Index

C_p statistic, 119
\( R^2 \), 85, 90, 118
\( R^2 \), adjusted, 118
0.05 p-value, 34
2-AAF, 192
5-FU, 243

" (tilde), 41, 70, 89, 95, 102

actuary, 245
adjusted \( R^2 \), 118
age, period, cohort model, 231
AIDS, 234
airlines, 95
alcohol, 75
alternative hypothesis, 32
alternative hypothesis, one-sided, 33
alternative hypothesis, two-sided, 33
analysis of variance (ANOVA), 87, 102
analysis of variance, total sum of squares, 88
answers to selected problems, 266
antagonism in drugs, 138
anti-fungal drugs, 138
arsenic, 76, 95, 120
ASPCA, 6
attendance at baseball games, 96, 120
auditions for orchestra, 206
autocorrelation, 215
Avedex, 46, 48
average, 2, 3

backward selection, 116
baseball, 120
baseball game attendance, 96
baseline hazard, 247
Bayesian statistics, 28
beetle mortality, 179, 183, 193
biased coin, 21
biased sample, 2
Big Bang, 28
bile duct, 243
bile duct cancer, 243

biliary cirrhosis, 253
binomial coefficient, 23
binomial distribution, 21–7, 213, 223
binomial response in glm, 182
blinded experiment, 62
Bonferroni correction, 36, 118
boxplot, 39, 41, 176, 177
boxplot notch, 41
brain cancer, 14
breast cancer, 79
Brewer’s yeast, 138
bubble plot, 105, 106
calculus, xii, 7, 30, 68, 246
cancer, 46, 83, 154, 237, 243
cancer deaths, 229
cancer incidence, 212
cancer mortality, 80
cancer of the bile duct, 243
cancer rates in Japan, 229
cars, characteristics in 1974, 124
cbind, 182
cell phones, 14
censored data, 125
censoring, 233
census, 2, 188
centenarian, 244
centenarian survival, 244
CGD, chronic granulomatous disease, 59
chi-squared distribution, 47, 264
chi-squared residual, reschi, 198
chi-squared test, 45, 48, 177
Cholangiocarcinoma, 243
choose symbol, 23
choose(), 23
cigarettes, 153
cirrhosis, 253
climate change, 210
clinical trial, 189, 237
cloth waste, 175
cohort, 230
coin toss, 21
complementary log–log link, 182
confidence interval, 25, 27, 43
confidence interval displacement, 202
confidence interval for binomial \( p \) parameter, 25
confounding relationships, 116
conjunctivitis, 14
Constitution, US, 2
continuity adjusted chi-squared, 48, 163
continuous events, 29
correlation, 69, 84
correlation matrix, 113
correlation of the global economy, 86
correlation plot, 113
corrplot, 113
covariance ratio, 112
Cox regression, 248, 249
cracker diet, 177
cross product ratio, 47
crossover experiment, 58
dark energy, 28
data mining, 119
dbinom(), 24
de Moivre, Abraham, 29
degrees of freedom, \( df \), 37, 51
demographic–economic paradox, 157
dependent variable, 57
derivative, xii, 7, 68
dermatologists’ fees, 78
deviance, 198, 220
deviance residual, 220, 251
deviance residuals, 200
df, 37, 51
dfbeta, 111, 151
dffits, 110, 156
diastolic blood pressure, exercise, 173
discrete events, 29
DNase in rats, 158
double blind experiment, 62
drought in United States, 72
drug interactions, 138
drug synergy, antagonism, 138
dummy variable, 131
e, \( e^x, \exp(x) \), 7
e, base of logarithm, xii
Ed50, 192
Efron, B, 190
El Salvador, 190
