Multivariate Statistical Methods in Quality Management

Kai Yang
Jayant Trewn

McGRAW-HILL ENGINEERING REFERENCE

Preface xiii

Chapter 1. Multivariate Statistical Methods and Quality 1

1.1 Overview of Multivariate Statistical Methods 1
1.1.1 Graphical multivariate data display and data stratification 3
1.1.2 Multivariate normal distribution and multivariate sampling distribution 3
1.1.3 Multivariate analysis of variance 4
1.1.4 Principal component analysis and factor analysis 5
1.1.5 Discriminant analysis 6
1.1.6 Cluster analysis 7
1.1.7 Mahalanobis Taguchi system (MTS) 7
1.1.8 Path analysis and structural model 8
1.1.9 Multivariate process control 10
1.2 Applications of Multivariate Statistical Methods in Business and Industry 10
1.2.1 Data mining 11
1.2.2 Chemometrics 12
1.2.3 Other applications 13
1.3 Overview of Quality Assurance and Possible Roles of Multivariate Statistical Methods 13
1.3.1 Stage 0: Impetus/ideation 13
1.3.2 Stage 1: Customer and business requirements study 15
1.3.3 Stage 2: Concept development 15
1.3.4 Stage 3: Product/service design/prototyping 15
1.3.5 Stage 4: Manufacturing process preparation/product launch 16
1.3.6 Stage 5: Production 16
1.3.7 Stage 6: Product/service consumption 17
1.3.8 Stage 7: Disposal 17
1.4 Overview of Six Sigma and Possible Roles of Multivariate Statistical Methods 18
1.4.1 Stage 1: Define the project and customer requirements (D or define step) 20
1.4.2 Stage 2: Measuring process performance 21
1.4.3 Stage 3: Analyze data and discover causes of the problem 21
1.4.4 Stage 4: Improve the process 22
1.4.5 Stage 5: Control the process 23

Chapter 2. Graphical Multivariate Data Display and Data Stratification 25

2.1 Introduction 25
2.2 Graphical Templates for Multivariate Data 26
2.2.1 Charts and graphs 26
2.2.2 Templates for displaying multivariate data 29
2.3 Data Visualization and Animation 33
2.3.1 Introduction to data visualization 33
2.4 Multivariate Data Stratification 38
2.4.1 Multi-vari chart technique 39
2.4.2 Graphical analysis of multivariate variation pattern 41

Chapter 3. Introduction to Multivariate Random Variables, Normal Distribution, and Sampling Properties 47

3.1 Overview of Multivariate Random Variables 47
3.2 Multivariate Data Sets and Descriptive Statistics 50
3.2.1 Multivariate data sets 50
3.2.2 Multivariate descriptive statistics 51
3.3 Multivariate Normal Distributions 55
3.3.1 Some properties of the multivariate normal distribution 56
3.4 Multivariate Sampling Distribution 57
3.4.1 Sampling distribution of $\bar{X}$ 57
3.4.2 Sampling distribution of S 58
3.4.3 Central limit theorem applied to multivariate samples 58
3.4.4 Hotelling's T^2 distribution 59
3.4.5 Summary 60
3.5 Multivariate Statistical Inferences on Mean Vectors 60
3.5.1 Small sample multivariate hypothesis testing on a mean vector 62
3.5.2 Large sample multivariate hypothesis testing on a mean vector 63
3.5.3 Small sample multivariate hypothesis testing on the equality of two mean vectors 64
3.5.4 Large sample multivariate hypothesis testing on the equality of two mean vectors 66
3.5.5 Overview of confidence intervals and confidence regions in multivariate statistical inferences 67
3.5.6 Confidence regions and intervals for a single mean vector with small sample size 68
3.5.7 Confidence regions and intervals for a single mean vector with large sample size 70
3.5.8 Confidence regions and intervals for the difference in two population mean vectors for small samples 71
3.5.9 Confidence regions and intervals for the difference in two population mean vectors for large samples 72
3.5.10 Other Cases 73
Appendix 3A: Matrix Algebra Refresher 73
A.1 Introduction 73
A.2 Notations and basic operations 73
A.3 Matrix operations 75

Chapter 4. Multivariate Analysis of Variance 81

4.1 Introduction 81
4.2 Univariate Analysis of Variance (ANOVA) 82
4.2.1 The ANOVA table 85

4.3 Multivariate Analysis of Variance 86
4.3.1 MANOVA model 86
4.3.2 The decomposition of total variation under MANOVA model 88
4.4 MANOVA Case Study 95

