

OpenIntro Statistics

Third Edition

David M Diez
Quantitative Analyst
david@openintro.org

Christopher D Barr
Graduate Student
Yale School of Management
chris@openintro.org

Mine Çetinkaya-Rundel
Assistant Professor of the Practice
Department of Statistics
Duke University
mine@openintro.org

Contents

1	Introduction to data	7
1.1	Case study: using stents to prevent strokes	7
1.2	Data basics	9
1.3	Overview of data collection principles	15
1.4	Observational studies and sampling strategies	19
1.5	Experiments	24
1.6	Examining numerical data	26
1.7	Considering categorical data	43
1.8	Case study: gender discrimination (special topic)	50
1.9	Exercises	55
2	Probability (special topic)	76
2.1	Defining probability (special topic)	76
2.2	Conditional probability (special topic)	88
2.3	Sampling from a small population (special topic)	102
2.4	Random variables (special topic)	104
2.5	Continuous distributions (special topic)	113
2.6	Exercises	116
3	Distributions of random variables	127
3.1	Normal distribution	127
3.2	Evaluating the normal approximation	137
3.3	Geometric distribution (special topic)	141
3.4	Binomial distribution (special topic)	145
3.5	More discrete distributions (special topic)	152
3.6	Exercises	158
4	Foundations for inference	168
4.1	Variability in estimates	169
4.2	Confidence intervals	174
4.3	Hypothesis testing	180
4.4	Examining the Central Limit Theorem	194
4.5	Inference for other estimators	197
4.6	Exercises	203

5	Inference for numerical data	219
5.1	One-sample means with the t -distribution	219
5.2	Paired data	228
5.3	Difference of two means	230
5.4	Power calculations for a difference of means (special topic)	239
5.5	Comparing many means with ANOVA (special topic)	246
5.6	Exercises	257
6	Inference for categorical data	274
6.1	Inference for a single proportion	274
6.2	Difference of two proportions	280
6.3	Testing for goodness of fit using chi-square (special topic)	286
6.4	Testing for independence in two-way tables (sp. topic)	297
6.5	Small sample hypothesis testing for a proportion (special topic)	302
6.6	Randomization test (special topic)	307
6.7	Exercises	312
7	Introduction to linear regression	331
7.1	Line fitting, residuals, and correlation	333
7.2	Fitting a line by least squares regression	340
7.3	Types of outliers in linear regression	349
7.4	Inference for linear regression	351
7.5	Exercises	356
8	Multiple and logistic regression	372
8.1	Introduction to multiple regression	372
8.2	Model selection	378
8.3	Checking model assumptions using graphs	382
8.4	Introduction to logistic regression	386
8.5	Exercises	395
A	End of chapter exercise solutions	405
B	Distribution tables	427
B.1	Normal Probability Table	427
B.2	t -Probability Table	430
B.3	Chi-Square Probability Table	432