Contents

Preface ix

1 Introduction 1
1.1 Cloud seeding 2
1.2 The basic paradigm 7

2 Diagnostic methods using residuals 10
2.1 The ordinary residuals 10
2.1.1 The hat matrix 11
2.1.2 The role of **V** *in data analyses* 14
2.1.3 Use of the ordinary residuals 15
Example 2.1.1 Illustration of bias 16
2.2 Other types of residuals 17
2.2.1 Studentized residuals 18
Internal Studentization 18
External Studentization 20
2.2.2 Mean shift outlier model 20
Example 2.2.1 Adaptive score data no. 1 22
Example 2.2.2 Cloud seeding no. 2 25
Accuracy of the Bonferroni bound for the outlier test 26
Example 2.2.3 Adaptive score data no. 2 27
Multiple cases 28
Computations 30
Example 2.2.4 Adaptive score data no. 3 31
Other outlier models 32
2.2.3 Predicted residuals 33
2.2.4 Uncorrelated residuals 34
BLUS residuals 35
Recursive residuals 36
2.3 Plotting methods 36
2.3.1 Standard residual plots 37
Example 2.3.1 Cloud seeding no. 3 38
Example 2.3.2 Old Faithful Geyser 40
Example 2.3.3 Residuals in place 43
2.3.2 Added variable plots 44
Non-null behavior 45

Computations 46
Example 2.3.4 Jet fighters no. 1 46
2.3.3 Partial residual plots 49
Example 2.3.5 Jet fighters no. 2 51
2.3.4 Probability plots 52
Example 2.3.6 Cloud seeding no. 4 56
2.4 Transformations 57
2.4.1 Families of transformations 60
2.4.2 Selecting a transformation 61
Invariance 63
Normality 64
Choice of model and scaling the predictors 64
Inference 65
Example 2.4.1 Tree data no. 1 66
2.4.3 Diagnostic methods 69
Atkinson's method 70
Andrews' method 72
Application to the power family 73
Tukey's test 73
Example 2.4.2 Tree data no. 2 74
Example 2.4.3 Jet fighters no. 3 74
2.4.4 Transforming the explanatory variables 78
Example 2.4.4 Tree data no. 3 81
Example 2.4.5 Cloud seeding no. 5 83
2.5 Residual analysis in two-way tables 86
2.5.1 Outliers 86
2.5.2 Nonadditivity 87
Example 2.5.1 Agricultural experiment 93

3 Assessment of influence **101**
3.1 Motivation 102
3.2 The influence curve 104
Example 3.2.1 The sample average 105
3.3 The influence curve in the linear model 106
3.4 Sample versions of the influence curve 108
3.4.1 Empirical influence curves 108
3.4.2 Sample influence curves 109
Example 3.4.1 Cloud seeding no. 6 111
Example 3.4.2 Partial F-tests 111
3.5 Applications of the sample influence curve 113
3.5.1 External scaling 114
$\mathbf{M} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$, $c = p'\hat{\sigma}^2$ 116
Example 3.5.1 Combinations of r_i^2, v_{ii} 118
Example 3.5.2 Rat data 118
Alternative full rank choices for $\mathbf{M}$, c 122
Lower dimensional norms 124
Example 3.5.3 Cloud seeding no. 7 126

Predictions 126
3.5.2 Internal scaling 129
Ordering using a multivariate outlier statistic 129
Jackknife method 131
Example 3.5.4 Cloud seeding no. 8 131
3.5.3 Graphical aids 131
Example 3.5.5 Cloud seeding no. 9 132
Example 3.5.6 Jet fighters no. 4 134
3.6 Multiple cases 135
Other norms 137
3.6.1 Generalized potential 137
Example 3.6.1 $m = 2$ 139
Alternative measures of potential 141
Example 3.6.2 Adaptive score data no. 4 142
3.6.2 Computing D_I 144
Example 3.6.3 Cloud seeding no. 10 144
Example 3.6.4 Drill data 148

4 Alternative approaches to influence **157**
4.1 Volume of confidence ellipsoids 158
4.2 The Andrews and Pregibon diagnostic 160
4.3 Predictive influence 161
4.3.1 Kullback–Leibler divergences and predictive densities with σ^2 known 162
4.3.2 Predictive influence functions, σ^2 known 163
Example 4.3.1 Predictive influence when $q = 1$, $m = 1$, σ^2 known 164
4.3.3 Predictive influence functions, σ^2 unknown 166
An alternative to choosing $\mathbf{X}_I = \mathbf{X}$ 168
4.4 A comparison of influence measures 169
Example 4.4.1 Cloud seeding no. 11 173
Calibration 173

5 Assessment of influence in other problems **176**
5.1 A general definition of residuals 177
Example 5.1.1 Leukemia data no. 1 178
5.2 A general approach to influence 182
Example 5.2.1 Leukemia data no. 2 185
5.3 Nonlinear least squares 186
Example 5.3.1 Duncan's data 188
5.4 Logistic regression and generalized linear models 191
Example 5.4.1 Leukemia data no. 3 193
Example 5.4.2 Leukemia data no. 4 199
5.5 Robust regression 200
Example 5.5.1 Cloud seeding no. 12 201
Example 5.5.2 Cloud seeding no. 13 204

5.6 Other problems 204
5.6.1 Correlation coefficient 205
5.6.2 Discriminant analysis 206
5.6.3 Linear regression with incomplete data 207

Appendix **209**
A.1 Weighted least squares 209
A.2 Updating formulae 210
A.3 Residual correlations 211

Bibliography **214**

Author and Subject indexes **225**