

Contents

PREFACE	vii
CONTENTS	xi
1. INTRODUCTION	1
1.1. Introduction	1
1.2. Notations	2
1.3. Standard Estimation Results in Least Squares	3
1.4. Assumptions	5
1.5. Iterative Regression Process	6
1.6. Organization of the Book	6
2. PREDICTION MATRIX	9
2.1. Introduction	9
2.2. Roles of P and (I - P) in Linear Regression	10
2.3. Properties of the Prediction Matrix	14
2.3.1. General Properties	14
2.3.2. Omitting (Adding) Variables	20
2.3.3. Omitting (Adding) an Observation	21
2.3.4. Conditions for Large Values of p_{ii}	25
2.3.5. Omitting Multiple Rows of X	27
2.3.6. Eigenvalues of P and (I - P)	28
2.3.7. Distribution of p_{ii}	31
2.4. Examples	34
3. ROLE OF VARIABLES IN A REGRESSION EQUATION	39
3.1. Introduction	39
3.2. Effects of Underfitting	40

3.3. Effects of Overfitting	43
3.4. Interpreting Successive Fitting	48
3.5. Computing Implications for Successive Fitting	49
3.6. Introduction of One Additional Regressor	49
3.7. Comparing Models: Comparison Criteria	50
3.8. Diagnostic Plots for the Effects of Variables	54
3.8.1. Added Variable (Partial Regression) Plots	56
3.8.2. Residual Versus Predictor Plots	56
3.8.3. Component-Plus-Residual (Partial Residual) Plots	57
3.8.4. Augmented Partial Residual Plots	58
3.9. Effects of an Additional Regressor	59

**4. EFFECTS OF AN OBSERVATION ON
A REGRESSION EQUATION 71**

4.1. Introduction	71
4.2. Omission Approach	72
4.2.1. Measures Based on Residuals	72
4.2.1.1 Testing for a Single Outlier	80
4.2.1.2 Graphical Methods	83
4.2.2. Outliers, High-leverage, and Influential Points	94
4.2.3. Measures Based on Remoteness of Points in X-Y Space	99
4.2.3.1. Diagonal Elements of P	99
4.2.3.2. Mahalanobis Distance	102
4.2.3.3. Weighted Squared Standardized Distance	102
4.2.3.4. Diagonal Elements of P_Z	104
4.2.4. Influence Curve	107
4.2.4.1. Definition of the Influence Curve	108
4.2.4.2. Influence Curves for $\hat{\beta}$ and $\hat{\sigma}^2$	109
4.2.4.3. Approximating the Influence Curve	113
4.2.5. Measures Based on the Influence Curve	116
4.2.5.1. Cook's Distance	117
4.2.5.2. Welsch-Kuh's Distance	120
4.2.5.3. Welsch's Distance	123
4.2.5.4. Modified Cook's Distance	124
4.2.6. Measures Based on the Volume of Confidence Ellipsoids	134
4.2.6.1. Andrews-Pregibon Statistic	136
4.2.6.2. Variance Ratio	138
4.2.6.3. Cook-Weisberg Statistic	140
4.2.7. Measures Based on the Likelihood Function	143

4.2.8. Measures Based on a Subset of the Regression Coefficients	148
4.2.8.1. Influence on a Single Regression Coefficient	149
4.2.8.2. Influence on Linear Functions of β	151
4.2.9. Measures based on the Eigenstructure of X	153
4.2.9.1. Condition Number and Collinearity Indices	154
4.2.9.2. Collinearity-Influential Points	158
4.2.9.3. Effects of an Observation on the Condition Number	161
4.2.9.4. Diagnosing Collinearity-Influential Observations	168
4.3. Differentiation Approach	178
4.4. Summary and Concluding Remarks	182
5. ASSESSING THE EFFECTS OF MULTIPLE OBSERVATIONS	185
5.1. Introduction	185
5.2. Measures Based on Residuals	187
5.3. Measures Based on the Influence Curve	190
5.3.1. Sample Influence Curve	191
5.3.2. Empirical Influence Curve	193
5.3.3. Generalized Cook's Distance	194
5.3.4. Generalized Welsch's Distance	195
5.4. Measures Based on Volume of Confidence Ellipsoids	195
5.4.1. Generalized Andrews-Pregibon Statistic	195
5.4.2. Generalized Variance Ratio	197
5.5. Measures Based on the Likelihood Function	198
5.6. Measures Based on a Subset of the Regression Coefficients	200
5.7. Identifying Collinearity-Influential Points	201
5.8. Identifying Influential Observations by Clustering	204
5.9. Example: Demographic Data	207
6. JOINT IMPACT OF A VARIABLE AND AN OBSERVATION	211
6.1. Introduction	211
6.2. Notations	213
6.3. Impact on the Leverage Values	215
6.4. Impact on Residual Sum of Squares	217
6.5. Impact on the Fitted Values	218
6.6. Partial F -Tests	221
6.7. Summary of Diagnostic Measures	226
6.8. Examples	226
6.9. Concluding Remarks	241

7. ASSESSING THE EFFECTS OF ERRORS OF MEASUREMENT	245
7.1. Introduction	245
7.2. Errors in the Response Variable	247
7.3. Errors in X: Asymptotic Approach	248
7.4. Errors in X: Perturbation Approach	251
7.5. Errors in X: Simulation Approach	261
8. STUDY OF MODEL SENSITIVITY BY THE GENERALIZED LINEAR MODEL APPROACH	263
8.1. Introduction	263
8.2. Generalized Linear Models (GLM)	264
8.3. Exponential Family	265
8.4. Link Function	267
8.5. Parameter Estimation for GLM	267
8.6. Judging the Goodness of Fit for GLM	271
8.7. Example	273
9. COMPUTATIONAL CONSIDERATIONS	281
9.1. Introduction	281
9.2. Triangular Decomposition	282
9.2.1. Definitions	282
9.2.2. Algorithm for Computing L and D	284
9.3. Q-R Decomposition	286
9.4. Properties of L and D	288
9.5. Efficient Computing of Regression Diagnostics	293
APPENDIX	297
A.1. Summary of Vector and Matrix Norms	297
A.2. Another Proof of Theorem 4.3	299
A.3. Proof of (4.60a) and (5.31a)	300
REFERENCES	301
INDEX	309