



Applied Statistics for Civil and Environmental Engineers

SECOND EDITION

Nathabandu T. Kottegoda and Renzo Rosso



**Blackwell
Publishing**

Contents

<i>Dedication</i>	xiii
<i>Preface to the First Edition</i>	xiv
<i>Preface to the Second Edition</i>	xvi
Introduction	1
1 Preliminary Data Analysis	3
1.1 Graphical Representation	3
1.1.1 Line diagram or bar chart	4
1.1.2 Dot diagram	4
1.1.3 Histogram	5
1.1.4 Frequency polygon	8
1.1.5 Cumulative relative frequency diagram	9
1.1.6 Duration curves	10
1.1.7 Summary of Section 1.1	11
1.2 Numerical Summaries of Data	11
1.2.1 Measures of central tendency	12
1.2.2 Measures of dispersion	15
1.2.3 Measure of asymmetry	19
1.2.4 Measure of peakedness	19
1.2.5 Summary of Section 1.2	19
1.3 Exploratory Methods	20
1.3.1 Stem-and-leaf plot	20
1.3.2 Box plot	22
1.3.3 Summary of Section 1.3	23
1.4 Data Observed in Pairs	23
1.4.1 Correlation and graphical plots	23
1.4.2 Covariance and the correlation coefficient	24
1.4.3 Q-Q plots	26
1.4.4 Summary of Section 1.4	27
1.5 Summary for Chapter 1	27
References	28
Problems	29
2 Basic Probability Concepts	38
2.1 Random Events	39
2.1.1 Sample space and events	39
2.1.2 The null event, intersection, and union	41
2.1.3 Venn diagram and event space	43
2.1.4 Summary of Section 2.1	49

2.2	Measures of Probability	50
2.2.1	Interpretations of probability	50
2.2.2	Probability axioms	52
2.2.3	Addition rule	53
2.2.4	Further properties of probability functions	55
2.2.5	Conditional probability and multiplication rule	56
2.2.6	Stochastic independence	61
2.2.7	Total probability and Bayes' theorems	65
2.2.8	Summary of Section 2.2	72
2.3	Summary for Chapter 2	72
	References	73
	Problems	74
3	Random Variables and Their Properties	83
3.1	Random Variables and Probability Distributions	83
3.1.1	Random variables	83
3.1.2	Probability mass function	84
3.1.3	Cumulative distribution function of a discrete random variable	85
3.1.4	Probability density function	86
3.1.5	Cumulative distribution function of a continuous random variable	88
3.1.6	Summary of Section 3.1	90
3.2	Descriptors of Random Variables	90
3.2.1	Expectation and other population measures	90
3.2.2	Generating functions	99
3.2.3	Estimation of parameters	103
3.2.4	Summary of Section 3.2	112
3.3	Multiple Random Variables	112
3.3.1	Joint probability distributions of discrete variables	113
3.3.2	Joint probability distributions of continuous variables	118
3.3.3	Properties of multiple variables	124
3.3.4	Summary of Section 3.3	132
3.4	Associated Random Variables and Probabilities	132
3.4.1	Functions of a random variable	133
3.4.2	Functions of two or more variables	135
3.4.3	Properties of derived variables	143
3.4.4	Compound variables	151
3.4.5	Summary of Section 3.4	154
3.5	Copulas	154
3.6	Summary for Chapter 3	157
	References	157
	Problems	160
4	Probability Distributions	165
4.1	Discrete Distributions	165
4.1.1	Bernoulli distribution	166
4.1.2	Binomial distribution	167
4.1.3	Poisson distribution	171
4.1.4	Geometric and negative binomial distributions	181

