

# ***An Introduction to Computational Statistics***

---

## ***Regression Analysis***

**Robert I. Jennrich**

*University of California, Los Angeles*



**PRENTICE HALL, Englewood Cliffs, New Jersey 07632**

# Contents

## Preface

xiii

## Chapter 1: A Quick Look at a Typical Regression Program 1

1.1	<i>Introduction</i>	1
1.2	<i>Some Concepts and Terminology</i>	1
1.3	<i>Problem Specification</i>	4
1.4	<i>The Output</i>	8
1.5	<i>Problem Specification Using Other Systems</i>	12
1.6	<i>Problems</i>	15

## Chapter 2: Simple Linear Regression 18

2.1	<i>The Simple Linear Regression Model</i>	18
2.2	<i>Least Squares Fitting</i>	19
2.2.1	Observation space representation	19
2.2.2	Variable space representation	20
2.2.3	Normal equations	21
2.3	<i>Statistical Properties of Least Squares Estimates</i>	25
2.3.1	Unbiased errors	26
2.3.2	Uncorrelated constant variance errors	26
2.3.3	Normally distributed errors	29
2.4	<i>Confidence and Prediction Intervals</i>	30
2.4.1	Confidence intervals for $\alpha$ and $\beta$	31

2.4.2	Confidence intervals for the expected response	32
2.4.3	Predicting a new response	33
2.4.4	Testing hypotheses	34
2.5	<i>Random Regressors</i>	36
2.6	<i>A Nonstandard Application: The Two-Sample t-test</i>	37
2.7	<i>Some Potential Problems in Applying Regression</i>	39
2.8	<i>The Regression Phenomenon</i>	44
2.9	<i>Problems</i>	46

## **Chapter 3: Applying Simple Linear Regression** **51**

3.1	<i>Transforming a Given Model to a Linear Model</i>	51
3.2	<i>Correlation and the Coefficient of Determination</i>	53
3.3	<i>Residual Analysis</i>	56
3.3.1	Plots of $\hat{e}$ on $\hat{y}$	56
3.3.2	Normal probability plots	61
3.3.3	Plots on external variables	67
3.4	<i>Weighted Regression</i>	67
3.5	<i>Structure of the Power Transforms</i>	71
3.6	<i>Some Remedial Actions</i>	74
3.6.1	Outliers	74
3.6.2	Searching for a linearizing transform	75
3.6.3	Transforming to stabilize variance	79
3.6.4	The relation between weighting and transforming	82
3.6.5	Nonnormality	83
3.6.6	How does one proceed?	84
3.7	<i>Correlated Errors</i>	85
3.7.1	The effect of correlated errors	85
3.7.2	Detecting correlation	86
3.7.3	Transforming to reduce autocorrelation	88
3.7.4	Confounding of correlation and bias	90
3.8	<i>Problems</i>	92

## **Chapter 4: Multiple Linear Regression** **97**

4.1	<i>The Multiple Linear Regression Model</i>	97
4.2	<i>Least Squares Fitting</i>	98
4.2.1	Observation and variable space representations	98
4.2.2	Normal equations	99
4.2.3	Multiple correlation and the coefficient of determination	101
4.3	<i>Statistical Properties of Least Squares Estimators</i>	102
4.3.1	Unbiased errors	103
4.3.2	Uncorrelated constant variance errors	103
4.3.3	Normally distributed errors	104

4.4	<i>Using the Computer</i>	104
4.5	<i>Confidence and Prediction Intervals</i>	106
4.5.1	Student <i>t</i> -statistics	107
4.5.2	Confidence intervals	108
4.5.3	Prediction intervals	111
4.6	<i>Goodness of Fit and Other Hypotheses</i>	112
4.6.1	The general linear hypothesis test	112
4.6.2	Goodness of fit using replicate observations	115
4.6.3	Robustness to nonnormality	118
4.7	<i>Collinearity</i>	119
4.7.1	Tolerance and variance inflation	119
4.7.2	Polynomial regression	122
4.7.3	Stepwise and ridge regression	123
4.8	<i>Regression Diagnostics</i>	125
4.8.1	Mahalanobis distance	125
4.8.2	Leverage	127
4.8.3	Studentized residuals and detecting outliers	131
4.8.4	Influence	135
4.8.5	Partial residual plots and partial regression plots	140
4.9	<i>Problems</i>	143

