

Contents

Preface	xiii
1 Introduction	1
1.1 Who Should Read This Book, 6	
1.2 How This Book is Organized, 6	
1.3 How to Read This Book and Learn from It, 7	
1.4 Note for Instructors, 8	
1.5 Book Web Site, 9	
2 Fundamentals of Statistics	11
2.1 Statistical Thinking, 11	
2.2 Data Format, 13	
2.3 Descriptive Statistics, 14	
2.3.1 Measures of Location, 14	
2.3.2 Measures of Variability, 16	
2.4 Data Visualization, 17	
2.4.1 Dot Plots, 17	
2.4.2 Histograms, 19	
2.4.3 Box Plots, 23	
2.4.4 Scatter Plots, 24	
2.5 Probability and Probability Distributions, 26	
2.5.1 Probability and Its Properties, 26	
2.5.2 Probability Distributions, 30	
2.5.3 Expected Value and Moments, 33	
2.5.4 Joint Distributions and Independence, 34	
2.5.5 Covariance and Correlation, 38	

- 2.6 Rules of Two and Three Sigma, 40
- 2.7 Sampling Distributions and the Laws of Large Numbers, 41
- 2.8 Skewness and Kurtosis, 44

3 Statistical Inference

51

- 3.1 Introduction, 51
- 3.2 Point Estimation of Parameters, 53
 - 3.2.1 Definition and Properties of Estimators, 53
 - 3.2.2 The Method of the Moments and Plug-In Principle, 56
 - 3.2.3 The Maximum Likelihood Estimation, 57
- 3.3 Interval Estimation, 60
- 3.4 Hypothesis Testing, 63
- 3.5 Samples From Two Populations, 71
- 3.6 Probability Plots and Testing for Population Distributions, 73
 - 3.6.1 Probability Plots, 74
 - 3.6.2 Kolmogorov–Smirnov Statistic, 75
 - 3.6.3 Chi-Squared Test, 76
 - 3.6.4 Ryan–Joiner Test for Normality, 76
- 3.7 Outlier Detection, 77
- 3.8 Monte Carlo Simulations, 79
- 3.9 Bootstrap, 79

4 Statistical Models

85

- 4.1 Introduction, 85
- 4.2 Regression Models, 85
 - 4.2.1 Simple Linear Regression Model, 86
 - 4.2.2 Residual Analysis, 94
 - 4.2.3 Multiple Linear Regression and Matrix Notation, 96
 - 4.2.4 Geometric Interpretation in an n-Dimensional Space, 99
 - 4.2.5 Statistical Inference in Multiple Linear Regression, 100
 - 4.2.6 Prediction of the Response and Estimation of the Mean Response, 104
 - 4.2.7 More on Checking the Model Assumptions, 107
 - 4.2.8 Other Topics in Regression, 110
- 4.3 Experimental Design and Analysis, 111
 - 4.3.1 Analysis of Designs with Qualitative Factors, 116
 - 4.3.2 Other Topics in Experimental Design, 124

Supplement 4A. Vector and Matrix Algebra, 125

Vectors, 125

Matrices, 127

Eigenvalues and Eigenvectors of Matrices, 130

Spectral Decomposition of Matrices, 130

Positive Definite Matrices, 131

A Square Root Matrix, 131

Supplement 4B. Random Vectors and Matrices, 132

Sphering, 134

5 Fundamentals of Multivariate Statistics

137

5.1 Introduction, 137

5.2 The Multivariate Random Sample, 139

5.3 Multivariate Data Visualization, 143

5.4 The Geometry of the Sample, 148

5.4.1 The Geometric Interpretation of the Sample Mean, 148

5.4.2 The Geometric Interpretation of the Sample Standard Deviation, 149

5.4.3 The Geometric Interpretation of the Sample Correlation Coefficient, 150

5.5 The Generalized Variance, 151

5.6 Distances in the p -Dimensional Space, 159

5.7 The Multivariate Normal (Gaussian) Distribution, 163

5.7.1 The Definition and Properties of the Multivariate Normal Distribution, 163

5.7.2 Properties of the Mahalanobis Distance, 166

6 Multivariate Statistical Inference

173

6.1 Introduction, 173

6.2 Inferences About a Mean Vector, 173

6.2.1 Testing the Multivariate Population Mean, 173

6.2.2 Interval Estimation for the Multivariate Population Mean, 175

6.2.3 T^2 Confidence Regions, 179

6.3 Comparing Mean Vectors from Two Populations, 183

6.3.1 Equal Covariance Matrices, 184

6.3.2 Unequal Covariance Matrices and Large Samples, 185

6.3.3 Unequal Covariance Matrices and Samples Sizes Not So Large, 186

- 6.4 Inferences About a Variance–Covariance Matrix, 187
- 6.5 How to Check Multivariate Normality, 188

7 Principal Component Analysis

193

- 7.1 Introduction, 193
- 7.2 Definition and Properties of Principal Components, 195
 - 7.2.1 Definition of Principal Components, 195
 - 7.2.2 Finding Principal Components, 196
 - 7.2.3 Interpretation of Principal Component Loadings, 200
 - 7.2.4 Scaling of Variables, 207
- 7.3 Stopping Rules for Principal Component Analysis, 209
 - 7.3.1 Fair-Share Stopping Rules, 210
 - 7.3.2 Large-Gap Stopping Rules, 213
- 7.4 Principal Component Scores, 217
- 7.5 Residual Analysis, 220
- 7.6 Statistical Inference in Principal Component Analysis, 227
 - 7.6.1 Independent and Identically Distributed Observations, 227
 - 7.6.2 Imaging Related Sampling Schemes, 228
- 7.7 Further Reading, 238

8 Canonical Correlation Analysis

241

- 8.1 Introduction, 241
- 8.2 Mathematical Formulation, 242
- 8.3 Practical Application, 245
- 8.4 Calculating Variability Explained by Canonical Variables, 246
- 8.5 Canonical Correlation Regression, 251
- 8.6 Further Reading, 256
- Supplement 8A. Cross-Validation, 256

9 Discrimination and Classification – Supervised Learning

261

- 9.1 Introduction, 261
- 9.2 Classification for Two Populations, 264
 - 9.2.1 Classification Rules for Multivariate Normal Distributions, 267
 - 9.2.2 Cross-Validation of Classification Rules, 277
 - 9.2.3 Fisher’s Discriminant Function, 280

9.3	Classification for Several Populations, 284	
9.3.1	Gaussian Rules, 284	
9.3.2	Fisher's Method, 286	
9.4	Spatial Smoothing for Classification, 291	
9.5	Further Reading, 293	
10	Clustering – Unsupervised Learning	297
10.1	Introduction, 297	
10.2	Similarity and Dissimilarity Measures, 298	
10.2.1	Similarity and Dissimilarity Measures for Observations, 298	
10.2.2	Similarity and Dissimilarity Measures for Variables and Other Objects, 304	
10.3	Hierarchical Clustering Methods, 304	
10.3.1	Single Linkage Algorithm, 305	
10.3.2	Complete Linkage Algorithm, 312	
10.3.3	Average Linkage Algorithm, 315	
10.3.4	Ward Method, 319	
10.4	Nonhierarchical Clustering Methods, 320	
10.4.1	<i>K</i> -Means Method, 320	
10.5	Clustering Variables, 323	
10.6	Further Reading, 325	
Appendix A	Probability Distributions	329
Appendix B	Data Sets	349
Appendix C	Miscellanea	355
References		365
Index		371