
Contents

1	Mathematical Models and Data Analysis	1
1.1	Introduction	1
1.2	Mathematical Models	3
1.2.1	Types of Data	3
1.2.2	What is a System Model?	4
1.2.3	Types of Models	5
1.2.4	Classification of Mathematical Models	6
1.2.5	Models for Sensor Response	10
1.2.6	Block Diagrams	11
1.3	Types of Problems in Mathematical Modeling	12
1.3.1	Background	12
1.3.2	Forward Problems	13
1.3.3	Inverse Problems	15
1.4	What is Data Analysis?	17
1.5	Types of Uncertainty in Data	18
1.6	Types of Applied Data Analysis and Modeling Methods	19
1.7	Example of a Data Collection and Analysis System	20
1.8	Decision Analysis and Data Mining	22
1.9	Structure of Book	22
	Problems	23
2	Probability Concepts and Probability Distributions	27
2.1	Introduction	27
2.1.1	Outcomes and Simple Events	27
2.1.2	Classical Concept of Probability	27
2.1.3	Bayesian Viewpoint of Probability	27
2.2	Classical Probability	28
2.2.1	Permutations and Combinations	28
2.2.2	Compound Events and Probability Trees	28
2.2.3	Axioms of Probability	30
2.2.4	Joint, Marginal and Conditional Probabilities	30
2.3	Probability Distribution Functions	32
2.3.1	Density Functions	32
2.3.2	Expectation and Moments	35
2.3.3	Function of Random Variables	35
2.4	Important Probability Distributions	37
2.4.1	Background	37
2.4.2	Distributions for Discrete Variables	37
2.4.3	Distributions for Continuous Variables	41

2.5	Bayesian Probability	47
2.5.1	Bayes' Theorem	47
2.5.2	Application to Discrete Probability Variables	50
2.5.3	Application to Continuous Probability Variables	52
2.6	Probability Concepts and Statistics	54
	Problems	56
3	Data Collection and Preliminary Data Analysis	61
3.1	Generalized Measurement System	61
3.2	Performance Characteristics of Sensors and Sensing Systems	62
3.2.1	Sensors	62
3.2.2	Types and Categories of Measurements	64
3.2.3	Data Recording Systems	66
3.3	Data Validation and Preparation	66
3.3.1	Limit Checks	66
3.3.2	Independent Checks Involving Mass and Energy Balances	67
3.3.3	Outlier Rejection by Visual Means	67
3.3.4	Handling Missing Data	68
3.4	Descriptive Measures for Sample Data	69
3.4.1	Summary Statistical Measures	69
3.4.2	Covariance and Pearson Correlation Coefficient	71
3.4.3	Data Transformations	72
3.5	Plotting Data	72
3.5.1	Static Graphical Plots	73
3.5.2	High-Interaction Graphical Methods	78
3.5.3	Graphical Treatment of Outliers	79
3.6	Overall Measurement Uncertainty	82
3.6.1	Need for Uncertainty Analysis	82
3.6.2	Basic Uncertainty Concepts: Random and Bias Errors	82
3.6.3	Random Uncertainty of a Measured Variable	83
3.6.4	Bias Uncertainty	84
3.6.5	Overall Uncertainty	84
3.6.6	Chauvenet's Statistical Criterion of Data Rejection	85
3.7	Propagation of Errors	86
3.7.1	Taylor Series Method for Cross-Sectional Data	86
3.7.2	Taylor Series Method for Time Series Data	89
3.7.3	Monte Carlo Method	92
3.8	Planning a Non-intrusive Field Experiment	93
	Problems	96
	References	100
4	Making Statistical Inferences from Samples	103
4.1	Introduction	103
4.2	Basic Univariate Inferential Statistics	103
4.2.1	Sampling Distribution and Confidence Limits of the Mean	103
4.2.2	Hypothesis Test for Single Sample Mean	106
4.2.3	Two Independent Sample and Paired Difference Tests on Means	108
4.2.4	Single and Two Sample Tests for Proportions	112

