INDEX

Note: Locators followed by “f” and “t” refer to figures and tables respectively.

<table>
<thead>
<tr>
<th>Function or Method</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>acf()</td>
<td>6, 39f</td>
</tr>
<tr>
<td>autocovariance estimation coding</td>
<td>38–39</td>
</tr>
<tr>
<td>background</td>
<td>37–38</td>
</tr>
<tr>
<td>and spectrum</td>
<td>68–69</td>
</tr>
<tr>
<td>for white noise errors</td>
<td>68f</td>
</tr>
<tr>
<td>acos()</td>
<td>6</td>
</tr>
<tr>
<td>AIC. See Akaike's information criteria</td>
<td></td>
</tr>
<tr>
<td>Akaike's information criteria (AIC)</td>
<td>106–107, 282</td>
</tr>
<tr>
<td>as cross-validation, NYC temperatures,</td>
<td>112</td>
</tr>
<tr>
<td>model selection with</td>
<td></td>
</tr>
<tr>
<td>anova()</td>
<td>10</td>
</tr>
<tr>
<td>arima.sim()</td>
<td>6, 130, 159–160</td>
</tr>
<tr>
<td>ARMA(2,2) model</td>
<td>166–167</td>
</tr>
<tr>
<td>AR(m) filtering matrix</td>
<td>187–189</td>
</tr>
<tr>
<td>filtering information</td>
<td>189</td>
</tr>
<tr>
<td>linear algebra</td>
<td>187</td>
</tr>
<tr>
<td>and lm()</td>
<td>188</td>
</tr>
<tr>
<td>to model MA(3)</td>
<td>181–184</td>
</tr>
<tr>
<td>standard computations</td>
<td>188</td>
</tr>
<tr>
<td>AR(1) model for irregular spacing</td>
<td></td>
</tr>
<tr>
<td>final analysis</td>
<td>268–269</td>
</tr>
<tr>
<td>method</td>
<td>266</td>
</tr>
<tr>
<td>motivation</td>
<td>265–266</td>
</tr>
<tr>
<td>results</td>
<td>266–267</td>
</tr>
<tr>
<td>sensitivity analysis</td>
<td>267</td>
</tr>
<tr>
<td>AR(m) structure, residuals for</td>
<td></td>
</tr>
<tr>
<td>data display</td>
<td>235–236, 235f–236f</td>
</tr>
<tr>
<td>filtering twice</td>
<td>236–238</td>
</tr>
<tr>
<td>ar.yw()</td>
<td>6, 171–174</td>
</tr>
<tr>
<td>asin()</td>
<td>6</td>
</tr>
<tr>
<td>Assumptions</td>
<td></td>
</tr>
<tr>
<td>equal variance</td>
<td></td>
</tr>
<tr>
<td>regression</td>
<td>31</td>
</tr>
<tr>
<td>two-sample t-test</td>
<td>31</td>
</tr>
<tr>
<td>independence</td>
<td>31</td>
</tr>
<tr>
<td>introduction</td>
<td>28–29</td>
</tr>
<tr>
<td>logarithmic transformations, illustration</td>
<td></td>
</tr>
<tr>
<td>of</td>
<td>32–34</td>
</tr>
<tr>
<td>normality</td>
<td>29–30</td>
</tr>
<tr>
<td>heavy tails</td>
<td>30–31</td>
</tr>
<tr>
<td>left skew</td>
<td>30</td>
</tr>
<tr>
<td>right skewed</td>
<td>30</td>
</tr>
</tbody>
</table>

