

**Texts in Statistical Science**

# **Understanding Advanced Statistical Methods**

**Peter H. Westfall**

Information Systems and Quantitative Sciences  
Texas Tech University, USA

**Kevin S. S. Henning**

Department of Economics and International Business  
Sam Houston State University, USA



**CRC Press**

Taylor & Francis Group

Boca Raton London New York

---

CRC Press is an imprint of the  
Taylor & Francis Group an **informa** business  
A CHAPMAN & HALL BOOK

---

# Contents

---

List of Examples .....	xiii
Preface.....	xix
Acknowledgments .....	xxiii
Authors .....	xxv
<b>1. Introduction: Probability, Statistics, and Science .....</b>	<b>1</b>
1.1 Reality, Nature, Science, and Models .....	1
1.2 Statistical Processes: Nature, Design and Measurement, and Data .....	3
1.3 Models .....	7
1.4 Deterministic Models .....	8
1.5 Variability.....	9
1.6 Parameters.....	11
1.7 Purely Probabilistic Statistical Models .....	12
1.8 Statistical Models with Both Deterministic and Probabilistic Components.....	16
1.9 Statistical Inference.....	18
1.10 Good and Bad Models.....	20
1.11 Uses of Probability Models.....	24
Vocabulary and Formula Summaries.....	30
Exercises .....	32
<b>2. Random Variables and Their Probability Distributions .....</b>	<b>37</b>
2.1 Introduction .....	37
2.2 Types of Random Variables: Nominal, Ordinal, and Continuous.....	37
2.3 Discrete Probability Distribution Functions.....	40
2.4 Continuous Probability Distribution Functions.....	44
2.5 Some Calculus—Derivatives and Least Squares.....	58
2.6 More Calculus—Integrals and Cumulative Distribution Functions.....	65
Vocabulary and Formula Summaries.....	74
Exercises .....	77
<b>3. Probability Calculation and Simulation.....</b>	<b>83</b>
3.1 Introduction .....	83
3.2 Analytic Calculations, Discrete and Continuous Cases.....	84
3.3 Simulation-Based Approximation .....	86
3.4 Generating Random Numbers.....	87
Vocabulary and Formula Summaries.....	90
Exercises .....	91
<b>4. Identifying Distributions .....</b>	<b>95</b>
4.1 Introduction .....	95
4.2 Identifying Distributions from Theory Alone.....	96
4.3 Using Data: Estimating Distributions via the Histogram.....	99
4.4 Quantiles: Theoretical and Data-Based Estimates .....	105
4.5 Using Data: Comparing Distributions via the Quantile–Quantile Plot .....	108
4.6 Effect of Randomness on Histograms and $q$ – $q$ Plots .....	110

Vocabulary and Formula Summaries.....	113
Exercises .....	114
<b>5. Conditional Distributions and Independence .....</b>	<b>117</b>
5.1 Introduction .....	117
5.2 Conditional Discrete Distributions .....	119
5.3 Estimating Conditional Discrete Distributions.....	121
5.4 Conditional Continuous Distributions .....	122
5.5 Estimating Conditional Continuous Distributions.....	124
5.6 Independence.....	125
Vocabulary and Formula Summaries.....	132
Exercises .....	133
<b>6. Marginal Distributions, Joint Distributions, Independence, and Bayes' Theorem.....</b>	<b>137</b>
6.1 Introduction .....	137
6.2 Joint and Marginal Distributions .....	139
6.3 Estimating and Visualizing Joint Distributions .....	145
6.4 Conditional Distributions from Joint Distributions .....	147
6.5 Joint Distributions When Variables Are Independent.....	150
6.6 Bayes' Theorem .....	153
Vocabulary and Formula Summaries.....	160
Exercises .....	161
<b>7. Sampling from Populations and Processes.....</b>	<b>165</b>
7.1 Introduction .....	165
7.2 Sampling from Populations.....	167
7.3 Critique of the Population Interpretation of Probability Models.....	172
7.3.1 Even When Data Are Sampled from a Population .....	172
7.3.2 Point 1: Nature Defines the Population, Not Vice Versa .....	172
7.3.3 Point 2: The Population Is Not Well Defined .....	173
7.3.4 Point 3: Population Conditional Distributions Are Discontinuous.....	173
7.3.5 Point 4: The Conditional Population Distribution $p(y x)$ Does Not Exist for Many $x$ .....	174
7.3.6 Point 5: The Population Model Ignores Design and Measurement Effects .....	175
7.4 The Process Model versus the Population Model .....	182
7.5 Independent and Identically Distributed Random Variables and Other Models .....	183
7.6 Checking the iid Assumption .....	187
Vocabulary and Formula Summaries.....	196
Exercises .....	198
<b>8. Expected Value and the Law of Large Numbers .....</b>	<b>201</b>
8.1 Introduction .....	201
8.2 Discrete Case .....	201
8.3 Continuous Case .....	204
8.4 Law of Large Numbers .....	207

