

Contents

Introduction	209
Sample Mean	210
3.2.1 Deviations	214
Expected Value	216
ABOUT THE AUTHOR	v
PREFACE	xix
ACKNOWLEDGMENTS	xxv
CHAPTER 1 Introduction to Statistics	1
1.1 Introduction	1
1.2 The Nature of Statistics	2
1.2.1 Data Collection	3
1.2.2 Inferential Statistics and Probability Models	4
1.3 Populations and Samples	5
1.3.1 Stratified Random Sampling	6
1.4 A Brief History of Statistics	7
1.5 Use of R	10
Key Terms	13
The Changing Definition of Statistics	13
Review Problems	14
CHAPTER 2 Describing Data Sets	19
2.1 Introduction	19
2.2 Frequency Tables and Graphs	20
2.2.1 Line Graphs, Bar Graphs, and Frequency Polygons	21
2.2.2 Relative Frequency Graphs	22
2.2.3 Pie Charts	26
2.2.4 Using R	26
Problems	29
2.3 Grouped Data and Histograms	33
2.3.1 Using R to Construct a Histogram	40
Problems	41
2.4 Stem-and-Leaf Plots	44
Problems	48
2.5 Sets of Paired Data	51

2.5.1	Using R to Construct a Scatter Diagram	54
	Problems	54
2.6	Some Historical Comments.....	57
	Key Terms.....	58
	Summary	59
	Review Problems	63

CHAPTER 3	Using Statistics to Summarize Data Sets	67
3.1	Introduction	67
3.2	Sample Mean.....	68
3.2.1	Deviations	73
	Problems	74
3.3	Sample Median	77
	Problems	81
3.3.1	Sample Percentiles	84
	Problems	88
3.4	Sample Mode.....	90
	Problems	93
3.5	Sample Variance and Sample Standard Deviation.....	94
	Problems	99
3.6	Normal Data Sets and the Empirical Rule.....	103
	Problems	108
3.7	Sample Correlation Coefficient	112
	Problems	119
3.8	The Lorenz Curve and Gini Index.....	124
3.8.1	The 80–20 and the Pareto Rules	131
	Problems	134
	Key Terms.....	134
	Summary	136
	Review Problems	138

CHAPTER 4	Probability.....	143
4.1	Introduction	143
4.2	Sample Space and Events of an Experiment	143
	Problems	147
4.3	Properties of Probability.....	150
	Problems	154
4.4	Experiments Having Equally Likely Outcomes	158
	Problems	162
4.5	Conditional Probability and Independence	164
	Problems	175
4.6	Bayes' Theorem	182

	Problems	186
4.7	Counting Principles	188
	Problems	194
	Key Terms	197
	Summary	199
	Review Problems	200
CHAPTER 5	Discrete Random Variables.....	209
5.1	Introduction	209
5.2	Random Variables.....	210
	Problems	214
5.3	Expected Value	216
5.3.1	Properties of Expected Values	219
	Problems	224
5.4	Variance of Random Variables.....	229
5.4.1	Properties of Variances.....	231
5.4.2	Expectation of a Function of a Random Variable	234
	Problems	237
5.5	Jointly Distributed Random Variables	240
5.5.1	Covariance and Correlation	241
	Problems	246
5.6	Binomial Random Variables	247
5.6.1	Expected Value and Variance of a Binomial Random Variable	252
	Problems	253
5.7	Hypergeometric Random Variables.....	256
	Problems	258
5.8	Poisson Random Variables	259
	Problems	262
	Key Terms.....	263
	Summary	263
	Review Problems	265
CHAPTER 6	Normal Random Variables.....	271
6.1	Introduction	271
6.2	Continuous Random Variables	272
	Problems	273
6.3	Normal Random Variables	275
	Problems	278
6.4	Probabilities Associated With a Standard Normal Random Variable	280
	Problems	282