INDEX

atan(), 6
Autocorrelation
AR(1), 41
AR(2), 41–42
estimation, 37
for MA(1) models, 51–52
for MA(2) models, 52
stationarity, 36
Autocovariance
AR(1), 41
AR(2), 41
ARMA\((m,l)\) model, 169–170
estimation, 37, 38
properties, 36–37
stationarity, 36
white noise, 37
Autoregressive model of order 1, AR(1), 35
adjustments, 125–151
implications, 151–152
skip method, 152
autocorrelation, 41
autocovariance, 41
definition, 40
elements (stable and unstable models), 44–46, 45f–46f
illustration, 42
Autoregressive model of order 2, AR(2), 35
autocorrelation, 41–42
autocovariance, 41
examples, 46–50, 46r, 47f–49f
and power spectrum, 68–70, 70–72
preliminary facts, 40
R code, 42–43
simulating data, 42–44
Backshift operator
and ARMA\((m,l)\) models, 162
definition, 161
examples, 162–164
stationary condition for AR(1) model, 161
Bayesian information criteria (BIC), 106–107
Best linear unbiased estimators (BLUES), 9
BIC. See Schwarz information criteria
BLUES. See Best linear unbiased estimators
Boise river flow data, 121f–122f
data splitting, 123
model selection with AIC, 122–123, 123f
model selection with filtering, 147–151
residuals, 123–124, 124f
Breast cancer, data analysis
background, 228–229
estrogen response negative, 228–229, 229f
estrogen response positive, 228–229, 229f
and female colon cancer, 241f
second data set (1975–2005)
background, 232–233
data structure, 232
data trend, 233–235, 234f–235f
regression analysis with filtered data, 238–243
residuals for AR\((m)\) structure,
235–238
statistical analysis, 233
Carrington, Richard, 214
Complex conjugates, 62
Complex numbers, 62
magnitude of, 62–63
Complex periodic model
accidental deaths
data splitting, 119–120
Fourier series structure, 116
model selection with AIC, 117–118
model selection with likelihood ratio
tests, 118–119
periodic data, comments on,
120–121
R Code, fitting large Fourier series,
116–117
residual, 120f
training set model, 120f
validation set model, 120f
monthly river flows, furnas 1931–1978
AR\((m)\) filtering matrix, 187–189
data, 185
data splitting, 191–192
model selection, 189, 189r, 192–193, 192f, 193f
periodic model, 123f
predictions for AR($m$), 190
saturated model, 186–187, 187f
Comprehensive R Archive Network (CRAN), 283
Coronal mass ejections, 213
cos(), 6, 66, 90
CRAN. See Comprehensive R Archive Network
Creek, Gregory, 248
Crosby, Ben, 245
Cross-validation, NYC temperatures
AIC for, 112
data splitting, 108–110, 109t, 110f
explained variation, $R^2$, 108
leave-one-out cross-validation, 110–112
Data import, 7
DataMarket, 21
export options, 276
homepage, 275f
licensing agreements, 280
login page, 275f
overview, 273–276
time series loading, 277–279
Data simulations, 9
Data splitting, 119–120, 191–192
d’c, 6
45-Degree line model, 104–106
dmlist(), 277
New York temperature data plot, 288f
dmseries(), 277
New York temperature data, 279
Endocrine disruptors, 228
Equal variance assumption
regression, 31
two-sample t-test, 31
ER+. See Estrogen response positive
ER-. See Estrogen response negative
Estrogen response negative (ER-), 228
breast cancer, 228–230
rates, 230–231
Euler’s formula, 63
expt(), 6
Explained variation, $R^2$, 108
Export options, 276
Fast Fourier transform (FFT), 75–77
Female colon cancer, 241f
FFT. See Fast Fourier transform
Filtering, 125, 133–134
and Boise river flow data, 147–151
comments on, 137–138
and global warming model, 136, 136f
floor(), 85
“For” statement, 5–6
Fourier series, 116–118, 119f
Fourier series structure, 116
Functions (R)
acos(), 6
asin(), 6
atan(), 6
cos(), 6
d’c, 6
expt(), 6
log(), 6
pi, 6
sin(), 6
sqrt(), 6
tan(), 6
see also Time series, functions
General ARMA models
arima.sim(), 159–160
and backshift operator, 162
texts, 160–161, 160f
mathematical formulation, 159
representative collection, 160f
spectrum for, 175–177
Geometric series, 63
Hat matrix, 111
Heavy tails, 30–31
help(), 10
help(“numericDeriv”), 289
High elevation (snow), 245
Homepage, 275f
Hyndman, Rob, 24
INDEX

“If” statement, 5
Impulse response operator
computation
coefficients computation, 165–166
definition, 165
plotting, 166–167
interpretation, 167
intuition, 164
utility, 167
Influential points, 13
Information criteria
Akaike's information criteria, 106–108
and model selection, 173–174, 173t
Schwarz information criteria, 106–108
Inquiry functions
anova(), 10–12
help(), 10–12
names(), 10–12
summary(), 10–12
International sunspot number, 213
Intervention model
directory assistance
concern, 199
data, 199
filtering information, 199–200, 202
model selection, 200, 200t
saturated model, 199–200
ozone levels in Los Angeles, 202–205, 203f, 204f, 204t
structure, 198