4.1.5	Log-series distribution	185
4.1.6	Multinomial distribution	187
4.1.7	Hypergeometric distribution	189
4.1.8	Summary of Section 4.1	192
4.2	Continuous Distributions	194
4.2.1	Uniform distribution	194
4.2.2	Exponential distribution	196
4.2.3	Erlang and gamma distribution	200
4.2.4	Beta distribution	203
4.2.5	Weibull distribution	205
4.2.6	Normal distribution	209
4.2.7	Lognormal distribution	215
4.2.8	Summary of Section 4.2	217
4.3	Multivariate Distributions	217
4.3.1	Bivariate normal distribution	219
4.3.2	Other bivariate distributions	222
4.4	Summary for Chapter 4	222
	References	223
	Problems	224
5	Model Estimation and Testing	230
5.1	A Review of Terms Related to Random Sampling	230
5.2	Properties of Estimators	231
5.2.1	Unbiasedness	231
5.2.2	Consistency	232
5.2.3	Minimum variance	232
5.2.4	Efficiency	234
5.2.5	Sufficiency	234
5.2.6	Summary of Section 5.2	235
5.3	Estimation of Confidence Intervals	236
5.3.1	Confidence interval estimation of the mean when the standard deviation is known	236
5.3.2	Confidence interval estimation of the mean when the standard deviation is unknown	239
5.3.3	Confidence interval for a proportion	242
5.3.4	Sampling distribution of differences and sums of statistics	242
5.3.5	Interval estimation for the variance: chi-squared distribution	243
5.3.6	Summary of Section 5.3	247
5.4	Hypothesis Testing	247
5.4.1	Procedure for testing	248
5.4.2	Probabilities of Type I and Type II errors and the power function	254
5.4.3	Neyman-Pearson lemma	256
5.4.4	Tests of hypotheses involving the variance	257
5.4.5	The F distribution and its use	258
5.4.6	Summary of Section 5.4	259
5.5	Nonparametric Methods	260
5.5.1	Sign test applied to the median	261
5.5.2	Wilcoxon signed-rank test for association of paired observations	262

5.5.3	Kruskal-Wallis test for paired observations in k samples	264
5.5.4	Tests on randomness: runs test	267
5.5.5	Spearman's rank correlation coefficient	268
5.5.6	Summary of Section 5.5	269
5.6	Goodness-of-Fit Tests	270
5.6.1	Chi-squared goodness-of-fit test	271
5.6.2	Kolmogorov-Smirnov goodness-of-fit test	273
5.6.3	Kolmogorov-Smirnov two-sample test	274
5.6.4	Anderson-Darling goodness-of-fit test	277
5.6.5	Other methods for testing the goodness-of-fit to a normal distribution	281
5.6.6	Summary of Section 5.6	282
5.7	Analysis of Variance	283
5.7.1	One-way analysis of variance	284
5.7.2	Two-way analysis of variance	288
5.7.3	Summary of Section 5.7	294
5.8	Probability Plotting Methods and Visual Aids	295
5.8.1	Probability plotting for uniform distribution	296
5.8.2	Probability plotting for normal distribution	297
5.8.3	Probability plotting for Gumbel or EV1 distribution	300
5.8.4	Probability plotting of other distributions	301
5.8.5	Visual fitting methods based on the histogram	303
5.8.6	Summary of Section 5.8	305
5.9	Identification and Accommodation of Outliers	305
5.9.1	Hypothesis tests	306
5.9.2	Test statistics for detection of outliers	307
5.9.3	Dealing with nonnormal data	309
5.9.4	Estimation of probabilities of extreme events when outliers are present	311
5.9.5	Summary of Section 5.9	312
5.10	Summary of Chapter 5	312
	References	313
	Problems	316
6	Methods of Regression and Multivariate Analysis	326
6.1	Simple Linear Regression	327
6.1.1	Estimates of the parameters	328
6.1.2	Properties of the estimators and errors	332
6.1.3	Tests of significance and confidence intervals	337
6.1.4	The bivariate normal model and correlation	339
6.1.5	Summary of Section 6.1	342
6.2	Multiple Linear Regression	342
6.2.1	Formulation of the model	343
6.2.2	Linear least squares solutions using the matrix method	343
6.2.3	Properties of least squares estimators and error variance	346
6.2.4	Model testing	350
6.2.5	Model adequacy	355
6.2.6	Residual plots	356
6.2.7	Influential observations and outliers in regression	358
6.2.8	Transformations	365

