

Index

- C_p statistic, 108
- R^2 , 77, 80, 108
- R^2 , adjusted, 108
- R^2 , partial, 105
- .05 p-value, 32
- 2-AAF, 182
- 5-FU, 235

- actuary, 237
- adjusted R^2 , 108
- age, period, cohort model, 222
- AIDS, 226
- airline “bumping” of passengers, 215, 217
- airline mergers, 127
- airlines, 84, 85*f*
- alcohol, 68
- alternative hypothesis, 30
- alternative hypothesis, one-sided, 31
- alternative hypothesis, two-sided, 31
- analysis of variance (ANOVA), 78, 93
- analysis of variance, total sum of squares, 79
- answers to selected problems, 263
- array statement in SAS, 148
- arsenic, 69, 85, 110
- attacks in Iraq, 208
- attendance at baseball game, 67, 110, 139
- auditions for orchestra, 198
- autocorrelation, 209
- Avadex, 42, 42*t*
- average, 2, 3

- backward selection, 105
- Baltimore Orioles, 67, 110, 139
- baseline hazard, 239
- Bayesian statistics, 26
- beetle mortality, 170*t*, 174*t*, 183
- biased coin, 21
- biased sample, 2
- Big Bang, 25
- bile duct, 235
- biliary cirrhosis, 244
- binomial coefficient, 22
- binomial distribution, 21, 23*f*, 205, 215

- blinded experiment, 56
- Bonferroni correction, 34, 107
- boxplot, 37, 165, 168
- brain cancer, 13
- breast cancer, 73
- bubble plot, 95

- calculus, viii, 7, 27, 62
- cancer, 42, 142, 229
- cancer deaths, 222
- cancer incidence, 204
- cancer of the bile duct, 235
- cancer rates in Japan, 221
- cars, characteristics in 1974, 114
- cell phones, 13
- censored data, 116
- censoring, 225
- census, 2, 179
- centenarian survival, 235
- CGD, chronic granulomatous disease, 54
- chi-squared distribution, 257, 258*t*
- chi-squared residual, *reschi*, 190, 213
- chi-squared test, 42, 168
- China buys U.S. debt, 16
- choose symbol, 22
- cigarettes, 140
- cirrhosis, 244
- class statement in SAS, 186
- clinical trial, 179, 229
- cloth waste, 165
- Cody and Smith, 49
- cohort, 222
- coin toss, 21
- complementary log-log link, 173
- confidence interval, 25, 40
- confidence interval displacement, 194
- confidence interval for binomial p parameter, 24
- conjunctivitis, 14
- Constitution, U.S., 2
- continuity adjusted chi-squared, 46, 152
- continuous events, 27
- convenience sample, 2
- Cook’s D, 101, 144

270 **Index**

- coronary bypass surgery, 112
 correlation, 63, 75
 correlation of the global economy, 77
 covariance ratio, 101
 Cox regression, 241
 cracker diet, 167
 cross product ratio, 43
 crossover experiment, 53
 crystal ball, 58

 dark energy, 25
 DASL, ix
 data mining, 109
 datalines statement in SAS, 51
 degrees of freedom, df, 35, 47
 Delwiche and Slaughter, 49
 demographic-economic paradox, 145
 dependent variable, 52
 derivative, viii, 8, 62
 dermatologists' fees, 71
 descending option in `proc logistic`, 186, 197
 deviance, 191, 210
 deviance residual, 242
 deviance residuals, 192
 df, 35, 47
 dfbeta, 101, 138
 dffits, 100, 144
 diastolic blood pressure, exercise, 163
 discrete events, 27
 DNase in rats, 147
 double blind experiment, 57
 drought in United States, 67
 dummy variable, 122, 186

 e , e^x , $\exp(x)$, 7
 e , base of logarithm, viii
 ED₅₀, 182
 Efron, B, 181
 El Salvador, 181
 ELISA, 147
 empirical data, 2
 equipoise, 235
 error distribution, 82
 error sum of squares, 79
 errors and residuals, 82
 exact test, 46, 49, 168
 Excel, 51
 explained sum of squares, 80
 exponential function, 7
 extrapolation, 16, 66, 236
 extrapolation, perils of, 65, 183

 F-ratio, 81
 factorial, 22
 failure rate, *also see* hazard function, 237
 fair coin, 21
 Farmer, 183
 fast food, 14
 female managers, 87
 fertility rates, 145
 Fisher exact test, 49, 152
 food imports, 86
 football, 56, 184, 201
 football, yards rushing, 220

 forward selection, 105
 friend, kind, patient, and understanding, 49, 51
 full rank, 124
 fungicide, 42
 fusion experiment, 35, 54, 163