Chapter 5. Principal Component Analysis and Factor Analysis 97

5.1 Introduction 97
5.2 Principal Component Analysis Based on Covariance Matrices 99
5.2.1 Two mathematical representations of principal component analysis 100
5.2.2 Properties of principal component analysis 101
5.2.3 Covariance and correlation between X and principal components Y 103
5.2.4 Principal component analysis on sample covariance matrix 103
5.3 Principal Component Analysis Based on Correlation Matrices 108
5.3.1 Principal component scores and score plots 111
5.4 Principal Component Analysis of Dimensional Measurement Data 114
5.4.1 Properties of the geometrical variation mode 117
5.4.2 Variation mode chart 118
5.4.3 Visual display and animation of principal component analysis 121
5.4.4 Applications for other multivariate data 122
5.5 Principal Component Analysis Case Studies 124
5.5.1 Improving automotive dimensional quality by using principal component analysis 124
5.5.2 Performance degradation analysis for IRLEDs (Yang and Yang, 2000) 131
5.6 Factor Analysis 141
5.6.1 Common factor analysis 143
5.6.2 Properties of common factor analysis 143
5.6.3 Parameter estimation in common factor analysis 147
5.7 Factor Rotation 148
5.7.1 Factor rotation for simple structure 149
5.7.2 Procrustes rotation 152
5.8 Factor Analysis Case Studies 152
5.8.1. Characterization of texture and mechanical properties of heat-induced soy protein gels (Kang, Matsumura, and Mori, 1991) 152
5.8.2 Procrustes factor analysis for automobile body assembly process 154
5.8.3 Hinge variation study using procrustes factor analysis 156

Chapter 6. Discriminant Analysis 161

6.1 Introduction 161
6.1.1 Discriminant analysis steps 162
6.2 Linear Discriminant Analysis for Two Normal Populations with Known Covariance Matrix 163
6.3 Linear Discriminant Analysis for Two Normal Population with Equal Covariance Matrices 167
6.4 Discriminant Analysis for Two Normal Population with Unequal Covariance Matrices 169
6.5 Discriminant Analysis for Several Normal Populations 170

6.5.1 Linear discriminant classification 170
6.5.2 Discriminant classification based on the Mahalanobis squared distances 171
6.6 Case Study: Discriminant Analysis of Vegetable Oil by Near-Infrared Reflectance Spectroscopy 175

Chapter 7. Cluster Analysis 181

7.1 Introduction 181
7.2 Distance and Similarity Measures 183
7.2.1 Euclidean distance 183
7.2.2 Standardized euclidean distance 183
7.2.3 Manhattan distance (city block distance) 184
7.2.4 Distance between clusters and linkage method 185
7.2.5 Similarity 189
7.3 Hierarchical Clustering Method 190
7.4 Nonhierarchical Clustering Method (K-Mean Method) 195
7.5 Cereal Brand Case Study 197

Chapter 8. Mahalanobis Distance and Taguchi Method 201

8.1 Introduction 201
8.2 Overview of the Mahalanobis-Taguchi System (MTS) 202
8.2.1 Stage 1: Creation of a baseline Mahalanobis space 203
8.2.2 Stage 2: Test and analysis of the Mahalanobis measure for abnormal samples 205
8.2.3 Stage 3 variable screening by using Taguchi orthogonal array experiments 206
8.2.4 Stage 4: Establish a threshold value (a cutoff MD) based on Taguchi's quality loss function and maintain a multivariate monitoring system 214
8.3 Features of the Mahalanobis-Taguchi System 216
8.4 The Mahalanobis-Taguchi System Case Study 216
8.4.1 Clutch disc inspection 217
8.5 Comments on the Mahalanobis-Taguchi System by Other Researchers and Proposed Alternative Approaches 221
8.5.1 Alternative approaches 221

Chapter 9. Path Analysis and the Structural Model 223

9.1 Introduction 223
9.2 Path Analysis and the Structural Model 225
9.2.1 How to use the path diagram and structural model 228
9.3 Advantages and Disadvantages of Path Analysis and the Structural Model 235
9.3.1 Advantages 235
9.3.2 Disadvantages 236
9.4 Path Analysis Case Studies 237
9.4.1 Path analysis model relating plastic fuel tank characteristics with its hydrocarbon permeation (Hamade, 1996) 237
9.4.2 Path analysis of a foundry process (Price and Barth, 1995) 241

Chapter 10. Multivariate Statistical Process Control 243

10.1 Introduction 243
10.2 Multivariate Control Charts for Given Targets 245
10.2.1 Decomposition of the Hotelling T^2 248
10.3 Two-Phase T^2 Multivariate Control Charts with Subgroups 251
10.3.1 Reference sample and new observations 251
10.3.2 Two-phase T^2 multivariate process control for subgroups 255
10.4 T^2 Control Chart for Individual Observations 259
10.4.1 Phase I reference sample preparation 260
10.4.2 Phase II: Process control for new observations 264
10.5 Principal Component Chart 265

Appendix Probability Distribution Tables 271

References 291

Index 295