INDEX

Added-variable plot, 103, 117–118, 327
Adjusted multiple correlation coefficient, 303
Akaike Information Criterion (AIC), 305, 308–309, 313, 327, 338, 345, 361
bias corrected, 305
Analysis of covariance, 16, 18
Analysis of variance, 16, 18, 74, 148
ANOVA, 148
multiple regression, 74
simple regression, 77
Anscombe quartet, 28, 99, 106
Assumption
constant variance, 94
homogeneity, 94
independent-errors, 95
linearity, 35, 94
normality, 41, 94
Auto-correlation, 22, 95, 192, 209, 213, 327
response variable, 336, 341
variable, 16, 336, 341
BLUE, 67, 90
C
Centering, 64
Classification, 354
Coefficient of determination, 46, 68
Collinearity-influential observations, 262
Collinearity, 22, 75, 96, 234
Computer repair data, 30, 34–35, 39, 46
Concordance Index, 346
Condition indices, 248, 251–252
Condition number, 253, 289, 306
Confidence
interval, 42
interval for $\beta_j$, 70
region, 41
Constant variance assumption, 94
Cook’s distance, 111
Corrected sums of squares and products matrix, 67
Correlated errors, 149
Correlation coefficient, 25, 28, 65
between $Y$ and $\hat{Y}$, 44, 68
population, 40
sample, 40, 43
Correlation matrix, 125, 252, 263–264
Cov(β_i, β_j), 90
Covariance, 25, 27
Cross-sectional data, 137, 226–227
Cumulative distribution function, 337

D

DASL Website, 2, 9
Data
collection, 15
reexpression, 17
transformation, 17
Dataset
advertising, 246, 248, 262
air pollution, 318–319
Anscombe quartet, 28, 106
bacteria deaths, 167
bankruptcy, 338–339
brain, 183–184
cigarette consumption, 1.86–87, 331
college expense, 194
collinear, 291
computer repair (expanded), 124
computer repair, 30–31
consumer expenditure, 210, 228
corn yields, 157
cost of education, 194
cost of health care, 12–13
diabetes, 348, 350
DJIA, 228
domestic immigration, 6
education expenditure, 150, 154, 197
Egyptian skulls, 9
equal educational opportunity, 235
examination, 82–83
exponential growth, 116
field goal kicking, 357
fuel consumption, 256
Hald, 277–278
Hamilton, 103
hard disk prices, 189
heights of husbands and wives, 51, 85
homicide, 314
housing starts, 218
import, 241, 254, 272
Industrial Production, 9
injury incidents, 173
labor force, 51
Longley, 289–290
magazine advertising, 187
milk production, 3
new drugs, 361
New York Rivers, 9, 107, 109, 113
newspapers, 53
nonlinear, 28
oil production, 189, 228
PCB, 366
preemployment testing, 140
presidential election, 158, 160, 188, 257, 331
quantitative sensory testing, 18
real estate, 328
right-to-work laws, 4
salary survey, 130–131
Scottish hills races, 118–119
ski sales, 148–149, 222–223
Space Shuttle, 12, 356
students information, 157, 159
supervised workers, 176, 192
supervisor performance, 58–59, 70, 310
wind chill factor, 187
Degrees of freedom, 37, 61, 97
DHTS, 112
Discriminant analysis, 354
Distance
Cook’s, 111
orthogonal, 33
perpendicular, 33
vertical, 33
Distribution
binomial, 360
chi-square, 341
gamma, 360
inverse Gaussian, 360
normal, 172
Poisson, 172, 176, 360
Dose-response curve, 192, 206
Draftman’s matrix, 102
Dummy variable, 129
Durbin-Watson statistic, 212

E

Eigenvalues, 252–253
Eigenvector, 253
Externally studentized residual, 98

F

F-Test, 72
Factors, 2
Fitted regression equation, 41
Fitted value, 19, 34, 61
Fitting models to data, 19
Forecast interval, 42
Forward selection, 307
Full model, 71
Function
cumulative distribution, 337
intrinsically nonlinear, 17
linear, 16
linearizable, 17
link, 361
logistic, 207
nonlinear, 16, 206

G
Generalized linear models, 359
Goodness-of-fit index, 46

H
Hadi’s influence measure, 113
Hat matrix, 90, 97
Heteroscedasticity, 94, 171, 191
High leverage, 108, 277
Homogeneity (Heteroscedasticity), 94, 171

I
Independent-errors assumption, 95
Indicator variable, 126, 129, 223
Influence measures, 111
Cook’s distance, 111, 114, 125
Hadi’s, 113–114, 125
Welsh and Kuh (DFIT’s), 112–114
Influential observations, 99, 106
Interaction effects, 133
Intercept, 61
Internally studentized residual, 98
Interpretation of regression coefficients, 62
Iterative process, 20

L
L-R plot, 115
Ladder of transformation, 183
Lagged variables, 227
Least absolute value, 55
Least squares, 19
estimates, 34
line, 34
method, 33–34, 61
properties, 67
weighted, 22, 179, 192
Leverage-residual plot, 115
Leverage values, 97
Linearity assumption, 35, 94
Link function, 361
Logistic regression, 360
binary, 22
multinomial, 22
ordered response variable, 23
polytomous, 347
Logistic
distribution, 337
function, 189, 207
models, 192, 207
regression, 16, 18, 22, 148, 335, 337
regression diagnostics, 341
response function, 336
Logit, 337
link function, 360
model, 207, 361
transformation, 337