 G-CSF, 179
 Galápagos Islands, 219
 gasoline price, 16, 132
 Gaussian distribution, 26
 global stock markets, 77
 glove use, 200
 Gosset, WS, 41
 graphical displays, 94
 gross domestic product, 17
 Guinness Brewery, 41

 halibut, 243, 243*t*
 Hamburger Helper lasagna, 121*t*
 hat diagonal, 194
 hat matrix diagonal, 96, 123, 139
 hazard function, 236, 237
 health-care costs, 112
 healthiest state, 157
 helium-filled football, 56
 Hessian weight, 213
 high leverage, 96
 high-dose chemotherapy, 179
 HIPPA, 248
 HIV, 12, 226
 housing, 9
 Hubble constant, 25
 hypothesis testing, 30

 ice cream flavors, 150
 if, then, else statements, 122
 IMF, International Monetary Fund, 11
 independence in a 2×2 table, 42
 index plots, `iplots`, 188, 190
 indicator variable, 120, 122, 186
 indicator variable as a t-test, 122
 influence, 99, 123
 informative censoring, 227
 integral of a function, viii, 9
 interactions, 128
 intercept, 7, 16, 64, 125
 intercept, zero, 71
 interval censoring, 226
 Iraq, 208

 jackknife, 99, 195
 January temperatures in U.S. cities, 91*t*, 111
 Japanese cancer rates, 221
 jittering, 39*f*, 39, 134, 142, 165, 242

 kangaroo flush, 53
 Kaplan-Meier curve, 227

 lackfit option, 189
 lambda (λ), 206
 last value carried forward, 138, 201
 LD₀₁, 182

271 Index

- LD₅₀, 249, 266
 least squares, 62, 93, 211
 left censoring, 226
 leverage, 84, 96, 160, 218, 251
 lightning strikes, 204
 likelihood ratio, 192
 linear functions, 7
 linear predictor *xbeta*, 212, 242
 link, 173
 link, log, 207
 link, log-log, 173
 link, logistic, 173
 link, probit, 173
Little SAS Book, 49
 liver cancer in mice, 182
 liver cirrhosis, 244
 LOESS, 160
 log link, 240
 log rank test, 232
 log transformation, 133
 log-log link, 173
 logarithm, viii, 7
 logistic regression, 169
 logistic regression in SAS, 173
 logistic regression, defined, 171
 logistic regression, deviance, 191
 logistic regression, extrapolation, 176
 logistic regression, residuals, 190
 logistic regression, SAS program, 174*t*
 logit, 171
 longitudinal data, 137, 139, 201
 lottery winners, 210*t*
 low-birth-weight babies, 49, 58, 96, 98, 105, 110
 LOWESS, 160

 macro data, 68
 main effects, 129
 mark-recapture, 179
 martingale residual, 242
 mathematics test scores, 153
 maximum likelihood estimates, 192
 mean, 2
 median, 3, 18, 37
 median test, 150
 Medicare, 68
 mental illness cases, 216
 merge two SAS datasets, 100
 meta-analysis, 26
 mice and cancer, 42, 142, 182
 Mickey Mouse, 217
 miracles, 49
 missing data, 138, 201
 model assumptions, 81
 model sum of squares, 80
 Mosteller, F., 91*t*, 260
 motorcycle accidents, 14
 MSCI EAFE, 77
 multicollinearity, 124, 129, 139
 multiple comparisons, 13, 34
 multiple myeloma, 245
 multiplicity of tests, 107
 multivariate displays, 94
 mutual funds, 116
 my dog ate it, 85

 naloxone, 222
 neutropenia, 179
 new home starts and sales, 9
 New York Mets, 177, 179
 nicotine, 140
 nodal involvement in prostate cancer, 187
 nonparametric regression, 160
 normal distribution, 3, 26
 not of full rank, 124
npar1way, 153, 156
 null hypothesis, 30
 null hypothesis, rejecting, 32
 nurses use of gloves, 200

 obesity, 14
 observational study, 12, 73
 odds ratio, 43, 171
 offset, 213, 221
 Olympic records, 15
 one-way ANOVA, 126, 165, 168
 outlier, 17, 40, 84, 152
 output delivery system, *ods*, 212

 p-value, *see also* statistical significance, 31
 Pagano, M., 50, 260
 paired t-test, 139
 panel data, 137
 partial *R*², 105
 partial correlation, 102, 104
 partial regression plot, 102
 Pearson residuals, 190
 percent body fat, 144
 pharmacokinetics, 164
 Phase I clinical trial, 229
 piano playing, ix
 pink eye, 14
 pirates, 237
 Pittsburgh Steelers, 201
 placebo, 235
 Poisson distribution, 204, 205, 206, 215
 Poisson distribution, illustrated, 207*f*
 Poisson regression, 204
 Poisson regression, chi-squared statistic, 209
 Poisson regression, deviance, 210
 Poisson regression, in SAS, 209
 Poisson regression, offset, 213, 215
 Poisson regression, parameter estimation, 211
 Poisson regression, residual, 212
 Poisson regression, scale parameter, 211
 Poisson regression, *obstats*, 211
 Poisson, Siméon-Denis, 204
 pooled sample, 151
 population, 1
 population mean, 3
 power, 33
 prediction, vii, 58
 prices of used cars, 143
 primary biliary cirrhosis, PBC, 244
 Pro Bowl appearances, 220
 probit regression, 172
 probit regression, SAS program, 174*t*
 probit transformation, 172
proc freq, 49
proc insight, 95

