

Contents

Preface	13
Software and Supporting Material	21
1 Introduction to Using MATLAB and Data Management	23
1.1 The MATLAB Interface	23
1.2 Running Code	27
1.3 Introduction to Arrays	29
1.3.1 Examples of Array Creation	29
1.3.2 Extracting Data from an Array	31
1.4 Data Types in MATLAB	32
1.5 Cell Arrays	35
1.5.1 Examples of Creating Cell Arrays	35
1.5.2 Extracting Data from a Cell Array	36
1.6 Structures	37
1.6.1 Examples of Creating Structs	37
1.6.2 Extracting Data from a Struct	38
1.7 Tables	38
1.7.1 Examples of Creating Tables	39
1.7.2 Importing a Table from External Files	40
1.7.3 Importing a Table via Tasks	42
1.7.4 Extracting Data from a Table	44
1.7.5 Extracting Data from a Table Based on Criteria . .	47
1.7.6 Adding Two Tables and Additional Table Metadata	49
1.8 Saving Results	50
1.9 Introduction to the Generation of Random Numbers with Spec- ified Distributions	52
1.10 Basic Programming Elements	53
1.10.1 The <code>if</code> Statements and Loops	53
1.10.2 The <code>assert</code> Statement	57

1.10.3	Some Built-In Constants and Special Functions . . .	58
1.11	The <code>dictionary</code>	59
1.11.1	Examples of Creating a <code>dictionary</code>	60
1.11.2	Extracting Data from a <code>dictionary</code>	60
1.12	MATLAB Help System	63
	Summary Exercises	64
2	Linear Algebra	67
2.1	Elementary Matrix Operations	67
2.2	Diagonal Matrices	69
2.3	Some Special Matrices	72
2.4	Linear Algebra Review	73
2.5	Systems of Linear Equations	74
2.6	Elementwise Matrix Operations	75
2.7	Implicit Expansion	77
2.7.1	Implicit Expansion in <code>table</code> Objects	81
2.8	The Norm of a Vector	83
2.9	The Inner (Dot) Product	84
2.10	Quadratic Forms	86
2.11	Vector Spaces	87
2.11.1	Linear Dependence, Independence, and Bases	87
2.11.2	Extracting Elements via Quadratic Forms	88
2.12	Rank	90
2.13	Eigenvalues and Eigenvectors	91
2.14	Characteristic Polynomial	91
2.15	Routines for Computing Eigenvalues and Eigenvectors	93
2.16	Spectral Decomposition	93
2.16.1	Spectral Decomposition via Symbolic Computation .	94
2.17	Orthogonal Projections	95
	Summary Exercises	99
3	Exploratory Data Analysis and Pivot Tables	107
3.1	Creating and Managing Categorical Arrays	108
3.2	Univariate Analysis of Categorical Variables	110
3.3	Univariate Analysis of Quantitative Variables	113
3.4	Univariate Plots for Categorical or Quantitative Data	115
3.5	Frequency Distributions of Quantitative Variables	116
3.6	Univariate Analysis of Quantitative Variables in the Presence of Subgroups	121
3.7	Confidence Intervals and Advanced Summaries	124

3.8	Pivot Tables via the Guided Interface	130
3.9	Advanced Pivot Tables	130
3.10	Boxplots Based on Multiple Grouping Variables	138
3.11	Violin Plots Based on Multiple Grouping Variables	140
	Summary Exercises	140
4	Importing Real-Time Data from the Web	147
4.1	Importing Data from a Web Page	147
4.2	Importing Data from a GitHub Page	148
4.2.1	Managing Files Downloaded via GitHub	150
4.3	Importing Large Datasets	153
4.4	Introduction to timetables	155
4.4.1	Extracting Records from a Timetable	156
4.4.2	Creating Sequences of Dates	158
4.5	Importing ISTAT Time Series	160
4.6	Custom Import Options	163
4.6.1	Duration and calendarDuration	165
4.6.2	Changing Series Frequency	165
4.7	Real-Time Import of Financial or Social-Media Time Series .	166
	Summary Exercises	168
5	Random Variables: Density, Distributions, and Parameter Estimation	173
5.1	The Gaussian or Normal Variable	173
5.2	The Uniform Random Variable	182
5.3	The Chi-Squared Random Variable	186
5.4	Student's t Random Variable	190
5.5	Parameter Estimation for Distributions	192
5.6	The Bivariate Normal Distribution	193
5.7	The Multivariate Normal Distribution	198
5.8	The Discrete Uniform Distribution	200
	Summary Exercises	201
6	Preliminary Data Processing	209
6.1	Missing Data and Strategies for Handling Them	209
6.2	Outliers and Strategies for Handling Them	213
6.3	Automatic Analysis of Missing and Univariate Outlier Patterns	215
6.4	Advanced Operations with Missing Values	218
6.5	Testing the Normality Assumption	220
	Summary Exercises	222

