INDEX

A
ABS function, 473
access and accesses files, 65
ACROSS option (REPORT), 451
ADJUST= option (multiple comparisons), 314, 322
Akaike Information Criteria (AIC), 343
ALPHA= option, 313, 363
analysis of covariance, 327 ff
analysis of variance (ANOVA/AOV), 311 ff
one-way see one-way ANOVA
two factor see two factor ANOVA
ANALYSIS option (REPORT), 449
ANOVA, 327 ff
ANGLE (A=) option, 401
ANNOTATE= option, 394
annotation data sets, 428–9
ANOVA/AOV see analysis of variance
(ANOVA/AOV)
appending data sets, 94
Apple Mac, 4
ARCOS function, 476
AR(1) covariance structure see autoregressive (AR(1))
covariance structure
area under the curve (AUC) see ROC curve
arithmetic/mathematical functions, 135, 473
arithmetic operators, 79
ARRAY statement, 163 ff
ARSin function, 476
ASCII file/format, 6, 15, 37, 48
assigning missing values see missing values, assign
asterisk (*) for comments, 116
ATAN function, 476
automated model selection, 301–3
autoregressive (AR(1)) covariance structure, 342
AXIS statement, 394
BACKGROUND = (color) option, 470
BACKWARD procedure, 301–3
bar charts, 405 ff
bar charts (using SGCHART), 424, 425
bar fill, 465
bar patterns, 465
binary data, 493
binary logistic regression see logistic regression
Bonferroni test, 314
box-and-whiskers plot, 317, 353, 414, 416–417
box plots see box-and-whiskers plot
BOXSTYLE= option, 415
BUBBLE Statement (SGPLOT), 394, 421
BY CLASS, 225
BY DESCENDING statement, 92
BY GROUP, 92, 121
BY statement, 91, 129
C
CALL SYMPUT, CALL SYMPUTN, 189, 486
CAT, CATS, CATT, CATX functions, 480
categorical data, 492–3
CAXIS= option, 397, 405
INDEX

CBOXES= option (BOXPLOT), 415
CBOXFILL= option (BOXPLOT), 415
CEIL function, 483
center output see NOCENTER
CFILL statement, 226–7
CFRAME option, 397, 405
character functions, 480
character variables, 24
chi-square test, 492
chi-square test (PROC FREQ), 253–60
choosing a SAS procedure, 491ff
CITRIX, 4, 53
CLASS statement, 218, 237, 342, 363
CLDIFF multiple comparisons, 314
Clear All command, 9
CLI statement, 298
CLM= option (GCHART), 406
COCHRAN option, 276
Cochran Q test, 492
code files, 15
color choices, 463
COLOR= option (C=), 398, 399, 401
column input, 29
COLUMN statement (REPORT), 444
comma separated values (CSVs), 42–4, 60–65
compact data entry method, 27
comparing K dependent samples, 355
comparing K independent samples, 350
comparing three or more means using one-way ANOVA, 312 ff
comparing three or more repeated measures, 492
comparing two dependent samples, 492
comparing two independent samples, 492
comparision statistics, choosing, 492
COMPBL function, 480
compile SAS code, 69
compound symmetry (CS) covariance structure, 342
COMPRESS function, 480
COMPUTE option(REPORT), 444, 445, 450
conditionally read a data set line, 87
conditional operators, 82
certainty limits, plotting, 404, 413
CONTENTS option, 67
contingency tables (PROC FREQ), 253–60
contingency Adj. chi-square, 258
CONTRAST statement, 323
Cook's D plot (regression), 295
Cook's D statistic, 307
C= option see COLOR= option (C=)
correlation analysis, 283 ff
correlation analysis using WITH option, 289
correlation coefficient (Pearson), 283 ff
correlations, factor analysis, 386
COS function, 476
COSH function, 476
COUTLINE option, 405
covariance structure, 342
creating a SAS library, 53 ff
creating new variables, 78 ff
creating SAS data set, 20, 24, 52 ff
crosshatch pattern, 465
cross-tabulation, 492
CS covariance structure see compound symmetry
CTABLE option, 364, 372
CTEXT option, 397, 405
Ctrl-Break, 12
cut off value, logistic regression, 372
D
DATALINES statement, 26
data sets see SAS data sets
DATA statement see DATA step
DATA step, 20 ff
data types see variable types
DATDIF function, 477
date and time functions, 476ff
DATE function, 477
date option see NODATE
DATEPART function, 477
dates in SAS, 476, 477
DATETIME function, 477
date variables, 24
day function, 477
day of week format see DOWNAME
DBase files, 65
DBMS sources, 65
DEFINE statement (REPORT), 444
density plots, 226, 421
dependent variables, 291, 297, 362
DEQUOTE function, 480
descending option, 362, 363
descriptive statistics, 491–2
descriptive statistics, choosing procedure, 491
dhms function, 156, 477
DIF function, 486
DISCRETE option (GCHART), 405, 418
DISPLAY statement (REPORT), 445
dlm option, 41
dlmstr, 41
DO LOOP, 167 ff
%DO Loop (Macro), 187
DONUT option (GCHART), 405
Dot (.) see missing values, assign
DO UNTIL loop, 167
do WHILE loop, 167, 169
downame format, 487
download sample files, 4
drinking analysis, 254–6
DROP statement, 88
dsd option, 41
Duncan multiple comparison test, 314
DUNNETT multiple comparisons, 314
Dunn's test (multiple comparisons), 357
**INDEX**

