
STATISTICS

PRINCIPLES AND METHODS

Revised Printing

Richard Johnson and Gouri Bhattacharyya
University of Wisconsin at Madison



JOHN WILEY & SONS

New York Chichester Brisbane Toronto Singapore

CONTENTS

1 INTRODUCTION	01
1 What Is Statistics?	1
2 Statistics in Our Everyday Life	1
3 Statistics in Aid of Scientific Inquiry	4
4 Two Basic Concepts—Population and Sample	7
5 Objectives of Statistics	8
2 ORGANIZATION AND DESCRIPTION OF DATA	10
1 Introduction	12
2 Main Types of Data	12
3 Describing Data by Tables and Graphs	14
4 Measures of Center	25
5 Measures of Variation	32
6 Concluding Remarks	41
7 Exercises	43
3 DESCRIPTIVE STUDY OF BIVARIATE DATA	55
1 Introduction	56
2 Summarization of Bivariate Categorical Data	57
3 Scatter Plot of Bivariate Measurement Data	60
4 The Correlation Coefficient—A Measure of Linear Relation	62
5 Prediction of One Variable from Another (Linear Regression)	70
6 Exercises	75
4 PROBABILITY	83
1 Introduction	85
2 Probability of an Event	86
3 Methods of Assigning Probability	92
4 Event Relations and Two Laws of Probability	101
5 Conditional Probability and Independence	108
6 Random Sampling from a Finite Population	116
7 Exercises	123

5	PROBABILITY DISTRIBUTIONS	131
	1 Introduction	132
	2 Random Variables	133
	3 Probability Distribution of a Discrete Random Variable	136
	4 Expectation (Mean) and Standard Deviation of a Probability Distribution	143
	5 Exercises	152
6	THE BINOMIAL DISTRIBUTION AND ITS APPLICATION IN TESTING HYPOTHESES	157
	1 Introduction	158
	2 Successes and Failures—Bernoulli Trials	159
	3 The Binomial Distribution	163
	4 Testing Hypotheses about a Population Proportion	171
	5 Exercises	186
7	THE NORMAL DISTRIBUTION	192
	1 Probability Model for a Continuous Random Variable	194
	2 The Normal Distribution—Its General Features	201
	3 Standard Normal Distribution	204
	4 Probability Calculations with Normal Distributions	209
	5 The Normal Approximation to the Binomial	213
	*6 Checking the Plausibility of a Normal Model	218
	*7 Transforming Observations to Attain Near Normality	220
	8 Exercises	224
8	VARIATION IN REPEATED SAMPLES—SAMPLING DISTRIBUTIONS	229
	1 Introduction	230
	2 The Sampling Distribution of a Statistic	232
	3 Distribution of the Sample Mean and the Central Limit Theorem	235
	4 Exercises	245
9	DRAWING INFERENCES FROM LARGE SAMPLES	249
	1 Introduction	250
	2 Point Estimation of a Population Mean	252
	3 Confidence Interval for μ	256
	4 Testing Hypotheses about μ	263
	5 Inferences about a Population Proportion	273

6	Deciding on the Sample Size	278
7	Exercises	283
10	SMALL-SAMPLE INFERENCES FOR NORMAL POPULATIONS	289
1	Introduction	290
2	Student's t Distribution	291
3	Confidence Interval for μ —Small Sample Size	295
4	Hypotheses Tests for μ	299
5	Relationship between Tests and Confidence Intervals	301
6	Inferences about the Standard Deviation σ (The Chi-Square Distribution)	304
7	Robustness of Inference Procedures	309
8	Exercises	311
11	COMPARING TWO TREATMENTS	319
1	Introduction	320
2	Independent Random Samples from Two Populations	324
3	Randomization and Its Role in Inference	339
4	Matched Pair Comparisons	342
5	Choosing between Independent Samples and a Matched Pair Sample	349
6	Exercises	353
12	REGRESSION ANALYSIS—I (Simple Linear Regression)	363
1	Introduction	364
2	Regression with a Single Predictor	366
3	A Straight-Line Regression Model	369
4	The Method of Least Squares	371
5	The Sampling Variability of the Least Squares Estimators— Tools for Inference	378
6	Important Inference Problems	379
7	The Strength of a Linear Relation	389
8	Remarks about the Straight-Line Model	393
9	Exercises	396
13	REGRESSION ANALYSIS—II (Multiple Linear Regression and Other Topics)	403
1	Introduction	404
2	Nonlinear Relations and Linearizing Transformations	405

3	Multiple Linear Regression	409
4	Residual Plots to Check the Adequacy of a Statistical Model	418
5	Exercises	422
14	ANALYSIS OF CATEGORICAL DATA	428
1	Introduction	430
2	Pearson's χ^2 Test for Goodness of Fit	433
3	Contingency Tables with One Margin Fixed (Test of Homogeneity)	436
4	Contingency Tables with Neither Margin Fixed (Test of Independence)	447
5	Exercises	456
15	ANALYSIS OF VARIANCE (ANOVA)	465
1	Introduction	466
2	Comparison of Several Treatments: The Completely Randomized Design	467
3	Population Model and Inferences for a Completely Randomized Design	475
4	Simultaneous Confidence Intervals	479
5	Graphical Diagnostics and Displays to Supplement ANOVA	484
6	Exercises	487
16	NONPARAMETRIC INFERENCE	491
1	Introduction	492
2	The Wilcoxon Rank-Sum Test for Comparing Two Treatments	492
3	Matched Pair Comparisons	504
4	Measure of Correlation Based on Ranks	514
5	Concluding Remarks	518
6	Exercises	520
APPENDIX A1	SUMMATION NOTATION	523
APPENDIX A2	EXPECTATION AND STANDARD DEVIATION—PROPERTIES	529
APPENDIX A3	THE EXPECTED VALUE AND STANDARD DEVIATION OF \bar{X}	535
APPENDIX B	TABLES	537
	Table 1 The Number of Combinations $\binom{n}{r}$	537
	Table 2 Cumulative Binomial Probabilities	538
	Table 3 Standard Normal Probabilities	547
	Table 4 Percentage Points of t Distributions	549

Table 5	Percentage Points of χ^2 Distributions	550
Table 6	Percentage Points of $F(\nu_1, \nu_2)$ Distributions	551
Table 7	Selected Tail Probabilities for the Null Distribution of Wilcoxon's Rank-Sum Statistic	553
Table 8	Selected Tail Probabilities for the Null Distribution of Wilcoxon's Signed-Rank Statistic	558

ANSWERS TO SELECTED ODD-NUMBERED EXERCISES	561
---------------------------------------------------	------------

INDEX	577
--------------	------------