Linear Models in Statistics

ALVIN C. RENCHER
Department of Statistics
Brigham Young University
Provo, Utah
Contents

Preface xiii
Acknowledgments xvii

1. Introduction 1
   1.1 Simple Linear Regression Model, 1
   1.2 Multiple Linear Regression Model, 2
   1.3 Analysis of Variance Models, 3

2. Matrix Algebra 5
   2.1 Matrix and Vector Notation, 5
      2.1.1 Matrices, Vectors, and Scalars, 5
      2.1.2 Matrix Equality, 6
      2.1.3 Transpose, 6
      2.1.4 Matrices of Special Form, 7
   2.2 Operations, 8
      2.2.1 Sum of Two Matrices or Two Vectors, 9
      2.2.2 Product of Two Matrices or Two Vectors, 9
   2.3 Partitioned Matrices, 14
   2.4 Rank, 17
   2.5 Inverse, 19
   2.6 Positive Definite Matrices, 21
   2.7 Systems of Equations, 25
   2.8 Generalized Inverse, 29
      2.8.1 Definition and Properties, 29
      2.8.2 Generalized Inverses and Systems of Equations, 32
   2.9 Determinants, 33
   2.10 Orthogonal Vectors and Matrices, 37
2.11 Trace, 39
2.12 Eigenvalues and Eigenvectors, 41
   2.12.1 Definition, 41
   2.12.2 Functions of a Matrix, 44
   2.12.3 Products, 45
   2.12.4 Symmetric Matrices, 46
   2.12.5 Positive Definite and Positive Semidefinite Matrices, 48
2.13 Idempotent Matrices, 48
2.14 Derivatives of Linear Functions and Quadratic Forms, 50

3. Random Vectors and Matrices 62
   3.1 Introduction, 62
   3.2 Means, Variances, Covariances, and Correlations, 63
   3.3 Mean Vectors and Covariance Matrices for Random Vectors, 67
      3.3.1 Mean Vector, 67
      3.3.2 Covariance Matrix, 68
      3.3.3 Generalized Variance, 69
      3.3.4 Standardized Distance, 69
   3.4 Correlation Matrices, 70
   3.5 Mean Vectors and Covariance Matrices for Partitioned Random
      Vectors, 70
   3.6 Linear Functions of Random Vectors, 71
      3.6.1 Means, 72
      3.6.2 Variances and Covariances, 73

4. Multivariate Normal Distribution 77
   4.1 Univariate Normal Density Function, 77
   4.2 Multivariate Normal Density Function, 78
   4.3 Moment-Generating Functions, 80
   4.4 Properties of the Multivariate Normal Distribution, 82
   4.5 Partial Correlation, 89

5. Distribution of Quadratic Forms in y 93
   5.1 Sums of Squares, 93
   5.2 Mean and Variance of Quadratic Forms, 95
   5.3 Noncentral Chi-Square Distribution, 99
   5.4 Noncentral F- and t-distributions, 101
      5.4.1 Noncentral F-distribution, 101
      5.4.2 Noncentral t-distribution, 102
   5.5 Distribution of Quadratic Forms, 103
5.6 Independence of Linear Forms and Quadratic Forms, 105

6. Simple Linear Regression
   6.1 The Model, 112
   6.2 Estimation of $\beta_0$, $\beta_1$, and $\sigma^2$, 113
   6.3 Hypothesis Test and Confidence Interval for $\beta_1$, 116
   6.4 Coefficient of Determination, 118

7. Multiple Regression: Estimation
   7.1 Introduction, 121
   7.2 The Model, 121
   7.3 Estimation of $\beta$ and $\sigma^2$, 125
      7.3.1 Least Squares Estimator for $\beta$, 125
      7.3.2 Properties of the Least Squares Estimator $\hat{\beta}$, 129
      7.3.3 An Estimator for $\sigma^2$, 133
   7.4 Geometry of Least Squares, 136
      7.4.1 Variable Space, 136
      7.4.2 Sample Space, 136
   7.5 The Model in Centered Form, 138
   7.6 Normal Model, 142
      7.6.1 Assumptions, 142
      7.6.2 Maximum Likelihood Estimators for $\beta$ and $\sigma^2$, 142
      7.6.3 Properties of $\hat{\beta}$ and $\hat{\sigma^2}$, 143
   7.7 $R^2$ In Fixed-x Regression, 145
   7.8 Generalized Least Squares: $\text{cov}(y) = \sigma^2 V$, 148
      7.8.1 Estimation of $\beta$ and $\sigma^2$ When $\text{cov}(y) = \sigma^2 V$, 149
      7.8.2 Misspecification of the Error Structure, 151
   7.9 Model Misspecification, 153
   7.10 Orthogonalization, 159