8.5	Law of Large Numbers for the Bernoulli Distribution .....	214
8.6	Keeping the Terminology Straight: Mean, Average, Sample Mean, Sample Average, and Expected Value .....	214
8.7	Bootstrap Distribution and the Plug-In Principle .....	216
	Vocabulary and Formula Summaries.....	218
	Exercises .....	220
<b>9.</b>	<b>Functions of Random Variables: Their Distributions and Expected Values .....</b>	<b>223</b>
9.1	Introduction .....	223
9.2	Distributions of Functions: The Discrete Case .....	223
9.3	Distributions of Functions: The Continuous Case .....	225
9.4	Expected Values of Functions and the Law of the Unconscious Statistician... ..	227
9.5	Linearity and Additivity Properties.....	228
9.6	Nonlinear Functions and Jensen’s Inequality.....	231
9.7	Variance .....	235
9.8	Standard Deviation, Mean Absolute Deviation, and Chebyshev’s Inequality .....	239
9.9	Linearity Property of Variance .....	244
9.10	Skewness and Kurtosis .....	248
	Vocabulary and Formula Summaries.....	254
	Exercises .....	256
<b>10.</b>	<b>Distributions of Totals .....</b>	<b>261</b>
10.1	Introduction .....	261
10.2	Additivity Property of Variance .....	261
10.3	Covariance and Correlation .....	267
10.4	Central Limit Theorem.....	272
	Vocabulary and Formula Summaries.....	277
	Exercises .....	279
<b>11.</b>	<b>Estimation: Unbiasedness, Consistency, and Efficiency .....</b>	<b>283</b>
11.1	Introduction .....	283
11.2	Biased and Unbiased Estimators .....	284
11.3	Bias of the Plug-In Estimator of Variance.....	287
11.4	Removing the Bias of the Plug-In Estimator of Variance .....	292
11.5	The Joke Is on Us: The Standard Deviation Estimator Is Biased after All.....	294
11.6	Consistency of Estimators.....	296
11.7	Efficiency of Estimators.....	298
	Vocabulary and Formula Summaries.....	303
	Exercises .....	304
<b>12.</b>	<b>Likelihood Function and Maximum Likelihood Estimates .....</b>	<b>307</b>
12.1	Introduction .....	307
12.2	Likelihood Function .....	307
12.3	Maximum Likelihood Estimates.....	318
12.4	Wald Standard Error.....	334
	Vocabulary and Formula Summaries.....	337
	Exercises .....	338

<b>13. Bayesian Statistics</b> .....	343
13.1 Introduction: Play a Game with Hans! .....	343
13.2 Prior Information and Posterior Knowledge .....	345
13.3 Case of the Unknown Survey .....	346
13.4 Bayesian Statistics: The Overview .....	349
13.5 Bayesian Analysis of the Bernoulli Parameter .....	350
13.6 Bayesian Analysis Using Simulation.....	356
13.7 What Good Is Bayes? .....	359
Vocabulary and Formula Summaries.....	368
Exercises .....	369
<b>14. Frequentist Statistical Methods</b> .....	373
14.1 Introduction .....	373
14.2 Large-Sample Approximate Frequentist Confidence Interval for the Process Mean .....	375
14.3 What Does <i>Approximate</i> Really Mean for an Interval Range? .....	381
14.4 Comparing the Bayesian and Frequentist Paradigms .....	384
Vocabulary and Formula Summaries.....	386
Exercises .....	387
<b>15. Are Your Results Explainable by Chance Alone?</b> .....	389
15.1 Introduction .....	389
15.2 What Does <i>by Chance Alone</i> Mean? .....	390
15.3 The <i>p</i> -Value.....	395
15.4 The Extremely Ugly " $pv \leq 0.05$ " Rule of Thumb .....	399
Vocabulary and Formula Summaries.....	405
Exercises .....	407
<b>16. Chi-Squared, Student's <i>t</i>, and <i>F</i>-Distributions, with Applications</b> .....	411
16.1 Introduction .....	411
16.2 Linearity and Additivity Properties of the Normal Distribution.....	412
16.3 Effect of Using an Estimate of $\sigma$ .....	413
16.4 Chi-Squared Distribution .....	416
16.5 Frequentist Confidence Interval for $\sigma$ .....	420
16.6 Student's <i>t</i> -Distribution.....	422
16.7 Comparing Two Independent Samples Using a Confidence Interval .....	426
16.8 Comparing Two Independent Homoscedastic Normal Samples via Hypothesis Testing .....	432
16.9 <i>F</i> -Distribution and ANOVA Test.....	435
16.10 <i>F</i> -Distribution and Comparing Variances of Two Independent Groups .....	441
Vocabulary and Formula Summaries.....	444
Exercises .....	448
<b>17. Likelihood Ratio Tests</b> .....	451
17.1 Introduction .....	451
17.2 Likelihood Ratio Method for Constructing Test Statistics.....	452
17.3 Evaluating the Statistical Significance of Likelihood Ratio Test Statistics .....	467

17.4	Likelihood Ratio Goodness-of-Fit Tests.....	474
17.5	Cross-Classification Frequency Tables and Tests of Independence.....	480
17.6	Comparing Non-Nested Models via the AIC Statistic .....	483
	Vocabulary and Formula Summaries.....	485
	Exercises .....	487
<b>18.</b>	<b>Sample Size and Power .....</b>	<b>491</b>
18.1	Introduction .....	491
18.2	Choosing a Sample Size for a Prespecified Accuracy Margin .....	493
18.3	Power .....	496
18.4	Noncentral Distributions .....	503
18.5	Choosing a Sample Size for Prespecified Power .....	506
18.6	Post Hoc Power: A Useless Statistic.....	508
	Vocabulary and Formula Summaries.....	510
	Exercises .....	511
<b>19.</b>	<b>Robustness and Nonparametric Methods .....</b>	<b>515</b>
19.1	Introduction .....	515
19.2	Nonparametric Tests Based on the Rank Transformation.....	517
19.3	Randomization Tests .....	519
19.4	Level and Power Robustness.....	522
19.5	Bootstrap Percentile- $t$ Confidence Interval .....	526
	Vocabulary and Formula Summaries.....	530
	Exercises .....	531
<b>20.</b>	<b>Final Words .....</b>	<b>533</b>
	<b>Index .....</b>	<b>535</b>