kappa(), 287
Leave-one-out cross-validation, 110–112
Left skew, 30
Leverage points, 13
Licensing agreements, 280
Likelihood ratio tests, 101–104
model selection with, 118–119
Linear model, 102–104, 102f
lm(), 9–10, 214
vs. nls(), 216
log(), 6
Login page, 275f
lowess() function, 81–82
Low (rain) elevation watersheds, 245
initial fits for, 247f
Matrix manipulation, in R
commands, 16
OLS, 15–16
mean(x), 4
Mid (mixed) elevation watersheds, 245–248
initial fits, 247f
Modeling
algorithm, 180
assumption, 180–181
example
AR(m) filter to model MA(3), 181–184
CO2 levels at Mauna Lau, 193–198
monthly river flow, 185–193
skip method, 184–185, 184t
Model selection
with AIC, 117–118
with likelihood ratio tests, 118–119
Monthly river flow, complex periodic model
AR(m) filtering matrix
filtering information, 189
fitting a model with lm(), 188
linear algebra, 187
standard computations, 188
data, 185
data splitting
computations, 191–192
linear algebra, 191
overview, 191
model selection, 189, 192–193
predictions for AR(m) model, 190
saturated model, 186–187
Moving average model, MA(1)
acf() plots, 55, 55f–56f
and AR(m) models, 52
autocorrelation for, 51–52
simulated examples, 52–54, 53f–54f
Moving average model, MA(2)
acf() plots, 54–55, 55f
autocorrelation for, 52
simulated examples, 54
Naïve analysis
CO2 and temperature change
association, 258–259
model selection, 258
saturated model, 256–257
INDEX

Naïve code, 72–74

name(), 10–11

Naming conventions, 4

Nested models, 99–101

Newton’s method (for nonlinear optimization), 284–285

nls(), 214–216, 289

vs. lm(), 216

Noise, 18

Nonlinear optimization, tutorial on general problem, 286

introduction, 284

Newton’s method for, 284–285

revisit, 286

Normality assumption, 29

heavy tails, 30–31

left skew, 30

right skew, 30

numericDeriv(), 289

NYC temperatures

application, 146–147

AR(1) prediction model, 144

cross-validation

Akaike’s information criterion, 112

data splitting, 108–110

explained variation, 108

leave-one-out cross-validation,

110–112

data, 142–144

outlier, 92, 92f

periodic function fitting, 91, 91t

prediction intervals, 142–144

simulation, 144–146

Observatory factor, 213

OLS. See Ordinary least squares

Ordinary least squares (OLS), 8–9

pacf(), 6, 174–175

Partial autocorrelation plot

hypothesis tests sequence, 174, 174t

pacf() function, 174–175, 175f

Periodic function fitting, 91

Periodic models, 90–91

complications

accidental deaths, 96

CO₂ data, 93–94

sunspot data, 94–96

daily average, 205–206, 206f

easy (NYC temperature data), 90–91, 91t

outlier, 92

periodic function fitting, 91

refitting, 92–93

monthly average, 206–207, 207f

weekly average, 205–206, 206f

Periodic transcendental functions, 64

Periodogram, 65

and acf() plot, 78f

example, 74–75

Naïve code for, 72–74

periodic analysis, 85–86

periodic behavior, 73f

for power spectrum, 68

and smoother, 79–80

and white noise, 73f

Personal reduction coefficient (K), 213

Phase, 89

Pi, 6

Power spectrum, 65

and acf() plot, 68f

for ARMA processes, 175–177

for AR(1) models, 68–70, 69f–70f

for AR(2) models, 70–72

and autocorrelation function, 66–67

definition, 66

and periodogram plot, 68f

for white noise, 68

Predictions for AR(m), 190

PRESS, 281–282

Prostate cancer, data analysis

background, 228–229

estrogen response negative, 228–229

estrogen response positive, 228–229


second data set (1975–2005)

background, 232–233

data structure, 232

data trend, 233–235

regression analysis with filtered data,

238–243

residuals for AR(m) structure,

235–238

statistical analysis, 233

Prostate-specific antigen (PSA), 234

PSA. See Prostate-specific antigen

Pseudo-periodic model, 161

p-values, 29, 100
INDEX

qqnorm(), 13, 14f, 15
Quadratic model, 101–104, 101t
Quasi-independent observations, 137

