

INDEX

SYMBOLS

Δ (delta), 29

' (prime), 32

A

accuracy. *See also* coefficient of determination
of logistic regression
analysis equation, 173–177
of multiple regression
equation, 119–126
adding matrices, 39–40
adjusted odds ratio,
192–194
adjusted R^2 , 124–126
alternative hypothesis
(H_a), 48
analysis of variance
(ANOVA). *See also*
hypothesis testing
logistic regression
analysis, 178–181
multiple regression
analysis, 128–132
regression analysis,
87–90
apparent error rate, 177
assumptions of normality,
85–86
autocorrelation, checking
for, 102–103
average, 72

B

bell curves, 53–54
best subsets regression,
139–140
binomial logistic regression
analysis. *See* logistic
regression analysis

C

calculus, differential. *See*
differential calculus
Canceling Exponentials
Rule, 22
categorical data, 46
converting numerical
data, 46–47
in logistic regression
analysis, 167
in multiple regression
analysis, 147–149
chi-squared (χ^2) distri-
butions, 54–55, 56,
204–206
coefficient of determi-
nation (R^2)
adjusted, 124–126
logistic regression
analysis, 173–177
multiple regression
analysis, 119–126
regression analysis,
81–82

coefficients. *See specific
coefficients by name*
columns, in matrices, 38
concentration matrix, 145
confidence coefficient,
92–93
confidence intervals,
calculating
multiple regression
analysis,
133–135, 146
for odds ratio, 194–195
regression analysis,
91–94
correlation coeffi-
cient (R), 70
general discussion,
64–65
multiple regression
analysis, 120
regression analysis,
78–82
critical value, 55

D

data. *See also*
categorical data
plotting, 64–65
types of, 46–47
degrees of freedom, 50–51
delta (Δ), 29
dependent variables, 14,
67, 149–152. *See also*
scatter plots

- differential calculus
 - differentiating, 31–36
 - general discussion, 24–30
- Durbin-Watson statistic, 102–103

- E**
- elements, in matrices, 38
- Euler's number, 19, 198–199
- event space, 53
- Excel functions, 198
- Exponentiation Rule, 22
- exponents, 19–23, 200
- extrapolation, 102

- F**
- F* distributions, 57–59, 206–209
- freedom, degrees of, 50–51
- F*-test, 129–133
- functions. *See also*
 - probability density functions
 - exponential, 19–23
 - inverse, 14–18
 - likelihood, 161–163, 171
 - logarithmic, 19–23
 - log-likelihood, 161–163, 171–172
 - natural logarithm, 20, 200–201

- G**
- graphs. *See also*
 - scatter plots
 - for inverse functions, 17–18
 - logistic regression analysis equation, 159

- H**
- H_0 (null hypothesis), 48
- H_a (alternative hypothesis), 48

- hypothesis testing, 85–90
 - logistic regression analysis, 178–181
 - multiple regression analysis, 128–132
 - with odds, 194

- I**
- identity matrices, 44
- independent variables, 14, 67
 - choosing best combination of, 138–140
 - determining influence on outcome variables, 149–152
 - logistic regression analysis, 164–167
 - multicollinearity, 149
 - structural equation modeling, 152
- interpolation, 102
- inverse functions, 14–18
- inverse matrices, 44, 202–204

- L**
- likelihood function, 161–163, 171
- likelihood ratio test, 179
- linear equations, turning nonlinear equations into, 104–106
- linear least squares regression, 71–76, 115
- linear regression analysis, 7
- linearly independent data, 47
- logarithms, 19–23
- logistic regression analysis, 8, 157
 - accuracy of equation, assessing, 173–177
 - adjusted odds ratio, 192–194

- confidence intervals
 - for odds ratios, 194–195
- equation for, calculating, 158–159, 170–173
- hypothesis testing, 178–181, 194
- maximum likelihood method, 159–163, 210–212
- odds ratio, 192–194
- predicting with, 182
- predictor variables, choosing, 164–167
- procedure, general discussion of, 168, 190
- relative risk, 195–196
- scatter plot, drawing, 169
- logit, 190–191
- log-likelihood function, 161–163, 171–172

- M**
- Mahalanobis distance, 133, 137, 144–146
- matrices
 - adding, 39–40
 - general discussion, 37–38
 - identity, 44
 - inverse, 44, 202–204
 - multiplying, 40–43, 201–202
 - prediction intervals, calculating, 144–146
- maximum likelihood estimate, 162–163
- mean, 49
- median, 49
- multicollinearity, 149
- multiple correlation coefficient
 - accuracy of multiple regression equation, 119–121
 - adjusted, 124–126
 - problems with, 122–123

multiple regression
 analysis, 7–8, 111
 accuracy of equation,
 assessing, 119–126
 analysis of variance,
 128–132
 categorical data, using
 in, 147–149
 confidence intervals,
 calculating,
 133–135
 equation for, calculating,
 115–119
 hypothesis testing, 127
 Mahalanobis distance,
 144–146
 multicollinearity, 149
 prediction intervals,
 calculating,
 136–137
 predictor variables,
 choosing, 138–140
 predictor variables,
 determining influ-
 ence on outcome
 variables, 149–152
 procedure, general dis-
 cussion of, 112, 142
 scatter plot, drawing,
 113–114
 standardized residuals,
 143–144
 multiplying matrices,
 40–43, 201–202