ELISA, 158
empirical data, 2
equipoise, 244
error distribution, 91
error sum of squares, 88
errors and residuals, 91
EVZIO, 231
exact distribution, 49, 53
exact methods, 27
exact test, 48, 52, 177
explained sum of squares, 89
exponential function, 7
extrapolation, 15, 72, 244
extrapolation, perils of, 71, 193
\( F \) ratio, 90
factorial, 22
factorial(), 23
failure rate, see hazard function
fair coin, 21
false discovery rate, 36
Farmer, 192
fast food, 14
female managers, 98
fertility rates, 157
firearm deaths, 97
Fisher exact test, 52, 163
Florida cancer mortality, 80
Food and Drug Administration, 36
food imports, 96
football, 61, 193, 209
football, yards rushing, 227
foreign owners of US debt, 17
forward selection, 116
friend, kind, patient, and understanding, 53
full rank, 135
fungicide, 46, 48
fusion experiment, 37, 60, 173
G-CSF, 189
Galápagos Islands, 227
gasoline price, 15, 144
Gauss, ICF, 29
Gaussian distribution, 29
generalized linear models, 182
ggplot2, 105
global stock markets, 86
global warming, 210
glove use, 207
Gosset, WS, 45
graphical displays, 104
Guinness Brewery, 45
halibut, 251, 252, 273
Hamburger Helper lasagna, 131
hat diagonal, 201, 259
hat diagonal in Poisson regression, 221
<table>
<thead>
<tr>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>hat matrix diagonal, 106, 133, 151</td>
</tr>
<tr>
<td>hazard function, 244, 245</td>
</tr>
<tr>
<td>hazard ratio, 250</td>
</tr>
<tr>
<td>healthcare costs, 123</td>
</tr>
<tr>
<td>healthiest state, 169</td>
</tr>
<tr>
<td>heart disease, 83</td>
</tr>
<tr>
<td>helium-filled football, 61</td>
</tr>
<tr>
<td>high leverage observations, 106</td>
</tr>
<tr>
<td>high-dose chemotherapy, 189</td>
</tr>
<tr>
<td>HIPPA, 257</td>
</tr>
<tr>
<td>histogram, 55, 56</td>
</tr>
<tr>
<td>HIV, 234</td>
</tr>
<tr>
<td>hot deck, 151</td>
</tr>
<tr>
<td>Hubble constant, 27</td>
</tr>
<tr>
<td>human immunodeficiency virus, 12</td>
</tr>
<tr>
<td>hypothesis testing, 32</td>
</tr>
<tr>
<td>ice cream flavors, 160</td>
</tr>
<tr>
<td>IMF, International Monetary Fund, 11</td>
</tr>
<tr>
<td>independence in a 2 × 2 table, 46</td>
</tr>
<tr>
<td>index plot, 198</td>
</tr>
<tr>
<td>indicator variable, 130, 131</td>
</tr>
<tr>
<td>indicator variable as a t-test, 133</td>
</tr>
<tr>
<td>influence, 110, 133</td>
</tr>
<tr>
<td>influence() , 109</td>
</tr>
<tr>
<td>informative censoring, 235</td>
</tr>
<tr>
<td>inner quartile range, 41</td>
</tr>
<tr>
<td>install.packages, 113</td>
</tr>
<tr>
<td>integral of a function, xii , 8</td>
</tr>
<tr>
<td>interactions, 140</td>
</tr>
<tr>
<td>intercept, 7, 15, 69</td>
</tr>
<tr>
<td>intercept as reference category, 135</td>
</tr>
<tr>
<td>intercept, zero, 77</td>
</tr>
<tr>
<td>interval censoring, 234</td>
</tr>
<tr>
<td>jackknife, 110, 204, 221</td>
</tr>
<tr>
<td>January temperatures in US cities, 102, 122</td>
</tr>
<tr>
<td>Japanese cancer rates, 229</td>
</tr>
<tr>
<td>jittering, 39, 41, 74, 147, 154, 175, 176, 251</td>
</tr>
<tr>