M
Masking problem, 109
Matrix
corrected sums of squares and products, 67
correlation, 252, 263–264
draftsman’s, 102
hat, 97
plot, 102, 125
projection, 97
variance-covariance, 293
Maximum likelihood, 19, 338, 360
Maximum likelihood method, 353, 347
Model
fitting, 19
full, 71
no-intercept, 195, 199
probit, 207, 337
random walk, 231
reduced, 71
through the origin, 46
Multicollinearity, 234
Multinomial logit model, 347
Multiple correlation coefficient, 68, 73, 77, 95,
303
Multiple regression, 18, 58, 68
ANOVA table, 74
assumptions, 67
Multiplicative effect, 133
Multivariate regression, 18

N
Nested models, 71, 344
No-intercept model, 46, 195, 199
Nominal model, 353
Nonlinear regression, 18
Nonparametric statistics, 212
Normal
distribution, 172
equations, 61, 89
scores, 105
Normality assumption, 37, 94
Normalizing transformations, 165

O
Odds ratio, 337
One-sample t-Test, 48
Ordered response category, 352
Ordered response variable, 23
Ordinal logistic regression, 352
Ordinal model, 353
Orthogonal predictors, 233
Orthogonal regression, 33, 55
Outliers, 108, 277

P

P-R plot, 115
P-value for
  F-Test, 73
  t-Test, 38, 69
Parameter estimation, 19
Parsimony, 75
Partial
  regression coefficients, 57, 62
  regression plot, 118
  residual plot, 118
Plot
  added-variable, 103, 117–118, 327
  L-R, 115
  matrix, 102, 125
  P-R, 115
  partial regression, 118
  partial residual, 118
  potential residual, 115, 125
  residual plus component, 103, 117–118, 327
  sequence, 211
Poisson regression, 23, 360
Potential residual plot, 115, 125
Predicted value, 19
Prediction errors, 231
Prediction interval, 42
Principal components, 19, 259–260, 292–293
Probit model, 207, 337
Projection matrix, 90, 97
Proportional odds model, 347, 352–353
Pure error, 196

R

R², 45, 68
Random walk model, 231
Reduced model, 71
Regression
  assumptions, 94
  coefficients, 2, 32, 57
  definition, 1
  elements of, 13
  examples of, 3
  line, 34
  linear, 18
  logistic, 16, 18, 22
  model through the origin, 46
  multiple, 17–18, 22, 58, 68
  multivariate, 18
  nonlinear, 18
  orthogonal, 33, 55
  parameters, 2, 32
  partial coefficients, 57
  Poisson, 23, 360
  ridge, 279

  robust, 123
  simple, 17–18, 22, 58, 70, 77, 81
  sum of squares, 44
  surrogate ridge, 288, 296
  trivial models, 48
  univariate, 18

Relationship between simple and multiple
  regression coefficients, 83

Residual
  adjusted, 113
  internally studentized, 98
  mean square, 48, 303
  ordinary least squares, 34, 61
  plus component plot, 103, 117–118, 327
  standardized, 97
  sum of squares, 37, 44, 97

Ridge
  bias of estimators, 295
  method, 19
  parameter, 280
  regression, 279
  trace, 279, 281
  variance of estimators, 295
Robust regression, 123, 359, 363
Runs test, 212

S

Sample standard deviation, 27
Sample
  mean, 26
  standard deviation, 48
  variance, 49

Sampling distribution of
  β₀, 41
  β₁, 37, 41

Scaling, 64
  unit-length, 65
Sequence plot, 211
Shrinkage estimators, 281
Simple regression, 18, 58, 70, 77, 81
  ANOVA table, 77
Simultaneous confidence region, 41
Standard error of estimate, 37, 90
Standard normal distribution, 98
Standardized
  deviance residuals, 341
  Person distribution, 341
  residual, 97
  variables, 66
Standardizing, 65
Stepwise method, 307
Stimulus-response relationships, 206
Sum of squared residuals, 61
Surrogate ridge regression, 288, 296
Swamping problem, 109

T
Test
  F-, 72–75, 77
t-, 37–38, 48, 69–70, 74, 79
durbin–watson, 212
  one-sample, 48
  runs, 212
  two-sample t, 49
Time series data, 138, 211, 226
Total
  mean square error, 294–296
  sum of squares, 44
  variance of OLS estimators, 281
  variance of ridge estimators, 280
Transformation, 17, 22, 36
  ladder of, 183
  to achieve linearity, 165
  to achieve normality, 165
  variance-stabilizing, 165
Trivial regression models, 48
Two-sample t-test, 49

U
Uniqueness of LS solution, 34, 61
Unit-length scaling, 65
Univariate regression, 18

V
Variable
  binary, 16, 336, 341
  dependent, 1
  dummy, 16, 129
  explanatory, 1
  independent, 2
  indicator, 16, 126, 129, 223
  lagged, 227
  predictor, 1
  qualitative, 15, 22
  quantitative, 15
  regressor, 2
  response, 1
  role of, 331
  selection, 22
  selection procedures, 299
  standardized, 66
Variance-covariance matrix, 293
Variance-stabilizing transformations, 165
Variance inflation factor, 248, 295, 306
Variance inflation factors, 248

W
Wald test, 340
Website
  book’s, 3
  case book, 2
  DASL, 3
  electronic dataset service, 3
Weighted
  least squares, 22, 179, 192
  leverages, 341
  welsch and kuh measure, 112–113