N

natural logarithm function,
 20, 200–201
 nonlinear regression,
 103–106
 normal distributions, 53–54
 normality, assumptions of,
 85–86
 null hypothesis (H_0), 48
 numerical data, 46–47

O

odds, 190
 hypothesis testing, 194
 logit, 190–191
 odds ratio (OR)
 adjusted, 192–194
 confidence intervals,
 calculating,
 194–195
 general discussion,
 191–192
 outcome variables, 14, 67,
 149–152
 outliers, 101, 144
 overfitting, 149

P

partial regression
 coefficients
 calculating with Excel,
 209–210
 general discussion,
 116–118
 hypothesis testing, 127,
 129–131
 Pearson product moment
 correlation coefficient,
 79. *See also* correla-
 tion coefficient
 plotting data, 64–65
 population mean, 91
 population regression, 86
 populations

assessing, 82–84
 confidence intervals,
 calculating,
 133–135

Power Rule, 21

predictions
 logistic regression
 analysis, 182
 multiple regression
 analysis, 136–137
 regression analysis,
 95–98

predictor variables, 14, 67
 choosing best combina-
 tion of, 138–140
 determining influence on
 outcome variables,
 149–152
 logistic regression
 analysis, 164–167
 multicollinearity, 149
 structural equation
 modeling, 152
 prime ($'$), 32
 probability density
 functions
 chi-squared distribution,
 54–55, 56, 204–206
 F distributions, 57–59,
 206–209
 general discussion,
 52–53
 normal distribution,
 53–54
 tables, 55–56
 Product Rule, 23
 pseudo- R^2 , 173–177

Q

qualitative data, 46
 quantitative data, 46
 Quotient Rule, 21

R

R (correlation coeffi-
 cient), 70
 general discussion,
 64–65
 multiple regression
 analysis, 120
 regression analysis,
 78–82

R^2 (coefficient of
 determination)
 adjusted, 124–126
 logistic regression
 analysis, 173–177

- R^2 (coefficient of determination), *continued*
 multiple regression analysis, 119–126
 regression analysis, 81–82
- regression analysis
 analysis of variance, 87–90
 assumptions of normality, 85–86
 autocorrelation, checking for, 102–103
 confidence intervals, calculating, 91–94
 correlation coefficient, calculating, 78–82
 equation, calculating, 71–77
 equation, general discussion, 66–67
 interpolation and extrapolation, 102
 nonlinear regression, 103–104
 prediction intervals, calculating, 95–98
 procedure, general discussion of, 68, 100
 samples and populations, 82–84
 scatter plot, drawing, 69–70
 standardized residual, 100–101
- regression diagnostics, 119–121
- regression equation
 calculating, 71–77
 general discussion, 66–67
 linear equations, turning nonlinear into, 104–106
- relative risk (RR), 195–196
- residual sum of squares, 73–74
- residuals, 71
 standardized, 100–101, 143–144
- round-robin method, 139–140
- rows, in matrices, 38
- RR (relative risk), 195–196
- §
- sample regression, 86
- sample variance, unbiased, 50
- samples, 82–84
- scatter plots
 differential calculus, 26
 for logistic regression analysis, 169
 for multiple regression analysis, 113–114
 plotting data, 64–65
 for regression analysis, 69–70
- SEM (structural equation modeling), 152
- squared deviations, sum of, 50
- standard deviation, 51–52
- standardized residuals, 100–101, 143–144
- statistically significant, 58
- statistics
 data types, 46–47
 hypothesis testing, 48
 variation, measuring, 49–52
- structural equation modeling (SEM), 152
- subsets regression, best, 139–140
- sum of squared deviations, 50
- T
- testing hypotheses, 85–90
 logistic regression analysis, 178–181
 multiple regression analysis, 128–132
 with odds, 194
- tolerance, 149
- U
- unbiased sample variance, 50
- V
- variables. *See* dependent variables; independent variables; scatter plots
- variance, 50–51
- variance, analysis of
 logistic regression analysis, 178–180
 multiple regression analysis, 128–132
 regression analysis, 87–90
- variance inflation factor (VIF), 149
- W
- Wald test, 180
- X
- x-bar, 72
- Y
- y-hat, 73