6.5	Finding Normal Probabilities: Conversion to the Standard Normal	283
6.6	Additive Property of Normal Random Variables	286
	Problems	287
6.7	Quantiles and Percentiles of Normal Random Variables	290
	Problems	293
	Key Terms.....	294
	Summary	295
	Review Problems	296
CHAPTER 7	Distributions of Sampling Statistics	301
7.1	A Preview	301
7.2	Introduction	302
7.3	Sample Mean.....	303
	Problems	307
7.4	Central Limit Theorem	308
7.4.1	Distribution of the Sample Mean	310
7.4.2	How Large a Sample Is Needed?	313
	Problems	314
7.5	Sampling Proportions From a Finite Population	316
7.5.1	Probabilities Associated With Sample Proportions: The Normal Approximation to the Binomial Distribution	319
	Problems	322
7.6	Distribution of the Sample Variance of a Normal Population	326
	Problems	328
	Key Terms.....	329
	Summary	329
	Review Problems	330
CHAPTER 8	Estimation.....	335
8.1	Introduction	335
8.2	Point Estimator of a Population Mean	336
	Problems	338
8.3	Point Estimator of a Population Proportion.....	339
	Problems	341
8.3.1	Estimating the Probability of a Sensitive Event.....	344
	Problems	345
8.4	Estimating a Population Variance.....	345
	Problems	347
8.5	Interval Estimators of the Mean of a Normal Population With Known Population Variance.....	350

8.5.1	Lower and Upper Confidence Bounds.....	357
	Problems	359
8.6	Interval Estimators of the Mean of a Normal Population With Unknown Population Variance.....	362
8.6.1	Lower and Upper Confidence Bounds.....	366
	Problems	367
8.7	Interval Estimators of a Population Proportion	371
8.7.1	Length of the Confidence Interval.....	373
8.7.2	Lower and Upper Confidence Bounds.....	375
	Problems	377
	Key Terms.....	381
	Summary	381
	Review Problems	384
CHAPTER 9	Testing Statistical Hypotheses	389
9.1	Introduction	389
9.2	Hypothesis Tests and Significance Levels	390
	Problems	393
9.3	Tests Concerning the Mean of a Normal Population: Case of Known Variance	395
	Problems	401
9.3.1	One-Sided Tests	404
	Problems	407
9.4	The t Test for the Mean of a Normal Population: Case of Unknown Variance	410
9.4.1	Use of R in Running a One Sample t-Test	417
	Problems	418
9.5	Hypothesis Tests Concerning Population Proportions	422
9.5.1	Two-Sided Tests of p	426
	Problems	429
	Key Terms.....	433
	Summary	434
	Review Problems and Proposed Case Studies	438
CHAPTER 10	Hypothesis Tests Concerning Two Populations	443
10.1	Introduction	443
10.2	Testing Equality of Means of Two Normal Populations: Case of Known Variances	445
	Problems	449
10.3	Testing Equality of Means: Unknown Variances and Large Sample Sizes	452
	Problems	458

10.4	Testing Equality of Means: Small-Sample Tests When the Unknown Population Variances Are Equal	461
10.4.1	Use of R in Running a Two Sample t -Test	464
	Problems	466
10.5	Paired-Sample t Test	470
	Problems	472
10.6	Testing Equality of Population Proportions	477
	Problems	487
	Key Terms	491
	Summary	491
	Review Problems	495
CHAPTER 11	Analysis of Variance	501
11.1	Introduction	501
11.2	One-Factor Analysis of Variance	503
	A Remark on the Degrees of Freedom	505
	Problems	506
11.3	Two-Factor Analysis of Variance: Introduction and Parameter Estimation	510
11.3.1	Using Matrices in R	512
	Problems	516
11.4	Two-Factor Analysis of Variance: Testing Hypotheses	518
	Problems	524
11.5	Final Comments	526
	Key Terms	527
	Summary	527
	Review Problems	530
CHAPTER 12	Linear Regression	535
12.1	Introduction	536
12.2	Simple Linear Regression Model	537
	Problems	539
12.3	Estimating the Regression Parameters	541
	Problems	545
12.4	Error Random Variable	550
	Problems	552
12.5	Testing the Hypothesis That $\beta = 0$	553
	Problems	556
12.6	Regression to the Mean	560
12.6.1	Why Biological Data Sets Are Often Normally Distributed	564
	Problems	566

12.7	Prediction Intervals for Future Responses	568
	Problems	569
12.8	Coefficient of Determination	572
	Problems	574
12.9	Sample Correlation Coefficient	576
	Problems	577
12.10	Analysis of Residuals: Assessing the Model.....	577
	Problems	578
12.11	Multiple Linear Regression Model.....	580
	12.11.1 Dummy Variables for Categorical Data	584
	Problems	586
12.12	Logistic Regression	588
	Key Terms.....	589
	Summary.....	591
	Review Problems.....	594

