INDEX

Added-variable plot, 95, 109
Adjusted multiple correlation coefficient, 285
AIC, 287, 290–291, 295, 308, 320, 327, 343
  bias corrected, 287
Akaike Information Criteria, 287
Analysis of covariance, 12, 15
Analysis of variance, 12, 15, 67, 140
ANOVA, 140
  multiple regression, 67
  simple regression, 71
Assumption
  constant variance, 86
  Homogeneity, 86
  independent-errors, 87
  linearity, 31, 86
  normality, 37, 86
Autocorrelation, 18, 87, 180, 197, 201

B

Backward elimination, 289
Bayes Information Criteria, 287
Best linear unbiased estimator, 61, 83
BIC, 287, 290–291, 295, 308, 320, 327
Binary
  response data, 195
  response variable, 318, 321
  variable, 12
BLUE, 61, 83

C

Centering, 239
Classification, 336
Coefficient of determination, 42, 61
Collinearity-influential observations, 246
Collinearity, 18, 88, 222
Computer repair data, 26, 30–31, 35, 42
Concordance Index, 328
Condition number, 244, 276, 288
Confidence
  interval, 38
  interval for β_j, 63
  region, 37
Constant variance assumption, 86
Cook's Distance, 103
Corrected sums of squares and products matrix, 60
Correlated errors, 141
Correlation coefficient, 21, 24, 240
  between Y and $\hat{Y}$, 40, 61
  population, 36
  sample, 36, 39
Correlation matrix, 117, 241, 243, 250, 252
Cov$(\hat{\beta}_i\, \hat{\beta}_j)$, 83
Covariance, 21, 23
Cross-sectional data, 128, 214–215

Regression Analysis by Example, Fourth Edition. By Samprit Chatterjee and Ali S. Hadi
Copyright © 2006 John Wiley & Sons, Inc.

Cumulative distribution function, 319

D

Data set
  advertising, 234, 236
  air pollution, 300, 302
  Anscombe's quartet, 24
  bacteria deaths, 155
  bankruptcy, 320, 322
  brain, 171–172
  cigarette consumption, 1, 79–80, 312
  college expense, 182
  computer repair (expanded), 116
  computer repair, 26–27
  consumer expenditure, 198, 216
  corn yields, 147
  cost of education, 182
  diabetes, 330, 332
  DJIA, 216
  domestic immigration, 6
  education expenditure, 141, 143, 185
  Egyptian skulls, 6
  equal educational opportunity, 223
  examination, 75–76
  exponential growth, 108
  field goal kicking, 339
  fuel consumption, 256
  Hald, 265–266
  Hamilton, 95
  hard disk prices, 177
  heights of husbands and wives, 47–48, 78
  homicide, 296
  housing starts, 206
  import, 229, 260
  injury incidents, 161
  labor force, 47
  Longley, 275–276
  magazine advertising, 174
  milk production, 3
  new drugs, 343
  New York rivers, 10, 6, 99, 101, 105
  newspapers, 48
  nonlinear, 24
  oil production, 176, 216
  PCB, 348
  preemployment testing, 132
  presidential election, 148–150, 176, 255, 312
  quantitative sensory testing, 14
  real estate, 309
  right-to-work laws, 4
  salary survey, 122–123
  Scottish Hills Races, 111
  ski sales, 140–141, 210–211
  Space Shuttle, 339
  students information, 147–148
  supervised workers, 164, 180
  supervisor performance, 54–55, 63–64, 292
  wind chill factor, 175
Data
  collection, 11
  re-expression, 13
  transformation, 13
Degrees of freedom, 33, 58, 89
DFITS, 104
Discriminant analysis, 336
Distance
  Cook's, 103
  orthogonal, 29
  perpendicular, 29
  vertical, 29
Distribution
  binomial, 342
  chi-square, 323
  gamma, 342
  inverse Gaussian, 342
  Poisson, 160, 164, 342
Dose-response curve, 180, 194
Draftsman's matrix, 94
Dummy variable, 121
Durbin-Watson statistic, 200

E

Eigenvalues, 243
Eigenvector, 243
Externally studentized residual, 90

F

F-test, 65
Fitted
  regression equation, 37
  value, 15, 30, 57
Fitting
  models to data, 14
Forecast interval, 38
Forward selection, 289
Full model, 65
Function
  cumulative distribution, 319
  intrinsically nonlinear, 13
  linear, 13
  linearizable, 13
  link, 343
  logistic, 195
  nonlinear, 13, 195

G

Generalized linear models, xv, 341
Goodness-of-fit index, 42

H

Hadi's influence measure, 105
Hat matrix, 83, 89

Heteroscedasticity, 86, 159, 179
High leverage, 100, 265
Homogeneity (Homoscedasticity), 86, 159

I

Independent-errors assumption, 87
Indicator variable, 118, 121, 211
Influence measures, 103
  Welsch and Kuh (DFITS), 106
  Cook's Distance, 103, 106, 117
  Hadi's, 105–106, 117
  Welsch and Kuh (DFITS), 104, 106
Influential observations, 91, 98
Interaction effects, 125
Intercept, 57
Internally studentized residual, 90
Interpretation of regression coefficients, 58
Iterative process, 16

L

L-R plot, 107
Ladder of transformation, 171
Lagged variables, 215
Least squares, 14
  estimates, 30
  line, 30
  method, 29–30, 57
  properties, 60
  weighted, 18, 167, 180
Leverage-residual plot, 107
Leverage values, 89
Linearity assumption, 31, 86
Link function, 343
Logistic regression, xv, 342
  binary, xv, 18
  multinomial, xv, 18
  polytomous, 329
  ordered response variable, xv, 18
Logistic
  distribution, 319
  function, 176, 195
  models, 180, 195
  regression, 12, 15, 18, 140, 317, 319
  regression diagnostics, 323
  response function, 318
Logit, 320
  link function, 342
  model, xv, 195, 343
  transformation, 319