<td>kangaroo flush, 58</td>
</tr>
<tr>
<td>Kaplan–Meier curve, 235</td>
</tr>
<tr>
<td>Kaplan–Meier estimate, 235</td>
</tr>
<tr>
<td>lambda (3), 214</td>
</tr>
<tr>
<td>last value carried forward, 151, 209</td>
</tr>
<tr>
<td>lattice package, 161</td>
</tr>
<tr>
<td>LD_{191}, 192</td>
</tr>
<tr>
<td>LD_{250}, 188, 258, 271</td>
</tr>
<tr>
<td>least squares, 68, 103, 220</td>
</tr>
<tr>
<td>left censoring, 234</td>
</tr>
<tr>
<td>leverage, 94, 106, 172, 227</td>
</tr>
<tr>
<td>lightning strikes, 212</td>
</tr>
<tr>
<td>linear functions, 6</td>
</tr>
<tr>
<td>lines() , 24</td>
</tr>
<tr>
<td>link, 182</td>
</tr>
<tr>
<td>link function, 248</td>
</tr>
<tr>
<td>link, log, 217</td>
</tr>
<tr>
<td>link, logistic, 182</td>
</tr>
<tr>
<td>link, probit, 182</td>
</tr>
<tr>
<td>liver cancer in mice, 192</td>
</tr>
<tr>
<td>liver cirrhosis, 253</td>
</tr>
<tr>
<td>LOESS, 170</td>
</tr>
<tr>
<td>log-rank test, 240</td>
</tr>
<tr>
<td>log transformation, 145</td>
</tr>
<tr>
<td>log–log link, 182</td>
</tr>
<tr>
<td>logarithm, xii , 7</td>
</tr>
<tr>
<td>logistic regression, 178</td>
</tr>
<tr>
<td>logistic regression in R, 182</td>
</tr>
<tr>
<td>logistic regression, defined, 181</td>
</tr>
<tr>
<td>logistic regression, deviance, 198</td>
</tr>
<tr>
<td>logistic regression, program, 183</td>
</tr>
<tr>
<td>logistic regression, residuals, 197</td>
</tr>
<tr>
<td>logit, 179</td>
</tr>
<tr>
<td>logit link, 182</td>
</tr>
<tr>
<td>longitudinal data, 149, 152, 207</td>
</tr>
<tr>
<td>lottery winners, 217</td>
</tr>
<tr>
<td>low-birth-weight babies, 54, 64, 107, 108, 116, 120</td>
</tr>
<tr>
<td>LOWESS, 170</td>
</tr>
<tr>
<td>Mackenzie, CA, 188</td>
</tr>
<tr>
<td>macro data, 75</td>
</tr>
<tr>
<td>main effects, 141</td>
</tr>
<tr>
<td>Mantel–Haenszel test, 242</td>
</tr>
<tr>
<td>marginal correlation, 116</td>
</tr>
<tr>
<td>mark–recapture, 188</td>
</tr>
<tr>
<td>martingale residual, 251</td>
</tr>
<tr>
<td>match(), 105</td>
</tr>
<tr>
<td>mathematics test scores, 165</td>
</tr>
<tr>
<td>matrix scatter plot, 55, 56, 139</td>
</tr>
<tr>
<td>maximum likelihood estimates, 200</td>
</tr>
<tr>
<td>maximum likelihood estimation, 39</td>
</tr>
<tr>
<td>mean, 2</td>
</tr>
<tr>
<td>median, 3, 20, 41</td>
</tr>
<tr>
<td>median test, 160</td>
</tr>
<tr>
<td>Medicare, 75</td>
</tr>
<tr>
<td>mental illness cases, 223</td>
</tr>
<tr>
<td>meta-analysis, 28</td>
</tr>
<tr>
<td>mice and cancer, 46, 154, 192</td>
</tr>
<tr>
<td>missing data, 150, 209</td>
</tr>
<tr>
<td>model assumptions, 91</td>
</tr>
<tr>
<td>model sum of squares, 89</td>
</tr>
<tr>
<td>mortality rates, 82</td>
</tr>
<tr>
<td>Mosteller, F, 102, 273</td>
</tr>
<tr>
<td>motorcycle accidents, 14</td>
</tr>
<tr>
<td>MSCI EAFE, 86</td>
</tr>
<tr>
<td>multicolinearity, 135, 141, 229</td>
</tr>
<tr>
<td>multiple comparisons, 13, 35</td>
</tr>
<tr>
<td>multiple myeloma, 254</td>
</tr>
<tr>
<td>multiplicity of tests, 118</td>
</tr>
<tr>
<td>multivariate displays, 104</td>
</tr>
</tbody>
</table>