## **Chapter 5: Computer Assisted Model Building 151**

5.1	<i>Introduction</i>	151
5.2	<i>Bias Versus Variance</i>	152
5.3	<i>Stepwise Methods</i>	152
5.3.1	Forward selection and backward elimination	153
5.3.2	Stepwise regression	154
5.4	<i>Best Subset Regression</i>	158
5.5	<i>Criteria for Evaluating Subsets</i>	160
5.6	<i>How Does One Proceed?</i>	167
5.7	<i>A Nonstandard Application</i>	168
5.8	<i>Problems</i>	169

## **Chapter 6: The General Linear Model 173**

6.1	<i>Introduction</i>	173
6.2	<i>Mean Vectors and Covariance Matrices</i>	174
6.3	<i>The Multivariate Normal Distribution</i>	175
6.4	<i>Projections</i>	176
6.5	<i>The <math>\chi^2</math>-Theorem</i>	180
6.6	<i>The General Linear Model</i>	181
6.6.1	Least squares estimates	181
6.6.2	Statistical properties	182

6.7	<i>The Linear Regression Model</i>	184
6.8	<i>The Fundamental F-test</i>	186
6.9	<i>The Gauss-Markov Theorem</i>	187
6.10	<i>Generalized Least Squares and Weighted Regression</i>	188
6.11	<i>Problems</i>	190

## **Chapter 7: Analysis of Variance and Covariance** **196**

7.1	<i>Introduction</i>	196
7.2	<i>The One-Way Model</i>	197
7.3	<i>Dummy Variables</i>	202
7.4	<i>The Simple Two-Way Additive Model</i>	203
7.5	<i>The General Two-Way Additive Model</i>	208
7.6	<i>The General Two-Way Model</i>	214
7.7	<i>Higher Order Models</i>	221
7.8	<i>Analysis of Covariance</i>	223
7.9	<i>Transformation and Model Diagnostics</i>	230
7.10	<i>Problems</i>	234

## **Chapter 8: Nonlinear Regression** **238**

8.1	<i>Why Nonlinear Models?</i>	238
8.2	<i>Overview of a Nonlinear Regression Program</i>	239
	8.2.1 Problem specification	240
	8.2.2 Output	241
8.3	<i>The Nonlinear Regression Model</i>	245
8.4	<i>Nonlinear Least Squares Fitting</i>	246
	8.4.1 Observation and variable space representations	246
	8.4.2 Normal equations	248
8.5	<i>The Gauss-Newton Algorithm</i>	248
	8.5.1 The algorithm	249
	8.5.2 Convergence of the unmodified algorithm	250
	8.5.3 Standard modifications	251
	8.5.4 Derivative free versions	255
8.6	<i>Using the Computer: Growth Curves</i>	258
8.7	<i>Statistical Properties</i>	266
	8.7.1 Standard errors	268
	8.7.2 Functions of parameters	270
	8.7.3 Goodness of fit and other hypotheses	276
8.8	<i>Problems</i>	278

<b>Chapter 9: Maximum Likelihood Analysis and Robust Estimation</b>	<b>284</b>
9.1 <i>Fitting Expectations Using Iterative Reweighting</i>	284
9.2 <i>Maximum Likelihood Analysis</i>	285
9.2.1 General results for the exponential family	287
9.2.2 Poisson models	288
9.2.3 Binomial models	293
9.2.4 Multinomial models	297
9.2.5 Exponential survival models	304
9.3 <i>Robust Estimation</i>	306
9.3.1 The location problem	306
9.3.2 Robust regression	308
9.4 <i>Problems</i>	313
<b>Appendix A: Statistical Tables</b>	<b>321</b>
Table I <i>Normal Curve Areas</i>	322
Table II <i>Critical values of <math>t</math></i>	323
Table III <i>Critical Values of <math>\chi^2</math></i>	324
Table IV <i>Percentage Points of the F-Distribution, <math>\alpha = .05</math></i>	326
Table V <i>Percentage Points of the F-Distribution, <math>\alpha = .01</math></i>	328
<b>Appendix B: Partial Answers to Selected Problems</b>	<b>330</b>
<b>Appendix C: References</b>	<b>348</b>
<b>Index</b>	<b>355</b>