4.2.5	Single and Two Sample Tests of Variance	113
4.2.6	Tests for Distributions	114
4.2.7	Test on the Pearson Correlation Coefficient	115
4.3	ANOVA Test for Multi-Samples	116
4.3.1	Single-Factor ANOVA	116
4.3.2	Tukey's Multiple Comparison Test	118
4.4	Tests of Significance of Multivariate Data	119
4.4.1	Introduction to Multivariate Methods	119
4.4.2	Hotteling T^2 Test	120
4.5	Non-parametric Methods	122
4.5.1	Test on Spearman Rank Correlation Coefficient	122
4.5.2	Wilcoxon Rank Tests—Two Sample and Paired Tests	123
4.5.3	Kruskall-Wallis—Multiple Samples Test	125
4.6	Bayesian Inferences	125
4.6.1	Background	125
4.6.2	Inference About One Uncertain Quantity	126
4.6.3	Hypothesis Testing	126
4.7	Sampling Methods	128
4.7.1	Types of Sampling Procedures	128
4.7.2	Desirable Properties of Estimators	129
4.7.3	Determining Sample Size During Random Surveys	130
4.7.4	Stratified Sampling for Variance Reduction	132
4.8	Resampling Methods	132
4.8.1	Basic Concept and Types of Methods	132
4.8.2	Application to Probability Problems	134
4.8.3	Application of Bootstrap to Statistical Inference Problems	134
	Problems	135
5	Estimation of Linear Model Parameters Using Least Squares	141
5.1	Introduction	141
5.2	Regression Analysis	141
5.2.1	Objective of Regression Analysis	141
5.2.2	Ordinary Least Squares	142
5.3	Simple OLS Regression	142
5.3.1	Traditional Simple Linear Regression	142
5.3.2	Model Evaluation	144
5.3.3	Inferences on Regression Coefficients and Model Significance	146
5.3.4	Model Prediction Uncertainty	147
5.4	Multiple OLS Regression	148
5.4.1	Higher Order Linear Models: Polynomial, Multivariate	149
5.4.2	Matrix Formulation	151
5.4.3	OLS Parameter Identification	151
5.4.4	Partial Correlation Coefficients	154
5.4.5	Beta Coefficients and Elasticity	154
5.5	Assumptions and Sources of Error During OLS Parameter Estimation	156
5.5.1	Assumptions	156
5.5.2	Sources of Errors During Regression	157
5.6	Model Residual Analysis	157
5.6.1	Detection of Ill-Conditioned Model Residual Behavior	157
5.6.2	Leverage and Influence Data Points	159

5.6.3	Remedies for Non-uniform Model Residuals	161
5.6.4	Serially Correlated Residuals	165
5.6.5	Dealing with Misspecified Models	166
5.7	Other OLS Parameter Estimation Methods	167
5.7.1	Zero-Intercept Models	167
5.7.2	Indicator Variables for Local Piecewise Models—Spline Fits	168
5.7.3	Indicator Variables for Categorical Regressor Models	169
5.7.4	Assuring Model Parsimony—Stepwise Regression	170
5.8	Case Study Example: Effect of Refrigerant Additive on Chiller Performance . . .	172
	Problems	175
6	Design of Experiments	183
6.1	Background	183
6.2	Complete and Incomplete Block Designs	184
6.2.1	Randomized Complete Block Designs	184
6.2.2	Incomplete Factorial Designs—Latin Squares	190
6.3	Factorial Designs	192
6.3.1	2^k Factorial Designs	192
6.3.2	Concept of Orthogonality	196
6.4	Response Surface Designs	199
6.4.1	Applications	199
6.4.2	Phases Involved	199
6.4.3	First and Second Order Models	200
6.4.4	Central Composite Design and the Concept of Rotation	201
	Problems	203
7	Optimization Methods	207
7.1	Background	207
7.2	Terminology and Classification	209
7.2.1	Basic Terminology and Notation	209
7.2.2	Traditional Optimization Methods	210
7.2.3	Types of Objective Functions	210
7.2.4	Sensitivity Analysis or Post Optimality Analysis	210
7.3	Calculus-Based Analytical and Search Solutions	211
7.3.1	Simple Unconstrained Problems	211
7.3.2	Problems with Equality Constraints	211
7.3.3	Lagrange Multiplier Method	212
7.3.4	Penalty Function Method	213
7.4	Numerical Search Methods	214
7.5	Linear Programming	216
7.6	Quadratic Programming	217
7.7	Non-linear Programming	218
7.8	Illustrative Example: Combined Heat and Power System	218
7.9	Global Optimization	221
7.10	Dynamic Programming	222
	Problems	226
8	Classification and Clustering Methods	231
8.1	Introduction	231
8.2	Parametric Classification Approaches	231