**E**
editor see enhanced editor
Editor window, 6
eigenvalues (factor analysis), 381, 388
enhanced editor, 6, 12
equality of variances test, 275
EQUAMAX option, 378
_ERROR_, 70
ERRORBAR= option (GCHART), 406
EVENT= option, 363
EXACT statement, 244, 348
Excel, 15, 22, 60, 65, 78, 443
EXCEL output destination see TAGSETS.EXCELXP
output destination
EXCEL, saving table to (TABULATE), 442
execute SAS code, 69
EXP function, 473
export see PROC EXPORT
EXPS option, 379
Extensible Markup Language (XML) output
destination, 194

**F**
FACT function, 473
factor analysis, 377ff
false negative, 373
false positive, 373
FILENAME command, 15
FILE PRINT statement, 458
FILLATTRS= option (SGPLOT), 421
financial functions, 484
FIND function, 155, 480
finding FIRST and LAST values, 107
FIRSTOBS= option, 42, 128, 131
FIRST values, find, 107
Fishers test, 245
fixed factors, 333
FLOOR function, 483
FLOWOVER option, 42
FLYOVER= option, 470
FONT_FACE= option, 470
FONT= option, 470
FONT_SIZE= option, 470
FONTS, SAS, 462
FONT_WEIGHT= option, 470
FOOTNOTE Statement, 113, 495
FOREGROUND= (color) option, 470
FORMAT= option, 432
formats, 34, 35
creating see PROC FORMAT
input, 469–70
library, 105–7
output, 471
permanent see PROC FORMAT
formatted input, 32
FORWARD procedure, 301–3
freeform list input, 24
FREQ statement, 313
frequency tables, 246, 249
Friedman’s test, 355
function keys see SAS function keys
functions, SAS, 134–9, 472ff
FUZZ function, 483

**G**
GAXIS option, 400
general linear model (GLM) see PROC GLM
goodness-of-fit analysis, 246–53, 492
GOPTIONS statement, 395, 403, 407, 408
graphs, 393ff
Graph template language (GTL), 428, 429
GRID option, 397
grouping factor, 220, 237, 311, 327, 340 ff
GROUP option (GCHART), 405
GROUP option (REPORT), 448
GTL see Graph template language (GTL)