8. Multiple Regression: Tests of Hypotheses and Confidence Intervals
   8.1 Test of Overall Regression, 170
   8.2 Test on a Subset of the $\beta$'s, 175
   8.3 $F$-Test in Terms of $R^2$, 183
   8.4 The General Linear Hypothesis Tests for $H_0: C\beta = 0$ and $H_0: C\beta = t$, 184
      8.4.1 The Test for $H_0: C\beta = 0$, 184
      8.4.2 The Test for $H_0: C\beta = t$, 189
   8.5 Tests on $\beta_j$ and $a'\beta$, 191
      8.5.1 Testing One $\beta_j$ or One $a'\beta$, 191
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.5.2 Testing Several $\beta_j$’s or $a_i'\beta$’s</td>
<td>192</td>
</tr>
<tr>
<td>8.6 Confidence Intervals and Prediction Intervals</td>
<td>197</td>
</tr>
<tr>
<td>8.6.1 Confidence Region for $\beta$</td>
<td>197</td>
</tr>
<tr>
<td>8.6.2 Confidence Interval for $\beta_j$</td>
<td>197</td>
</tr>
<tr>
<td>8.6.3 Confidence Interval for $a_i'\beta$</td>
<td>198</td>
</tr>
<tr>
<td>8.6.4 Confidence Interval for $E(y)$</td>
<td>199</td>
</tr>
<tr>
<td>8.6.5 Prediction Interval for a Future Observation</td>
<td>200</td>
</tr>
<tr>
<td>8.6.6 Confidence Interval for $\sigma^2$</td>
<td>202</td>
</tr>
<tr>
<td>8.6.7 Simultaneous Intervals</td>
<td>203</td>
</tr>
<tr>
<td>8.7 Likelihood Ratio Tests</td>
<td>204</td>
</tr>
<tr>
<td>9. Multiple Regression: Model Validation and Diagnostics</td>
<td>214</td>
</tr>
<tr>
<td>9.1 Residuals</td>
<td>214</td>
</tr>
<tr>
<td>9.2 The Hat Matrix</td>
<td>217</td>
</tr>
<tr>
<td>9.3 Outliers</td>
<td>219</td>
</tr>
<tr>
<td>9.4 Influential Observations and Leverage</td>
<td>222</td>
</tr>
<tr>
<td>10. Multiple Regression: Random $x$’s</td>
<td>228</td>
</tr>
<tr>
<td>10.1 Multivariate Normal Regression Model</td>
<td>229</td>
</tr>
<tr>
<td>10.2 Estimation in Multivariate Normal Regression</td>
<td>229</td>
</tr>
<tr>
<td>10.3 $R^2$ in Multivariate Normal Regression</td>
<td>237</td>
</tr>
<tr>
<td>10.4 Tests and Confidence Intervals</td>
<td>240</td>
</tr>
<tr>
<td>10.5 Effect of Each Variable on $R^2$</td>
<td>245</td>
</tr>
<tr>
<td>10.6 Prediction for Nonnormal Data</td>
<td>247</td>
</tr>
<tr>
<td>10.7 Sample Partial Correlations</td>
<td>249</td>
</tr>
<tr>
<td>11. Analysis of Variance Models</td>
<td>259</td>
</tr>
<tr>
<td>11.1 Non-Full-Rank Models</td>
<td>259</td>
</tr>
<tr>
<td>11.1.1 One-Way Model</td>
<td>259</td>
</tr>
<tr>
<td>11.1.2 Two-Way Model</td>
<td>262</td>
</tr>
<tr>
<td>11.2 Estimation</td>
<td>265</td>
</tr>
<tr>
<td>11.2.1 Estimability of $\beta$</td>
<td>265</td>
</tr>
<tr>
<td>11.2.2 Estimable Functions of $\beta$</td>
<td>268</td>
</tr>
<tr>
<td>11.3 Estimators</td>
<td>272</td>
</tr>
<tr>
<td>11.3.1 Estimators of $\lambda'\beta$</td>
<td>272</td>
</tr>
<tr>
<td>11.3.2 Estimator of $\sigma^2$</td>
<td>277</td>
</tr>
<tr>
<td>11.3.3 Normal Model</td>
<td>278</td>
</tr>
<tr>
<td>11.4 Reparameterization</td>
<td>279</td>
</tr>
<tr>
<td>11.5 Side Conditions</td>
<td>281</td>
</tr>
<tr>
<td>11.6 Testing Hypotheses</td>
<td>284</td>
</tr>
</tbody>
</table>
11.6.1 Testable Hypotheses, 284  
11.6.2 Full and Reduced Model, 285  
11.6.3 General Linear Hypothesis, 287  
11.7 An Illustration of Estimation and Testing, 289  
11.7.1 Estimable Functions, 290  
11.7.2 Testing a Hypothesis, 291  
11.7.3 Orthogonality of Columns of $X$, 293