8.2.1	Distance Between Measurements	231
8.2.2	Statistical Classification	232
8.2.3	Ordinary Least Squares Regression Method	234
8.2.4	Discriminant Function Analysis	235
8.2.5	Bayesian Classification	238
8.3	Heuristic Classification Methods	240
8.3.1	Rule-Based Methods	240
8.3.2	Decision Trees	240
8.3.3	k Nearest Neighbors	241
8.4	Classification and Regression Trees (CART) and Treed Regression	243
8.5	Clustering Methods	245
8.5.1	Types of Clustering Methods	245
8.5.2	Partitional Clustering Methods	246
8.5.3	Hierarchical Clustering Methods	248
	Problems	249
9	Analysis of Time Series Data	253
9.1	Basic Concepts	253
9.1.1	Introduction	253
9.1.2	Terminology	255
9.1.3	Basic Behavior Patterns	255
9.1.4	Illustrative Data Set	256
9.2	General Model Formulations	257
9.3	Smoothing Methods	257
9.3.1	Arithmetic Moving Average (AMA)	258
9.3.2	Exponentially Weighted Moving Average (EWA)	259
9.4	OLS Regression Models	261
9.4.1	Trend Modeling	261
9.4.2	Trend and Seasonal Models	261
9.4.3	Fourier Series Models for Periodic Behavior	263
9.4.4	Interrupted Time Series	266
9.5	Stochastic Time Series Models	267
9.5.1	Introduction	267
9.5.2	ACF, PACF and Data Detrending	268
9.5.3	ARIMA Models	271
9.5.4	Recommendations on Model Identification	275
9.6	ARMAX or Transfer Function Models	277
9.6.1	Conceptual Approach and Benefit	277
9.6.2	Transfer Function Modeling of Linear Dynamic Systems	277
9.7	Quality Control and Process Monitoring Using Control Chart Methods	279
9.7.1	Background and Approach	279
9.7.2	Shewart Control Charts for Variables and Attributes	280
9.7.3	Statistical Process Control Using Time Weighted Charts	284
9.7.4	Concluding Remarks	285
	Problems	286
10	Parameter Estimation Methods	289
10.1	Background	289
10.2	Concept of Estimability	289

10.2.1	Ill-Conditioning	290
10.2.2	Structural Identifiability	291
10.2.3	Numerical Identifiability	293
10.3	Dealing with Collinear Regressors During Multivariate Regression	294
10.3.1	Problematic Issues	294
10.3.2	Principle Component Analysis and Regression	295
10.3.3	Ridge Regression	298
10.3.4	Chiller Case Study Analysis Involving Collinear Regressors	299
10.3.5	Stagewise Regression	302
10.3.6	Case Study of Stagewise Regression Involving Building Energy Loads	303
10.3.7	Other Methods	307
10.4	Non-OLS Parameter Estimation Methods	307
10.4.1	General Overview	307
10.4.2	Error in Variables (EIV) and Corrected Least Squares	308
10.4.3	Maximum Likelihood Estimation (MLE)	310
10.4.4	Logistic Functions	312
10.5	Non-linear Estimation	315
10.5.1	Models Transformable to Linear in the Parameters	315
10.5.2	Intrinsically Non-linear Models	317
10.6	Computer Intensive Methods	318
10.6.1	Robust Regression	318
10.6.2	Bootstrap Sampling	320
	Problems	321
11	Inverse Methods	327
11.1	Inverse Problems Revisited	327
11.2	Calibration of White Box Models	327
11.2.1	Basic Notions	327
11.2.2	Example of Calibrated Model Development: Global Temperature Model	328
11.2.3	Analysis Techniques Useful for Calibrating Detailed Simulation Models	331
11.2.4	Case Study: Calibrating Detailed Building Energy Simulation Programs to Utility Bills	334
11.3	Model Selection and Identifiability	340
11.3.1	Basic Notions	340
11.3.2	Local Regression—LOWESS Smoothing Method	342
11.3.3	Neural Networks—Multi-Layer Perceptron (MLP)	343
11.3.4	Grey-Box Models and Policy Issues Concerning Dose-Response Behavior	347
11.3.5	State Variable Representation and Compartmental Models	348
11.3.6	Practical Identifiability Issues	351
11.4	Closure	354
11.4.1	Curve Fitting Versus Parameter Estimation	354
11.4.2	Non-intrusive Data Collection	354
11.4.3	Data Fusion and Functional Testing	355
	Problems	355

12 Risk Analysis and Decision-Making	359
12.1 Background	359
12.1.1 Types of Decision-Making Problems and Applications	359
12.1.2 Engineering Decisions Involving Discrete Alternatives	361
12.2 Decision-Making Under Uncertainty	362
12.2.1 General Framework	362
12.2.2 Modeling Problem Structure Using Influence Diagrams and Decision Trees	363
12.2.3 Modeling Chance Events	366
12.2.4 Modeling Outcomes	368
12.2.5 Modeling Risk Attitudes	368
12.2.6 Modeling Multi-attributes or Multiple Objectives	371
12.2.7 Analysis of Low Epistemic but High Aleatory Problems	373
12.2.8 Value of Perfect Information	374
12.2.9 Bayesian Updating Using Sample Information	375
12.3 Risk Analysis	377
12.3.1 Formal Treatment of Risk Analysis	377
12.3.2 Context of Statistical Hypothesis Testing	378
12.3.3 Context of Environmental Risk to Humans	380
12.3.4 Other Areas of Application	381
12.4 Case Study Examples	383
12.4.1 Risk Assessment of Existing Buildings	383
12.4.2 Decision Making While Operating an Engineering System	389
Problems	393
ERRATUM	E1
Appendix	397
A: Statistical Tables	397
B: Large Data Sets	411
C: Solved Examples and Problems with Practical Relevance	420
Index	423