CHAPTER 13 Chi-Squared Goodness-of-Fit Tests 601

13.1	Introduction	601
13.2	Chi-Squared Goodness-of-Fit Tests	604
	Problems	609
13.3	Testing for Independence in Populations Classified According to Two Characteristics	615
	Problems	620
13.4	Testing for Independence in Contingency Tables With Fixed Marginal Totals	624
	Problems	627
	Key Terms.....	630
	Summary.....	630
	Review Problems.....	632

CHAPTER 14 Nonparametric Hypotheses Tests..... 639

14.1	Introduction	639
14.2	Sign Test	640
	14.2.1 Testing the Equality of Population Distributions When Samples Are Paired	642
	14.2.2 One-Sided Tests	644
	Problems	646
14.3	Signed-Rank Test	648
	Problems	652
14.4	Rank-Sum Test for Comparing Two Populations.....	656
	14.4.1 Comparing Nonparametric Tests With Tests That Assume Normal Distributions	662

14.5	Runs Test for Randomness.....	665
	Problems	669
14.6	Testing the Equality of Multiple Probability Distributions.....	671
14.6.1	When the Data Are a Set of Comparison Rankings	673
	Problems	676
14.7	Permutation Tests	677
14.7.1	The Two Sample Permutation Test	680
	Problems	684
	Key Terms.....	684
	Summary	685
	Review Problems	688
CHAPTER 15	Quality Control	691
15.1	Introduction	691
15.2	The $\bar{X}$ Control Chart for Detecting a Shift in the Mean.....	692
	Problems	697
15.2.1	When the Mean and Variance Are Unknown	698
15.2.2	S Control Charts	701
	Problems	703
15.3	Control Charts for Fraction Defective.....	706
	Problems	708
15.4	Exponentially Weighted Moving-Average Control Charts	708
	Problems	712
15.5	Cumulative-Sum Control Charts	713
	Problems	716
	Key Terms.....	716
	Summary	716
	Review Problems	717
CHAPTER 16	Machine Learning and Big Data.....	719
16.1	Introduction	719
16.2	Late Flight Probabilities	720
16.3	The Naive Bayes Approach	721
16.3.1	A Variation of Naive Bayes	724
	Problems	727
16.4	Distance Based Estimators the k -Nearest Neighbors Rule .	728
16.4.1	A Distance Weighted Method	729
	Problems	731
16.5	Assessing the Approaches.....	731
16.5.1	Combining Estimators	732
16.5.2	Cross Validation	733

Problems	733
16.6 Choosing the Best Probability: A Bandit Problem	733
Problems	736
APPENDIX A A Data Set	737
APPENDIX B Mathematical Preliminaries	741
B.1 Summation	741
B.2 Absolute Value	741
B.3 Set Notation	742
ANSWERS TO ODD-NUMBERED PROBLEMS	743
INDEX	787

Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.

Instructor Site:

H. G. Wells (1864-1946)

<https://educate.elsevier.com/9780443438905>

In today's complicated world, very few issues are clear-cut and without controversy. In order to understand and form an opinion about an issue, one must usually gather information, or data. To learn from data, one must know something about statistics, which is the art of learning from data.

This introductory statistics text is written for college-level students in any field of study. It can be used in a quarter, semester, or full-year course. Its only prerequisite is high school algebra. Our goal in writing it is to present statistical concepts and techniques in a manner that will teach students not only how and when to utilize the statistical procedures developed, but also to understand why these procedures should be used. As a result we have made a great effort to explain the ideas behind the statistical concepts and techniques presented. Concepts are motivated, illustrated, and explained in a way that attempts to increase one's intuition. It is only when a student develops a feel or intuition for statistics that she or he is really on the path toward making sense of data.

To illustrate the diverse applications of statistics and to offer students different perspectives about the use of statistics, we have provided a wide variety of real examples and problems to be worked by students. Most refer to real-world issues, such as gun control, stock price models, health issues, driving age limits, school admission ages, public policy issues, gender issues, use of helmets, sports, disputed authorship, scientific fraud, and Vitamin C, among many others. Many of them use data that not only are real but are themselves of interest. The examples have been posed in a clear and concise manner and include many thought-provoking problems that emphasize thinking and problem-solving skills. In addition, some of the problems are designed to be open-ended and can be used as starting points for term projects.