6.2.9	Confidence intervals on mean response and prediction	366
6.2.10	Ridge regression	368
6.2.11	Other methods and discussion of Section 6.2	370
6.3	Multivariate Analysis	373
6.3.1	Principal components analysis	373
6.3.2	Factor analysis	379
6.3.3	Cluster analysis	383
6.3.4	Other methods and summary of Section 6.3	385
6.4	Spatial Correlation	386
6.4.1	The estimation problem	387
6.4.2	Spatial correlation and the semivariogram	387
6.4.3	Some semivariogram models and physical aspects	389
6.4.4	Spatial interpolations and Kriging	391
6.4.5	Summary of Section 6.4	394
6.5	Summary of Chapter 6	394
	References	395
	Problems	398
7	Frequency Analysis of Extreme Events	405
7.1	Order Statistics	406
7.1.1	Definitions and distributions	406
7.1.2	Functions of order statistics	409
7.1.3	Expected value and variance of order statistics	411
7.1.4	Summary of Section 7.1	415
7.2	Extreme Value Distributions	415
7.2.1	Basic concepts of extreme value theory	415
7.2.2	Gumbel distribution	422
7.2.3	Fréchet distribution	429
7.2.4	Weibull distribution as an extreme value model	432
7.2.5	General extreme value distribution	435
7.2.6	Contagious extreme value distributions	439
7.2.7	Use of other distributions as extreme value models	445
7.2.8	Summary of Section 7.2	450
7.3	Analysis of Natural Hazards	453
7.3.1	Floods, storms, and droughts	453
7.3.2	Earthquakes and volcanic eruptions	461
7.3.3	Winds	465
7.3.4	Sea levels and highest sea waves	470
7.3.5	Summary of Section 7.3	473
7.4	Summary of Chapter 7	474
	References	474
	Problems	478
8	Simulation Techniques for Design	487
8.1	Monte Carlo Simulation	488
8.1.1	Statistical experiments	488
8.1.2	Probability integral transform	493
8.1.3	Sample size and accuracy of Monte Carlo experiments	495
8.1.4	Summary for Section 8.1	501
8.2	Generation of Random Numbers	501

8.2.1	Random outcomes from standard uniform variates	501
8.2.2	Random outcomes from continuous variates	506
8.2.3	Random outcomes from discrete variates	511
8.2.4	Random outcomes from jointly distributed variates	513
8.2.5	Summary of Section 8.2	514
8.3	Use of Simulation	514
8.3.1	Distributions of derived design variates	514
8.3.2	Sampling statistics	517
8.3.3	Simulation of time- or space-varying systems	519
8.3.4	Design alternatives and optimal design	524
8.3.5	Summary of Section 8.3	530
8.4	Sensitivity and Uncertainty Analysis	530
8.5	Summary and Discussion of Chapter 8	531
	References	531
	Problems	533
9	Risk and Reliability Analysis	541
9.1	Measures of Reliability	542
9.1.1	Factors of safety	542
9.1.2	Safety margin	547
9.1.3	Reliability index	550
9.1.4	Performance function and limiting state	558
9.1.5	Further practical solutions	568
9.1.6	Summary of Section 9.1	577
9.2	Multiple Failure Modes	577
9.2.1	Independent failure modes	578
9.2.2	Mutually dependent failure modes	584
9.2.3	Summary of Section 9.2	592
9.3	Uncertainty in Reliability Assessments	592
9.3.1	Reliability limits	592
9.3.2	Bayesian revision of reliability	593
9.3.3	Summary of Section 9.3	597
9.4	Temporal Reliability	597
9.4.1	Failure process and survival time	597
9.4.2	Hazard function	602
9.4.3	Reliable life	605
9.4.4	Summary of Section 9.4	606
9.5	Reliability-Based Design	606
9.6	Summary for Chapter 9	612
	References	613
	Problems	615
10	Bayesian Decision Methods and Parameter Uncertainty	623
10.1	Basic Decision Theory	624
10.1.1	Bayes' rules	624
10.1.2	Decision trees	627
10.1.3	The minimax solution	630
10.1.4	Summary of Section 10.1	632
10.2	Posterior Bayesian Decision Analysis	632
10.2.1	Subjective probabilities	633

10.2.2	Loss and utility functions	634
10.2.3	The discrete case	635
10.2.4	Inference with conditional binomial and prior beta	636
10.2.5	Poisson hazards and gamma prior	638
10.2.6	<i>Inferences with normal distribution</i>	639
10.2.7	Likelihood ratio testing	642
10.2.8	Summary of Section 10.2	643
10.3	Markov Chain Monte Carlo Methods	643
10.4	James-Stein Estimators	650
10.5	Summary and Discussion of Chapter 10	653
	References	653
	Problems	656
Appendix A: Further mathematics		659
A.1	Chebyshev Inequality	659
A.2	Convex Function and Jensen Inequality	659
A.3	Derivation of the Poisson distribution	659
A.4	Derivation of the normal distribution	660
A.5	MGF of the normal distribution	661
A.6	Central limit theorem	662
A.7	Pdf of Student's T distribution	663
A.8	Pdf of the F distribution	664
A.9	Wilcoxon signed-rank test: mean and variance of the test statistic	664
A.10	Spearman's rank correlation coefficient	665
Appendix B: Glossary of Symbols		667
Appendix C: Tables of Selected Distributions		673
Appendix D: Brief Answers to Selected Problems		684
Appendix E: Data Lists		687
	<i>Index</i>	707