

Applied Multivariate Analysis and Related Topics with R

Lang Wu and Jin Qiu

Contents

1	Introduction	3
1.1	Goal of Statistics	3
1.2	Univariate Analysis	5
1.3	Multivariate Analysis	9
1.4	Multivariate Normal Distribution	18
1.5	Unsupervised Learning and Supervised Learning	22
1.6	Data Analysis Strategies and Statistical Thinking	25
1.7	Outline	28
2	Principal Components Analysis (PCA)	30
2.1	The Basic Idea	30
2.2	The Principal Components	31
2.3	Choose Number of Principal Components	35
2.4	Considerations in Data Analysis	36
2.5	Examples with R	37
3	Factor Analysis	46
3.1	The Basic Idea	46
3.2	The Factor Analysis Model	47
3.3	Methods for Estimation	48
3.4	Examples in R	50
4	Discriminant Analysis and Cluster Analysis	56
4.1	Introduction	56
4.2	Discriminant Analysis	56
4.3	Cluster Analysis	60
4.4	Examples in R	63
5	Inference for a Multivariate Normal Population	71
5.1	Introduction	71
5.2	Inference for Multivariate Means	72

5.3	Inference for Covariance Matrices	75
5.4	Large Sample Inferences about a Population Mean Vector	76
5.5	Examples in R	76
6	Discrete or Categorical Multivariate Data	81
6.1	Discrete or Categorical Data	81
6.2	The Multinomial Distribution	82
6.3	Contingency Tables	83
6.4	Associations Between Discrete or Categorical Variables	85
6.5	Logit Models for Multinomial Variables	88
6.6	Loglinear Models for Contingency Tables	89
6.7	Example in R	91
7	Copula Models	96
7.1	Introduction	96
7.2	Copula Models	97
7.3	Measures of Dependence	100
7.4	Applications in Actuary and Finance	101
7.5	Applications in Longitudinal and Survival Data*	103
7.6	Example in R	104
8	Linear and Nonlinear Regression Models	108
8.1	Introduction	108
8.2	Linear Regression Models	109
8.3	Model Selection	111
8.4	Model Diagnostics	112
8.5	Data Analysis Examples with R	113
8.6	Nonlinear Regression Models	118
8.7	More on Model Selection	122
9	Generalized Linear Models	127
9.1	Introduction	127
9.2	The Exponential Family	127

9.3	The General Form of a GLM	129
9.4	Inference for GLM	131
9.5	Model Selection and Model Diagnostics	132
9.6	Logistic Regression Models	135
9.7	Poisson Regression Models	139
10	Multivariate Regression and MANOVA Models	146
10.1	Introduction	146
10.2	Multivariate Regression Models	147
10.3	MANOVA Models	149
10.4	Examples in R	151
11	Longitudinal Data, Panel Data, and Repeated Measurements	156
11.1	Introduction	156
11.2	Methods for Longitudinal Data Analysis	157
11.3	Linear Mixed Effects Models	158
11.4	GEE Models	163
12	Methods for Missing Data	167
12.1	Missing Data Mechanisms	167
12.2	Methods for Missing Data	169
12.3	Multiple Imputation Methods	172
12.4	Multiple Imputation by Chained Equations	174
12.5	The EM Algorithm	175
12.6	Example in R	178
13	Robust Multivariate Analysis	183
13.1	The Need for Robust Methods	183
13.2	General Robust Methods	185
13.3	Robust Estimates of the Mean and Standard Deviation	189
13.4	Robust Estimates of the Covariance Matrix	190
13.5	Robust PCA and Regressions	192
13.6	Examples in R	194

14 Selected Topics	199
14.1 Likelihood Methods	199
14.2 Bootstrap Methods	202
14.3 MCMC Methods and the Gibbs Sampler	203
14.4 Survival Analysis	204
14.5 Data Science, Big Data, and Data Mining	207