M

Masking problem, 101
Matrix, 241
  corrected sums of squares and products, 60
  correlation, 241, 243, 250, 252
  draftsman's, 94
  hat, 89
  plot, 94, 117
  projection, 89
  variance-covariance, 241, 257
Maximum likelihood, 14, 320, 342
Maximum likelihood method, 335, 330
Model
  fitting, 14
  full, 65
  no-intercept, 183, 186
  probit, 195, 319
  random walk, 219
  reduced, 65
  through the origin, 42
Multicollinearity, 68, 222
Multinomial logit model, 329
Multiple correlation coefficient, 61, 67, 70, 87, 285
Multiple regression, 15, 54, 61
  ANOVA table, 67
  assumptions, 60
Multiplicative effect, 125
Multivariate regression, 14–15

N

Nested models, 65, 326
No-intercept model, 42, 183, 186
Nominal model, 335
Nonlinear regression, 15
Nonparametric statistics, 200
Normal
  equations, 57, 82
  property of, 160
  scores, 97
Normality assumption, 33, 86
Normalizing transformations, 153

O

Odds ratio, 319
One-sample t-test, 44
Ordered response category, 334
Ordered response variable, xv, 18
Ordinal logistic regression, 334
Ordinal model, 335
Orthogonal
  predictors, 221, 239
  regression, 29
Outliers, 100, 265

P

P-R plot, 107
P-value for
  F-test, 66
  t-test, 63

P-value
for t-test, 34
Parameter estimation, 14
Parsimony, 69
Partial
regression coefficients, 53, 58
regression plot, 110
residual plot, 110
Plot
added-variable, 95, 109
L-R, 107
matrix, 94, 117
P-R, 107
partial regression, 110
partial residual, 110
potential-residual, 107, 117
residual plus component, 95
sequence, 199
Poisson regression, xv, 18, 342
Potential-residual plot, 107, 117
Predicted value, 15
Prediction
errors, 219
interval, 38
Principal components, 14, 239, 243, 255, 257
Probit model, 195, 319
Projection matrix, 83, 89
Proportional odds model, 329, 334–335
Pure error, 184

R

R^2, 41, 61
Random walk model, 219
Reduced model, 65
Regression
assumptions, 86
coefficients, 2, 28, 53
definition, 1
elements of, 7
examples of, 3
line, 30
linear, 15
logistic, 12, 15, 18
model through the origin, 42
multiple, 13, 15, 17, 54, 61
multivariate, 14–15
nonlinear, 15
parameters, 2, 28
partial coefficients, 53
Poisson, xv, 18, 342
ridge, 268
robust, 115
simple, 13, 15, 17, 54, 63, 71, 74
sum of squares, 40
trivial models, 44
univariate, 14–15
Relationship between simple and multiple regression coefficients, 76
Residual
adjusted, 105
internally studentized, 90
mean square, 44, 285
ordinary least squares, 30, 57
plots, 90
plus component plot, 95, 109–110
standardized, 89
sum of squares, 33, 40, 89
Ridge
bias of estimators, 278
method, 14
parameter, 268
regression, 268
trace, 268–269
variance of estimators, 278
Robust regression, xv, 115, 341, 345

S

Sample
mean, 22
standard deviation, 44
variance, 44
Sampling distribution of
$\hat{\beta}_0$, 37
$\hat{\beta}_1$, 33, 37
Scaling, 239
unit length, 240
Sequence plot, 199
Shrinkage estimators, 269
Simple regression, 15, 54, 63, 71, 74
ANOVA table, 71
Simultaneous confidence region, 37
Standard deviation, 33
sample, 23, 44
Standard error of
$\hat{\beta}$, 83
estimate, 33
Standard normal distribution, 90
Standardized
deviance residuals, 323
Personian residuals, 323
residual, 89
variables, 240
Standardizing, 240
Stepwise method, 289
Stimulus-response relationships, 195
Sum of squared residuals, 58
Swamping problem, 101

T

Test
t-, 44

Durbin-Watson, 200
F-, 65–68, 70
one-sample t, 44
runs, 200
t-, 33–34, 63–64, 67, 72
two-sample t, 45
Time series data, 130, 199, 214
Total
mean square error, 277–279
sum of squares, 40
variance, 268
variance of OLS estimators, 269
variance of ridge estimators, 268
Transformation, 13, 18, 32
ladder of, 171
to achieve linearity, 153
to achieve normality, 153
variance-stabilizing, 153
Trivial regression models, 44
Two-sample t-test, 45

U

Uniqueness of LS solution, 30, 57
Univariate regression, 14–15

V

Variable
binary, 12, 318, 321
dependent, 1
dummy, 12, 121
explanatory, 1
independent, 2
indicator, 12, 118, 121, 211
lagged, 215
predictor, 1
qualitative, 12, 18
quantitative, 12
regressor, 2
response, 1
role of, 313
selection, 18
selection procedures, 281
standardized, 240
Variance-covariance matrix, 241, 257
Variance-stabilizing transformations, 153
Variance
inflation factor, 236, 278, 288
of ridge estimators, 278
sample, 44

W

Wald Test, 321
Web site
book's, 3, xiv
Case book, 2
DASL, 3
Electronic Dataset Service, 3
Weighted
least squares, 18, 167, 180
leverages, 323
Welsch and Kuh measure, 104