**H**
HAXIS option, 397, 400
HBAR3D option (GCHART), 405
HBAR option (GCHART), 405, 421
HBOUND (arrays), 172
help, 15
HISTOGRAM (PROC UNIVARIATE), 235
HISTOGRAM (SGPLOT), 421ff
HMS function, 478
homogeneity hypothesis test, 492
horizontal bar charts (HBARs), 405
 Hosmer and Lemeshow test, 364, 373
HOUR function, 478
HREF option (GCHART), 397, 406
HTML= option (GCHART), 406, 418ff
hyperlinks see HyperText Markup Language (HTML)
HyperText Markup Language (HTML), 194

**I**
ID statement, 122
IF-THEN-ELSE statements, 82–8
IF-THEN for missing values, 84
IF-THEN to subset data sets, 85
IF to conditionally read a data set, 87
importing data, 60
importing data use code see PROC IMPORT
importing data using import wizard, 60
INCLUDE=k statement, 364
%INCLUDE statement, 186, 357
independent variables, 291, 297, 362
INDEX function, 481
INFILE statement, 37, 42, 495
informs, 33
INPUT function, 155, 485
INSET statement, 225, 238, 405
INSIDE= option (GCHART), 406
INTCK function, 136–7, 478
interaction effects, 333–5
INTERPOLATION= option (I=), 398, 399, 403
interquartile range, 417
inter-rater reliability see kappa
INT function, 483
INTNX function, 478
I= option see INTERPOLATION= option (I=)
IRR function, 484

J
JOBSCORE Example, 298ff
JUSTIFICATION= option, 401
JUST= option, 470

K
kappa, 245, 262–7, 492
KEEP statement, 88
KEYLEGEND option (SGPLOTS), 421
Kruskal–Wallis, multiple comparisons, 357
Kruskal–Wallis (KW) test, 350, 492
Kurtosis, 217, 230

L
labeling variables, 76–8
LABEL= option, 402
LABEL statement, 123, 495
LACKFIT option, 364, 371
LAG function, 486, 488
landscape orientation see ORIENTATION= option
LAST value, find, 107
launching SAS, 5
LBOUND (arrays), 172
LEFT function, 481
LEGEND= option, 397
LENGTH function, 481
libraries see SAS library
library nicknames, 54, 58
linear regression, 492
multiple, 297ff
simple, 289ff
LINE= option, 398, 399, 407, 408
line plots, 393, 404
LINESIZE option, 131
LINE style see LINE= option
LISTING output destination, 194
log file, 15
LOG function, 473
logistic analysis
multiple, 369
simple, 365
logistic equation, 361, 368
logistic regression, 360, 493
logistic regression, graphs, 367ff
Log window, 6
Lotus 1-2-3 files, 65
LOWCASE function, 481
LRECL option, 41
LSD multiple comparisons, 314
LSMEANS statement, 313, 314, 322, 344

M
macro routine, %INCLUDE, 186
macro routines, 182ff
macros, SAS see SAS macros
macro variable, creating with CALL SYMPUT, 189
macro variables, 178–82
main effects, 333
MAJOR= option, 401
Mann–Whitney Test see Wilcoxon–Mann–Whitney
matrix of histograms, 228–39
matrix of scatterplots, 287–8
MAXDEC=n option, 180–181
MAX function, 136, 473
maximum likelihood estimates, 368
MAXIS option, 400
McNemar test, 492
MDY function, 137, 478
mean bars (GCHART), 412ff
MEAN function, 473
MEANS statement (multiple comparisons), 314
MEDIAN function, 136, 473
median test, 348, 359
menus see SAS menus
merge, few-to-many, 98
merging data sets, 94, 96–100
messy data, cleaning, 146–59
Microsoft Access, 23, 60, 65
Microsoft Excel see Excel
MIDPOINTS option, 226
MINEIGEN=n option, 378
MIN function, 136, 473
MINOR= option, 401
MINUTE function, 478
missing values, assign, 84
MISSOVER option, 41, 45
MIXED model see PROC MIXED
mode, 217
model I ANOVA, 334
model II ANOVA, 334
model III ANOVA, 334, 335
MODEL statement, 292, 297, 312–13, 329, 362
MOD function, 473
moments, 229
MONNAME format, 487
MONTH function, 478
month name format see MONNAME
MORT function, 484
MS Access see access
MS Excel see Excel
multiple comparisons, 314
multiple linear regression, 297ff
MUO= option (Univariate), 225