R (programming language)
  code, 3
  common functions, 6
  console code, 277
  conventions, 5
  data sources, 24–26
  inquiry functions, 10–12
  matrix manipulation, 15–16
  model parameters estimation, 9–12
  smoothers in 
    lowess(), 81–82
    smooth.spline(), 82–83
  structures, 5
R Code, fitting large Fourier series, 116–117
rdatamarket package, 283
read.csv(), 7
read.delim(), 7
read.table(), 7, 25
Real data, 21
Refitting, 92–93
Regression, 31
Regression model
  matrix representation, 9
  OLS estimates, 9
  ordinary least squares, 8–9
  for periodic data, 89–96
Relative sunspot number, 213
Residuals analysis, 14f
  influential points, 13
  lack of fit, 13
  nonwhite noise error, 13–14
  normality, 13
  outliers, 13
  plots, 14–15, 14f
  unequal variance, 14
Richer models, 116
Right skew, 30
  p-values, 30
  and unknown period, 95f
R^2 pred, 111

Saturated model, 117, 193–194
  and data fit, 197–198
  and filter, 193–194
  naïve analysis, 256–257
  pruning, 196–197
  residuals, 195f
  scan(), 7, 25
Semmelweis, Ignaz Philipp, 138
Semmelweis data, 22–24, 25–26
Semmelweis intervention
  data, 138–139
  filtered analysis, 140–142
  inferences, 142
  serial correlation, 139
  transformations, 142
  vs. patch/uncut case, 139–140
Serial correlation
  and Semmelweis intervention, 139
Signals, 18
Simple mean model, 104–106
Simple regression, 20
  analysis of variance, 100t
  hypothesis tests, 99–101
  ratio tests, 101–104
  Sleuth case, global warming
    analysis, 135–136, 135f, 136t
    data, 132–133
    filtering, 133–134, 137–138
    simulation, 134–135, 135t
Simulated data, 20–21
sin(), 6
Skip method, 152, 184–185
Smoothers, 80
  lowess() function, 81–82
  for series
    known period, 85–86, 86f
    unknown period, 86–87, 86f
  smooth.spline() function, 82–83
smooth.spline() function, 82–83
Solar flares, 213–214
solve(), 16
span(), 84–85
spec.pgram(), 6, 75–77
sqrt(), 6
SSE, 98
SST, 282
Standard errors, 13
Statistical operations, 4
Straight-line model, 104–106
summary(), 10, 12
sum(x), 4
INDEX

\[ \tan() \], 6
Tennant, Christopher, 245
Time series, 19
- assumptions, 143
- data, 21–24
- extrapolation, 143
- prediction intervals, 143
Time Series Data Library, 21, 24
Time series function (R)
- `acf()`, 6
- `arima.sim()`, 6
- `ar.yw()`, 6
- `pacf()`, 6
- `spec.pgram()`, 6
- `ts()`, 6
Time series loading, 277–279
Transcendental series, 63
- `ts()`, 6, 25
- t-tests, 23
Two-sample t-test, 23, 29, 31
- adjustment for AR(1), 128–129
- assumption, 31
- simulation example, 129–130, 131f
- Sleuth data, 125–128
- Sleuth data analysis, 131–132
Variable lag, 263
Vostok ice core data
- alignment
- issues, 254
- matched dates, 254, 255f
- need, 254
- patterns, 254, 255f
- time stamps, 255–256, 256f
AR(1) model for irregular spacing
- final analysis, 268–269
- method, 266
- motivation, 265–266
- results, 266–267
- sensitivity analysis, 267, 268f
- naïve analysis
- CO\(_2\) and temperature change
- association, 258–259
- model selection, 258, 258f
- saturated model, 256–257, 257f
related simulation
- code, 260
- model, 259–260, 261
- sample of 283, 262–265, 265f
- source, 251–252, 252f
Watersheds data
- averaging data, 248–249
- fitting Fourier series
- data structure, 246
- data to physical processes, connecting patterns in, 246–248
- Fourier series fits to data, 246
- high elevation (snow), 245–248
- low (rain) elevation, 245–248
- mid (mixed) elevation, 245–248
- results, 250
White noise, 18, 23, 24
- and autocovariance, 37
- and power spectrum, 68
Wolf, Rudolf, 213
Wolf number, 213
- amplitude, instability in, 217–218
- background, 213–214
- data splitting (for prediction)
  - approach, 220–222
  - AR-adjusted predictions, 223
  - AR correction, 222
  - fitting one step ahead, 222, 223f, 224f
  - model selection, 223–224
  - predictions two steps ahead, 224–225, 225f, 225f
- mean, instability in, 217–218
- `nls()` function, 214–216
- period, instability in, 217–218, 218f
- period determination, 216–217
- sunspot data, 217f
- for unknown period, 214
Yule–Walker equations
- AR(m) and, 170–174
- errors sequence, 172f
- model selection (using information criteria), 173–174, 173f
Zürich number, 213