7	Correlation and Cograduation	225
7.1	The Covariance Matrix and the Correlation Matrix	228
7.2	The Sampling Distribution of the Correlation Coefficient when $\rho = 0$	231
7.3	The Sampling Distribution of the Correlation Coefficient when $\rho \neq 0$	236
7.4	Spearman's Rank Correlation	240
7.4.1	Rank Correlation with Ties	240
7.5	The Relationship Between Correlation and Rank Correlation	243
7.6	Correlation and Rank Correlation in the Presence of Missing Values	244
	Summary Exercises	246
8	Graphical Representations	251
8.1	Plots for Univariate Series	251
8.2	Waterfall Charts	252
8.3	Plots for Bivariate Series	253
8.4	Funnel Charts	255
8.5	Univariate Density Estimation	256
8.6	Custom Scatter Plots	259
8.7	Scatter Plots with Histograms or Boxplots at the Margins .	260
8.8	Plots with Bivariate Histograms	263
8.9	Exploratory Plots for Regression Analysis	265
8.10	Three-Dimensional Plots	268
8.11	The Balloon Plot	271
8.12	Bubble Charts	272
8.12.1	Geographic Bubble Maps	274
8.13	The <code>geoplot</code>	278
8.13.1	The <code>geoplotFS</code> App	280
8.13.2	The <code>geoiconchart</code>	281
8.14	Swarm Charts	283
8.15	The Scatter Plot Matrix	285
8.15.1	Scatter Plot Matrix with Bivariate Contours and Regression Lines	285
8.15.2	Scatter Plot Matrix with Numbers, Symbolic Display, and Background Colour	288
8.15.3	Scatter Plot Matrix with a Grouping Variable	290
8.16	Heatmaps	293
8.17	Polar Plots	296
8.17.1	The <code>compassplot</code>	297

8.18	Polar Histogram	298
8.18.1	The <code>polarhistogramFS</code> Function	299
8.19	Parallel Coordinate Plots	301
8.20	Parallel Coordinate Plots with a Grouping Variable	303
8.21	Iconic Representations (Glyphs)	305
8.21.1	Stars	305
8.21.2	Chernoff Faces	306
	Summary Exercises	308
9	Association	315
9.1	Introduction	315
9.2	Association indices for 2×2 tables	315
9.2.1	Measures using Pearson's chi-squared	318
9.2.2	Measures based on the cross-product ratio	321
9.3	Association Indices for $I \times J$ Tables	323
9.4	Association Indices Based on Error Reduction	326
9.5	Association Indices Based on Reduction of Heterogeneity	327
9.6	The <code>corrNominal</code> Function	328
9.7	Association Indices for Ordinal Variables	329
9.8	The <code>corrOrdinal</code> Function	332
	Summary Exercises	334
10	Distances and Similarity Indices	339
10.1	Definition of Distances	339
10.2	Some Types of Distance	340
10.3	Distance and Dissimilarity Indices	344
10.4	Weighted Euclidean Space	345
10.5	Polygons and Distances	347
10.6	Mahalanobis Distance	350
10.6.1	Properties of Mahalanobis Distance	354
10.7	Measurement Scale of Distances	355
10.8	Similarity Indices	356
10.8.1	Similarity for Binary Phenomena	356
10.8.2	Similarity with Mixed Variables	359
	Summary Exercises	361
11	Dimension Reduction	365
11.1	Principal Component Analysis (PCA): Introduction	365
11.2	The First PC as a Linear Combination of the Original Variables	365

11.3	The First k PCs as Linear Combinations of the Original Variables	368
11.3.1	Relationship among Eigenvalues, Trace, and Determinant	369
11.4	Singular Value Decomposition (SVD)	370
11.5	The First k PCs as the Best Rank- k Representation of the Original Variables	372
11.6	PC as Orthogonal Projection of the Points onto a Reduced-Dimension Subspace	374
11.6.1	Regression Line Versus the First-PC Line	376
11.6.2	Reconstructing the Original Matrix by a Lower-Rank Approximation	378
11.6.3	Principal-Axis Line versus Regression Line	380
11.6.4	Principal Components as a Rotation of the Cartesian Axes	381
11.7	Practical Principal Component Analysis	388
11.8	The Biplot	396
11.8.1	First Representation of Row Points and Column Points	399
11.8.2	Second Representation of Row Points and Column Points	400
11.9	Projection Quality for Each Point	405
11.10	The Function <code>pcaFS</code>	409
11.11	Principal Components on \tilde{X} or on Z	416
11.12	Principal Components with Geographic Data	417
	Summary Exercises	422
12	Suggestions for Further Reading	431
	Bibliography	433