N
_N_, 70
naming conventions, 21, 23
NFACTORS option, 377
N function, 136, 473
INDEX 515

#n indicator, 39
NMISS function, 136, 473
NOCENTER option, 131
NODATE option, 131
NOFRAME option, 397
NOLEGEND option, 397, 406
nonparametric analysis, 347ff
nonparametric multiple comparisons, 355
NONUMBER option, 131
NOPRINT option, 313
NOPROB option, 286
N option, 126
normal curve on graph, 226, 421
normality, test for, 234
normal probability plot, 232
NOSIMPLE option, 284
NOTCHES = option (BOXPLOT), 415
NPV function, 484
numeric variables, 24

O
OBS= option, 42, 128
Odds ratios (OR), 360ff
ODDS RATIO statement, 363
ODS CLOSE, 195
ODS EXCLUDE statement, 198–201
ODS GRAPHICS, 204–7, 401
OFFSET = option, 401
olympic data analysis, 384ff
one-sample t-test, 270–74
one-way ANOVA, 312ff, 492
open VMS, 4
options see system options
OPTIONS FMTSEARCH, 106
OPTIONS NOFMTERR, 106
OR see odds ratios (OR)
ORDER = option, 313, 401, 432
ORDER option (REPORT), 445, 446
ordinal data, 491–3
ORDINAL function, 473
ORIENTATION = option, 131
output delivery system (ODS), 193ff, 495
destinations, 193–4
interactive plot, 418, 429
OUTPUT statement, 201–4
PREFERENCES statement, 195
reset defaults see output delivery system (ODS),
   PREFERENCES statement
SELECT statement, 198–201
STYLE attributes, 469, 470
styles, 196
TAGSETS, 443
OUTPUT OUT= statement, 218, 363
output window, 13–14
OUTROC = option, 364
OUTSIDE = option (GCHART), 406

P
PAD option, 41, 45
PAGESIZE= option, 131
paired (dependent) samples, 492
paired t-test, 279–81, 492
pairwise comparisons, 314, 320, 324, 330, 352, 356–7
patterns for GPLOT/PROC UNIVARIATE, 407ff
PATTERN statement, 394, 405, 407, 464–7
PDIFF multiple comparisons, 314
PDV see program data vector
Pearson correlation coefficient (r), 283ff, 492
permanent SAS data sets and, 51–3, 59
PFILL statement, 238
pie charts, 405ff
PIE3D option (GCHART), 405
PIE option (GCHART), 405
PLOTS option, 313, 363
PLOT statement, 344, 394ff, 405
plotting symbols, 469
pointer control commands, 41
Portable Document Format (PDF), 194
portrait orientation see ORIENTATION = option
Postscript (PS) see Postscript (PS) output destination
Postscript (PS) output destination, 194
predicted values
linear regression, 294, 303–6
logistic regression, 371ff
prediction bands, linear regression, 294
predictions, linear regression, 303–6
principal components analysis, 378
Print Command Language (PCL), 194
PRIORS =option, 377
PROBSIG = option, 131
PROC ANOVA, 312ff, 497
PROC APPEND, 100
PROC BOXPLOT, 279, 414, 498
PROC CATALOG, 106
PROC CONTENTS, 67–8
PROC CORR, 283–9, 497
PROC DATASETS, 67
PROC DOCUMENT, 209
PROC EXPORT, 66
PROC FACTOR, 377ff, 498
PROC FORMAT, 243ff, 356, 496
PROC GCHART, 405ff, 498
PROC GLM, 312, 318–25, 329, 333, 335, 337, 497
PROC GPLOT, 344, 394, 397, 498
PROC GSLIDE, 185
PROC IMPORT, 64
PROC LOGISTIC, 377ff, 498
PROC MEANS, 215ff, 344, 496
PROC MIXED, 337, 339ff, 497
PROC NPAR1WAY, 347, 497
PROC PLOT, 207–9
PROC PRINT, 125–32, 497
PROC PRINT Traffic Lighting, 207–9
PROC REG, 289ff, 497
PROC REPORT, 444ff, 497
INDEX