272 Index

- proc `logistic`, /n option, 186, 190
 proc `logistic`, class statement, 186
 proc `logistic`, descending option, 186, 190
 proc `logistic`, `iplots` or index plots, 190
 proc `npar1way`, 156
 proc `print` after every data step, 51, 100, 190, 209
 proportional hazards model, 239
 proportional hazards regression, 241
 proportional to, model, 71
 prostate cancer, 187, 188*t*
 puromycin, 141*t*
- quartiles, 37
- random effects model, 137, 142
 randomized study, 12
 rank-sum test, 155
 raw residual, 190
 raw residual, `Resraw`, 212
 reference category, 125
 regression diagnostics, 96, 99
 regression to the mean, 138, 140
`reschi`, chi-squared residual, 190, 213
 residual, 60, 61, 212, 242
 residual plot, 60, 81, 82, 96
 residual sum of squares, 79, 160
 residuals and errors, 82
 residuals for logistic regression, 190
 response variable, 52
 right censoring, 225
 risk set, 231
 robust estimator, 150
 robust statistics, 152
 root mean squared error, 81
 run-up, 165
- S&P 500, 77
 saddle point, 8
 sample average, 3
 sample median, 3
 sample variance, 4
 SAS, `array` statement, 148
 SAS, `datalines`, 51
 SAS, `data` step, 50
 SAS Help file, 49, 173
 SAS, `infile`, 51
 SAS, `input` in `data` step, 51
 SAS, `label` in `data` step, 51
 SAS, learn by copying, 49
 SAS log file, 49
 SAS, `output` statement, 212
 SAS, overview, 49
 SAS, `proc`, 50
 SAS `proc logistic`, 173
 SAS, `proc print`, 51, 190
 SAS, `title`, 51
 Satterthwaite adjustment, 41
 scale parameter, 211
 scatter-plot matrix, 94
 sea monsters, 237
 SEER, 32, 252
 SEM, 3
- sex discrimination, 198
 shark attacks, 204
 shoe tying, 226
 signal-to-noise ratio, 81
 significance level, *see also* statistical significance, 32
 slope, 7, 16
 slope of a line, 8
 spaghetti plot, 139, 142
 species diversity, 219
 sqrt function, 139
 square root transformation, 136, 139
 standard deviation, 3, 24
 standard error, 3
 standardized Student residual, 101
 statistical inference, 1
 statistical significance, viii, 41, 153
 statistical significance, versus clinical significance, 33
 statistics in sports, 178
 statistics, no need for, 78
 statistics, what is it?, 1
 StatLib, ix, 243
 stem and leaf plot, 151*t*
 stepwise regression, 105
 stock market returns, 18
 stormy seas, 237
 Student, 41
 Student t-test, 3
 Student t-test and indicator in regression, 122
 sum of squares, 79, 160
 Super Bowl, 184
 survival curve, 227
 Swiss fertility rates, 145
 symphony orchestras, 198
 synergy, 94, 128
- t-test, 3, 155
 test scores, 153
 testicular cancer, 222
 time series, 209
 Tmax, exercise, 164
 tobacco, 68
 topotecan, 179
 total sum of squares, 79
 Tourette's syndrome, 222, 224*t*
 toxoplasmosis, 181, 181*t*
 traffic accidents, 14
 transformations, 132
 transitional data, 137
 trees' heights, vii
 tuberculosis, TB, 11
 Tukey, J. W., 91*t*
 tumors in mice, 142
 twins, 204
 type I and type II analysis, 105
- unbiased estimator, 24
 unequal variances, 41
 unexplained sum of squares, 79
 unexplained variability, 79
 univariate, SAS procedure, 4
 university endowments, 110
 U.S. attitudes by foreigners, 139

273 **Index**

U.S. Constitution, 2
U.S. presidents, 17, 18
used car prices, 143
useless data, 90

variable, 49
variance, 4, 24, 41
variances, test of equality, 41
variances, unequal, 41
violin, 198
vital rates, 74
vital statistics, 221

waste run-up, 165
weather map, 4
wells, in ELISA, 147
whiskers, 39
white noise, 63
Wilcoxon test, 155, 234
wine prices, 116, 117*t*, 134, 140
women managers, 87

Yates corrected chi-squared, 46, 152

zero intercept, 71