© in this web service Cambridge University Press & Assessment  www.cambridge.org
multivariate methods, 113
mutual funds, 128
my dog ate it, 95
NA, missing values in R, 80
naloxone, 231
NARCAN, 231
neutropenia, 189
New York Mets, 186, 188
nicotine, 153
nodal involvement in prostate cancer, 195
nonparametric regression, 170
normal distribution, 3, 29
not of full rank, 135
null hypothesis, 32
null hypothesis, rejecting, 34
nurses’ use of gloves, 207
obesity, 14
observational study, 12, 80, 82
odds ratio, 47, 50, 179
offset, 221
olsrr library, 117
Olympic records, 14
one-way ANOVA, 137, 176, 177
outlier, 18, 41, 93, 164

p-value, see also statistical significance Pagano, M, 54
paired t-test, 152
panel data, 149
partial correlation, 113, 115
pcor, 115
Pearson residual, 219
Pearson residuals, 198
percent body fat, 156
Peto–Peto test, 242
pharmacokinetics, 174
Phase I clinical trial, 237
piano playing, xiii
pink eye, 14
pirates, 245
Pittsburgh Steelers, 209
placebo, 243
plot, 104
plot(), 24
pnorm(), 30
Poisson distribution, 212, 213, 223
Poisson distribution, illustrated, 214
Poisson regression, 212
Poisson regression diagnostics, 218
Poisson regression, deviance, 220
Poisson regression, offset, 221, 223
Poisson regression, parameter estimation, 220
Poisson, Siméon-Denis, 154
puromycin, 154
quantile function, 263
quartiles, 41

R Help file, 54
R, overview, 52
random dot stereogram, 37
random effects model, 128
randomized study, 13
rank sum test, 167
raw residual, 198
reading text in R, 101
reference category, 135
regression diagnostics, 106, 110
regression to the mean, 150–2
reschi, chi-squared residual, 198
residual, 66, 67
residual in proportional hazards, 251
residual plot, 66, 67
S&P 500, 86
sampling, 3
scatter plot, 55, 56

population, 1
population mean, 3
power, 34
prop.cor, 115
prediction, xi, 64
prices of used cars, 155
primary biliary cirrhosis, 253
Pro Bowl appearances, 227
probit link, 182
probit regression, 181
probit regression, program, 183
probit transformation, 181
prop.test(), 26, 59
proportional hazards model, 247
proportional hazards regression, 249
proportional to, model, 77
prostate cancer, 195, 196

© in this web service Cambridge University Press & Assessment
www.cambridge.org
Index

scatter plot matrix, 104
sd(), 4
SEER, 34, 261
SEM (standard error of the mean), 3
sex discrimination, 206
shark attacks, 212
shoe tying, 234
signal-to-noise ratio, 90
significance level, see also statistical significance versus clinical significance
slope of a line, 7, 15
Smith, D, 189
spaghetti plot, 152, 154, 268
species diversity, 227
sqrt function, 151
square root transformation, 149, 151
standard deviation, 3, 25
standard error, 3
statistical inference, 1
statistical significance, xii, 44, 165
statistical significance, versus clinical significance, 35
statistics in sports, 187
statistics, no need for, 87
statistics, what is it?, 1
stem and leaf plot, 161
stepwise regression, 116
stereogram, 37
stock market returns, 18
stormy seas, 245
Student, 45
Student t-test, 3, 36
Student t-test and indicator in regression, 133
sum of squares, 88, 170
Super Bowl, 193
survival curve, 235
survival object in R, 250
Swiss fertility rates, 157
symphony orchestras, 206
text(), 105
time series, 215
T max , exercise, 174
tobacco, 75
topotecan, 189
total sum of squares, 88
Tourette’s syndrome, 231
toxoplasmosis, 190
traffic accidents, 14
transformations, 63, 144
transitional data, 149
trees’ heights, xi
tuberculosis, 12
tuberculosis, TB, 11
Tuft ER, 105, 273
Tukey, JW, 102
tumors in mice, 154
twins, 212
US attitudes by foreigners, 152
US Constitution, 2
US presidents, 17, 18
unbiased estimator, 25
unequal variances, 44
unexplained sum of squares, 88
unexplained variability, 88
university endowments, 120
US Food and Drug Administration, 36
US murder rates, 97, 98
used car prices, 155
useless data, 101
\texttt{var()}, 4
variable, 54
variance, 4, 25
variances, unequal, 44
violence rates, 97
violin, 206
vital rates, 82
vital statistics, 229
Washington, DC, weather, 210
waste run-up, 175
weather map, 4
wells, in ELISA, 158
whiskers, 41
white noise, 69
Wilcoxon test, 167
Wilkinson, L, 273
wine prices, 125, 126, 146, 152
women managers, 98
Yates corrected chi-squared, 48, 163
yeast, 138
zero intercept, 77