PROC SGPLOT, 421ff
PROC SORT, 91, 344, 496
PROC SQL, 100
PROC statement syntax, 118–19
PROC TABULATE, 431ff, 498
PROC TEMPLATE, 196
PROC TRANSPOSE, 139–43
PROC TTEST, 270–80, 496
PROC UNIVARIATE, 224ff, 354, 465
PROC UNIVARIATE INSET, 228–9, 271–3, 496
PROC code see code files
program data vector, 68–71
program code
program data vector, 68–71
PROMAX option, 378
PUT function, 485
PUT statement
PUT options, 457
PUT reports, 455, 499
P-values in multiple comparisons, 314
Q
Q–Q plot, 240–241, 274, 279
QTR function, 478
quantiles, 231
quantitative data, 215 ff, 283–4, 312, 327, 334, 361, 491–3
QUARTIMAX option, 378
QUIT statement, 118, 296
QUOTE function, 481
R
random factors, 333, 337–40
random numbers, creating, 473
range, 230
RANNOR function, 473
RANUNI function, 473
RAXIS option, 400
reading data into SAS, 24, 50
reading/writing data, 50
receiver operating characteristic (ROC) see ROC
curve
references, 508
REG option (SGPLOT), 421
regression
linear, 492
logistic, 493
relational analyses, choosing, 493
relative risk (PROC FREQ), 245, 261–4
RELRISK risk measure, 245, 261
repeated measures, 492
ANOVA, 319ff
ANOVA with grouping factor, 340
REPEATED statement, 313, 342
REPEAT function, 481
residual analysis, regression, 306–8
residual by predicted value plot, 294
residual by quartile plot, 295
residuals by percent plot, 295
response profile table, 366
results viewer, 6
RETAIN statement, 173ff
REVERSE function, 481
Rich Text Format (RTF), 126, 131, 194
RIGHT function, 481
RISKLIMITS option, 364, 370, 372
ROC curve, 364, 374
ROTATE= option, 377, 382
ROUND function, 136, 483
R-Square, 293
RStudent by Leverage Plot, 295
RStudent by Predicted Value Plot, 294
RTF see Rich Text Format (RTF)
running man icon, 7, 17
RUN statement, 118
S
SAS7BDAT file, 15, 21
SAS Colors, 463, 464
SASDATA folder, 4
SAS data sets, 20
SAS fonts, 462
SAS function keys, 13
SAS graphs, 393ff
SAS help, 15
SAS library, 53
SAS macros, 177ff, 357
SAS menus, 14
SAS University Edition, 501
SAS Windows Interface, 4
Save As command, 8–9, 17
SCAN function, 481
SCANOVER option, 42
SCATTERPLOT option (SGPLOT), 421, 426, 427
scatterplots, 393, 396, 399
Scheffe multiple comparisons, 314, 331
SCHEMATIC boxplot option, 416, 417
SCREE plot and option, 377, 381, 384, 388
SECOND function, 478
selecting variables (regression), 296, 301–3, 364, 369
SELECTION= statement, 298, 364
SELECT statement, 144 ff
sensitivity, 374
SET statement, 89
signed rank test, 231, 353, 492
SIGN function, 473
sign test, 231, 492
simple linear regression, 289ff
SIN function, 476
SINH function, 476
skewness, 217
SLENTRY= option, 298, 363, 370
slope, test, 290
SLENTRY= option, 298, 363, 370
SMC (factor analysis), 380, 381
SNK multiple comparisons, 314
sorting data see PROC SORT
SOUNDEX function, 481
SPACE= option (GCHART), 406
Spearman correlations, 284–6, 347, 492
special use functions, 485
specificity, 374
SPLIT= statement, 128–9
splitting titles, 128–30
SQRT function, 473
squared multiple correlations see SMC (factor analysis)
square root function (SQRT) see SQRT function
stacked bar charts, 410
standard deviation (STDDEV), 217
STAR option (GCHART), 405
STATISTICNAME option (REPORT), 445
stem-and-leaf plot, 225, 232–4
STEPWISE procedure, 301–3
STOPOVER option, 42
Student’s t-test, 217, 231, 270 ff
STYLE= option, 432
STYLE option (ODS), 196
SUBGROUP option (GCHART), 406
subsetting data sets, 85 ff
SUBSTR function, 155, 481
SUM function, 136, 472–3
summarized frequencies (PROC FREQ), 249–52
summarized two-way frequencies, 257
SUM statement, 81
SUMVAR option (GCHART), 406
SYMBOLn = option, 344, 397, 399, 403, 469
SYMPUT, 448
SYMPUT, SYMPUTN, 486
syntax for PROC statements, 118–19
system options, 131