12. One-Way Analysis of Variance: Balanced Case  
12.1 The One-Way Model, 300  
12.2 Estimable Functions, 301  
12.3 Estimation of Parameters, 302  
12.3.1 Solving the Normal Equations, 302  
12.3.2 An Estimator for $\sigma^2$, 304  
12.4 Testing the Hypothesis $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$, 305  
12.4.1 Full and Reduced Model, 305  
12.4.2 General Linear Hypothesis, 309  
12.5 Expected Mean Squares, 312  
12.5.1 Full and Reduced Model, 313  
12.5.2 General Linear Hypothesis, 315  
12.6 Contrasts, 317  
12.6.1 Hypothesis Test for a Contrast, 317  
12.6.2 Orthogonal Contrasts, 318  
12.6.3 Orthogonal Polynomial Contrasts, 323

13. Two-Way Analysis of Variance: Balanced Case  
13.1 The Two-Way Model, 337  
13.2 Estimable Functions, 338  
13.3 Estimators of $\lambda'\beta$ and $\sigma^2$, 342  
13.3.1 Solving the Normal Equations and Estimating $\lambda'\beta$, 342  
13.3.2 An Estimator for $\sigma^2$, 344  
13.4 Testing Hypotheses, 345  
13.4.1 Test for Interaction, 345  
13.4.2 Tests for Main Effects, 354  
13.5 Expected Mean Squares, 362  
13.5.1 Sums of Squares Approach, 362  
13.5.2 Quadratic Form Approach, 363

14. Analysis of Variance: Unbalanced Data  
14.1 Introduction, 371
14.2 One-Way Model, 372
   14.2.1 Estimation and Testing, 372
   14.2.2 Contrasts, 374
14.3 Two-Way Model, 377
   14.3.1 Unconstrained Model, 378
   14.3.2 Constrained Model, 384

15. Analysis of Covariance 392
   15.1 Introduction, 392
   15.2 Estimation and Testing, 393
      15.2.1 The Analysis of Covariance Model, 393
      15.2.2 Estimation, 395
      15.2.3 Testing Hypotheses, 397
   15.3 One-Way Model with One Covariate, 397
      15.3.1 The Model, 398
      15.3.2 Estimation, 398
      15.3.3 Testing Hypotheses, 399
   15.4 Two-Way Model with One Covariate, 405
      15.4.1 Tests for Main Effects and Interactions, 406
      15.4.2 Test for Slope, 410
      15.4.3 Test for Homogeneity of Slopes, 411
   15.5 One-Way Model with Multiple Covariates, 412
      15.5.1 The Model, 412
      15.5.2 Estimation, 413
      15.5.3 Testing Hypotheses, 416
   15.6 Analysis of Covariance with Unbalanced Models, 421

   16.1 Introduction, 426
   16.2 Estimation of $\lambda^{'}\beta$ and Prediction of $a$ in $y = X\beta + Za + e$, 429
      16.2.1 Best Linear Unbiased Estimator of $\lambda^{'}\beta$, 429
      16.2.2 Best Linear Unbiased Predictor of the Random Vector $a$, 430
   16.3 Estimation of Variance Components, 432
      16.3.1 Expected Mean Squares, 433
      16.3.2 ANOVA Estimators, 436
   16.4 Hypothesis Tests, 436

17. Additional Models 440
   17.1 Nonlinear Regression, 440
   17.2 Logistic Regression, 441
17.3 Loglinear Models, 444
17.4 Poisson Regression, 445
17.5 Generalized Linear Models, 446

A. Answers and Hints to Selected Problems 449

B. Data Sets and SAS Files 561

Bibliography 563