settings	16-49	Comments on statistical control	17-31
Section 16.5 Summary	16-52	Don't confuse control with capability!	17-34
Section 16.5 Exercises	16-53	Section 17.2 Summary	17-35
Chapter 16 Exercises	16-56	Section 17.2 Exercises	17-36
Chapter 16 Notes and Data Sources	16-58	17.3 Process Capability Indexes	17-41
CHAPTER 17		The capability indexes C_p and C_{pk}	17-43
Statistics for Quality: Control and Capability	17-1	Cautions about capability indexes	17-46
Introduction	17-1	Section 17.3 Summary	17-48
Use of data to assess quality	17-2	Section 17.3 Exercises	17-48
17.1 Processes and Statistical Process Control	17-3	17.4 Control Charts for Sample Proportions	17-52
Describing processes	17-3	Control limits for p charts	17-53
Statistical process control	17-6	Section 17.4 Summary	17-57
$\bar{x}$ charts for process monitoring	17-8	Section 17.4 Exercises	17-57
s charts for process monitoring	17-12	Chapter 17 Exercises	17-58
Section 17.1 Summary	17-17	Chapter 17 Notes and Data Sources	17-60
Section 17.1 Exercises	17-18	Tables	T-1
17.2 Using Control Charts	17-22	Answers to Odd-Numbered Exercises	A-1
$\bar{x}$ and R charts	17-23	Notes and Data Sources	N-1
Additional out-of-control rules	17-24	Photo Credits	C-1
		Index	I-1