T
TABLES statement, 244, 432
TAGSETS.EXCELXP output destination, 194
TAN function, 476
TANH function, 476
temporary SAS data sets, 51
test for location, 354
time function, 478
timepart function, 478
tips and tricks, 11
TITLE statement, 113, 462, 463, 495
TODAY function, 478
Traffic lighting (PROC PRINT), 207–9
TRANSLATE function, 481
TRANSPARANCY= option (SGPLOT), 421
transposing data sets see PROC TRANSPOSE
TRANWRD option, 481
trigonometric functions, 476
TRIM, TRIMN functions, 481
truncation functions, 483
TRUNCOVER option, 41, 45
T-tests see also Student’s t-test
one-sample, 270–274	paired, 279–81
two-sample, 274–9
TUKEY multiple comparisons, 314, 322
two factor ANOVA, 333, 337
two independent samples, comparing, 347
two-sample t-test, 274–9, 492
two-way ANOVA using, 274 ff
two-way tables (PROC FREQ), 253–60
TYPE= option (GCHART), 406

U
uncorrected sum of squares (USS), 217
UN covariance structure see unstructured (UN) covariance structure
UNIFORM option, 394
University Edition, SAS, 501
UNIX, 4
unstructured (UN) covariance structure, 342
UPCASE function, 481

V
VALUE= option (V=), 398, 401
Van der Waerden test, 348
variable name rules, 23
variable selection (regression), 296, 301–3
variable types, 24
variance components (VC) covariance structure, 342
VARIMAX option, 378
VAR statement, 120
VAXIS option, 400
VBAR3D option (GCHART), 405
VBAR option (GCHART), 405, 421
VC covariance structure see variance components (VC) covariance structure
VERIFY function, 481
viewing data, 57
viewtable, 57
vocabulary data analysis, 377ff
VREF option, 397
VSCHOOL option, 226

W
Washington, DC, crime analysis, 308–9
WAXIS option, 226
WBARLINE option, 226
WEEKDAY function, 478
weighted kappa, 264–7
WEIGHT statement, 244
WHERE statement, 124
WIDTH= option, 401, 405
Wilcoxon–Mann–Whitney, 348, 492
WILCOXON option, 348
windows interface see SAS windows interface
Wireless Markup Language (WML), 194
WITH statement, 289
WML see Wireless Markup Language (WML)
WOUTLINE option, 405
WPGM see enhanced editor
XLS and XLSX files, 65
XML output destination see Extensible Markup Language (XML) output destination

Yate's chi-square, 259
YEARCUTOFF= option, 131
YEAR function, 478, 488

YRDIF function, 139, 478
YYQ function, 478

Z
ZIPCITYDISTANCE function, 486, 489
ZIPCITY function, 486, 489
ZIPSTATE function, 486, 489
Z/OS, 4