

DAVID FREEDMAN
ROBERT PISANI
ROGER PURVES

Statistics

Fourth Edition

W • W • NORTON & COMPANY

NEW YORK • LONDON



Contents

Preface **xv**

PART I. DESIGN OF EXPERIMENTS

Chapter 1. Controlled Experiments **3**

1. The Salk Vaccine Field Trial 3
2. The Portacaval Shunt 7
3. Historical Controls 8
4. Summary 10

Chapter 2. Observational Studies **12**

1. Introduction 12
2. The Clofibrate Trial 13
3. More Examples 15
4. Sex Bias in Graduate Admissions 17
5. Confounding 20
6. Review Exercises 24
7. Summary and Overview 27

PART II. DESCRIPTIVE STATISTICS

Chapter 3. The Histogram **31**

1. Introduction 31
2. Drawing a Histogram 35
3. The Density Scale 38
4. Variables 42
5. Controlling for a Variable 45
6. Cross-Tabulation 47
7. Selective Breeding 48
8. Review Exercises 50
9. Summary 56

Chapter 4. The Average and the Standard Deviation **57**

1. Introduction 57
2. The Average 58
3. The Average and the Histogram 61
4. The Root-Mean-Square 66
5. The Standard Deviation 67
6. Computing the Standard Deviation 71
7. Using a Statistical Calculator 74
8. Review Exercises 74
9. Summary 76

Chapter 5. The Normal Approximation for Data	78
1. The Normal Curve	78
2. Finding Areas under the Normal Curve	82
3. The Normal Approximation for Data	85
4. Percentiles	88
5. Percentiles and the Normal Curve	90
6. Change of Scale	92
7. Review Exercises	93
8. Summary	96
Chapter 6. Measurement Error	97
1. Introduction	97
2. Chance Error	97
3. Outliers	102
4. Bias	103
5. Review Exercises	104
6. Special Review Exercises	105
7. Summary and Overview	108
Chapter 7. Plotting Points and Lines	110
1. Reading Points off a Graph	110
2. Plotting Points	112
3. Slope and Intercept	113
4. Plotting Lines	114
5. The Algebraic Equation for a Line	115
 PART III. CORRELATION AND REGRESSION 	
Chapter 8. Correlation	119
1. The Scatter Diagram	119
2. The Correlation Coefficient	125
3. The SD Line	130
4. Computing the Correlation Coefficient	132
5. Review Exercises	134
6. Summary	139
Chapter 9. More about Correlation	141
1. Features of the Correlation Coefficient	141
2. Changing SDs	144
3. Some Exceptional Cases	147
4. Ecological Correlations	148
5. Association is Not Causation	150
6. Review Exercises	153
7. Summary	157
Chapter 10. Regression	158
1. Introduction	158
2. The Graph of Averages	162

3. The Regression Method for Individuals	165	
4. The Regression Fallacy	169	
5. There Are Two Regression Lines	174	
6. Review Exercises	176	
7. Summary	178	
Chapter 11. The R.M.S. Error for Regression		180
1. Introduction	180	
2. Computing the R.M.S. Error	185	
3. Plotting the Residuals	187	
4. Looking at Vertical Strips	190	
5. Using the Normal Curve Inside a Vertical Strip	195	
6. Review Exercises	198	
7. Summary	201	
Chapter 12. The Regression Line		202
1. Slope and Intercept	202	
2. The Method of Least Squares	208	
3. Does the Regression Make Sense?	211	
4. Review Exercises	213	
5. Summary and Overview	216	
PART IV. PROBABILITY		
Chapter 13. What Are the Chances?		221
1. Introduction	221	
2. Conditional Probabilities	226	
3. The Multiplication Rule	228	
4. Independence	230	
5. The Collins Case	233	
6. Review Exercises	234	
7. Summary	236	
Chapter 14. More about Chance		237
1. Listing the Ways	237	
2. The Addition Rule	241	
3. Two FAQs (Frequently Asked Questions)	243	
4. The Paradox of the Chevalier De Méré	248	
5. Are Real Dice Fair?	252	
6. Review Exercises	252	
7. Summary	254	
Chapter 15. The Binomial Formula		255
1. Introduction	255	
2. The Binomial Formula	259	
3. Review Exercises	261	
4. Special Review Exercises	263	
5. Summary and Overview	268	

PART V. CHANCE VARIABILITY

Chapter 16. The Law of Averages	273
1. What Does the Law of Averages Say?	273
2. Chance Processes	278
3. The Sum of Draws	279
4. Making a Box Model	281
5. Review Exercises	285
6. Summary	287
Chapter 17. The Expected Value and Standard Error	288
1. The Expected Value	288
2. The Standard Error	290
3. Using the Normal Curve	294
4. A Short-Cut	298
5. Classifying and Counting	299
6. Review Exercises	304
7. Postscript	307
8. Summary	307
Chapter 18. The Normal Approximation for Probability Histograms	308
1. Introduction	308
2. Probability Histograms	310
3. Probability Histograms and the Normal Curve	315
4. The Normal Approximation	317
5. The Scope of the Normal Approximation	319
6. Conclusion	325
7. Review Exercises	327
8. Summary	329

PART VI. SAMPLING

Chapter 19. Sample Surveys	333
1. Introduction	333
2. The <i>Literary Digest</i> Poll	334
3. The Year the Polls Elected Dewey	337
4. Using Chance in Survey Work	339
5. How Well Do Probability Methods Work?	342
6. A Closer Look at the Gallup Poll	343
7. Telephone Surveys	346
8. Chance Error and Bias	348
9. Review Exercises	351
10. Summary	353
Chapter 20. Chance Errors in Sampling	355
1. Introduction	355
2. The Expected Value and Standard Error	359
3. Using the Normal Curve	362

4. The Correction Factor	367	
5. The Gallup Poll	370	
6. Review Exercises	371	
7. Summary	373	
Chapter 21. The Accuracy of Percentages		375
1. Introduction	375	
2. Confidence Intervals	381	
3. Interpreting a Confidence Interval	383	
4. <i>Caveat Emptor</i>	387	
5. The Gallup Poll	389	
6. Review Exercises	391	
7. Summary	394	
Chapter 22. Measuring Employment and Unemployment		395
1. Introduction	395	
2. The Design of the Current Population Survey	396	
3. Carrying out the Survey	398	
4. Weighting the Sample	401	
5. Standard Errors	402	
6. The Quality of the Data	404	
7. Bias	404	
8. Review Exercises	405	
9. Summary	407	
Chapter 23. The Accuracy of Averages		409
1. Introduction	409	
2. The Sample Average	415	
3. Which SE?	422	
4. A Reminder	424	
5. Review Exercises	425	
6. Special Review Exercises	428	
7. Summary and Overview	436	
PART VII. CHANCE MODELS		
Chapter 24. A Model for Measurement Error		441
1. Estimating the Accuracy of an Average	441	
2. Chance Models	445	
3. The Gauss Model	450	
4. Conclusion	454	
5. Review Exercises	455	
6. Summary	457	
Chapter 25. Chance Models in Genetics		458
1. How Mendel Discovered Genes	458	
2. Did Mendel's Facts Fit His Model?	463	
3. The Law of Regression	465	

4. An Appreciation of the Model	468
5. Review Exercises	470
6. Summary and Overview	471

PART VIII. TESTS OF SIGNIFICANCE

Chapter 26. Tests of Significance	475
1. Introduction	475
2. The Null and the Alternative	477
3. Test Statistics and Significance Levels	478
4. Making a Test of Significance	482
5. Zero-One Boxes	483
6. The <i>t</i> -Test	488
7. Review Exercises	495
8. Summary	500
Chapter 27. More Tests for Averages	501
1. The Standard Error for a Difference	501
2. Comparing Two Sample Averages	503
3. Experiments	508
4. More on Experiments	512
5. When Does the <i>z</i> -Test Apply?	517
6. Review Exercises	518
7. Summary	521
Chapter 28. The Chi-Square Test	523
1. Introduction	523
2. The Structure of the χ^2 -Test	530
3. How Fisher Used the χ^2 -Test	533
4. Testing Independence	535
5. Review Exercises	540
6. Summary	544
Chapter 29. A Closer Look at Tests of Significance	545
1. Was the Result Significant?	545
2. Data Snooping	547
3. Was the Result Important?	552
4. The Role of the Model	555
5. Does the Difference Prove the Point?	560
6. Conclusion	562
7. Review Exercises	563
8. Special Review Exercises	565
9. Summary and Overview	576

Notes	A3
Answers to Exercises	A43
Tables	A104
Index	A107