

INDEX

A

- aborting SAS processes, 11
- Active Libraries window, 58, 64–65
- advanced programming: arrays and DO loops, 357–364; Dynamic Data Exchange, 352–357; RETAIN statement, 357–359; SAS macros, 369–372; transposing data sets, 365–369
- AIC (Akaike Information Criteria), 299–300
- ALPHA= option, PROC UNIVARIATE, 138
- Analysis: of counts and tables. *see* counts and tables; hands-on example, 9; preparing data for. *see* data, preparing for analysis; running first, 5–8
- analysis of covariance (ANCOVA): example of, 279–284; overview of, 278–279
- analysis of variance (ANOVA), 202–216, 277–304; analysis of covariance, 278–284; graphing means comparison, 213–215; one-way, comparing three or more means, 203–209; output for PROC REG, 232; repeated measures with grouping factor, 296–303; repeated measures with multiple comparisons, 209–212; two-way, using PROC GLM, 284–293; two-way, using PROC MIXED, 293–295
- appending (combining) data sets, 97–99
- area under the curve (AUC), 274, 276
- arithmetic functions, SAS reference, 386–387
- ARRAY statement: example of, 362–363; overview of, 359–360; using DO loops, 360–362; using DO UNTIL loop, 363–364; using DO WHILE loop, 364
- ASCII file/format, 4, 15
- ASE (Asymptotic Standard Error), 180–181
- asterisk (*), comments, 115
- Asymptotic Standard Error (ASE), 180–181
- at least ordinal (choosing SAS procedure), 397–401
- AUC (area under the curve), multiple logistic regression, 274, 276
- automated model selection, 231–232
- axes: customizing, 310–313; modifying bar chart, 317–319
- AXIS statement, 310–313

B

- backslash (/), entering multiple-line data, 42
- BACKWARD procedure, MODEL statement: automated model selection, 233–234; multiple linear regression, 231; multiple logistic analysis, 270; simple logistic model, 263–264
- bar characteristics, HISTOGRAM statement, 141–142
- bar charts: ; bar and block patterns for, 380–381; creating, 315–319; with error bars representing means by group, 321–323; stacked, 319–321
- blanks: column input and, 31–32; freeform list input and, 27; in SAS statements, 11; SAS variable names and, 23–24
- Bonferroni test, 282
- box-and-whiskers plot, 149
- boxplots: changing appearance, 325–326; creating simple, 323–325; means comparison using, 197
- BY DESCENDING statement, PROC SORT, 95
- BY group option, PROC MEANS, 136–137
- BY statement, SAS procedures, 118–120
- BY *variable(s)* statement: defined, 117; PROC FREQ, 160; PROC PTEST, 194; PROC UNIVARIATE, 139

C

- case sensitivity: SAS statements, 11; sorting with PROC SORT, 96; variable name rules, 23–24
- CENTER option, 111
- central limit theorem, one-sample t-tests, 188
- CFILL statement, PROC UNIVARIATE, 154

character (text, string) variables: entering data using column input, 31–32; overview of, 24; using freeform list input, 27–28

character functions, SAS reference, 391–393

chi-square test, 171–176, 271–275

CLASS option, PROC MEANS, 136–137

CLASS statement: logistic analysis, 262–264; multiple logistic analysis, 270; PROC PTEST, 194–195

CLASS *variable* statement: one-way ANOVA, 204; PROC MEANS, 133–134; PROC UNIVARIATE, 139, 153; repeated measures ANOVA, 210

Clear All command, windows, 8

CLI statement, PROC REG, 230

clipboard, reading Excel data from Windows, 352–354

CLM statement, PROC REG, 230

COCHRAN option, PROC PTEST, 194

code files, SAS: creating SAS library using, 64–65; importing Excel data using, 65–67, 71–72; including comments in, 115; SAS data sets vs., 51; *.sas* extension for, 15

coef variation, PROC UNIVARIATE, 143

color options: AXIS statement, 311–313; SAS Enhanced Editor, 12; using HISTOGRAM statement of PROC; UNIVARIATE, 140–141; using PATTERN statement, 380–381

column input, entering data, 30–33, 44–45

columns, SAS data set, 23

comments, 115

compact method, data entry, 28–30

comparison of means, graphing, 330

comparison tests, decision table for, 399–400

COMPRESS option, PROC FREQ, 160

constant values, creating new variables as, 80

contingency table analysis: calculating relative risk measures, 177–178; creating from raw data, 175–176

Continuity Adj. Chi-Square, 176

Cook's D statistic, 239

Copy, Graph window, 307

corrected SS, PROC UNIVARIATE, 143

correlation analysis, 217–244: basics of, 218; calculating using WITH statement, 222; example of, 219–220; producing matrix of scatterplots, 220–221; using PROC CORR for, 218–219

counts and tables, 159–186: inter-rater reliability (kappa), 178–181; one-way frequency tables, 163–166; one-way frequency tables, creating from summarized data, 167–171; relative risk measures, 177–178; two-way tables, 171–176; using PROC FREQ, 160–163; weighted kappa, 181–184

covariance, analysis of (ANCOVA), 278–284

cross-tabulation tables: analyzing, 171–172; calculating weighted Kappa, 181–184; creating contingency table from raw data, 175–176; example of, 173–174

C: \SASDATA folder, 2–3

CSV (comma separated values): *.csv* extension, 15; data entry method for, 44

Ctrl+Break, aborting SAS processes, 11

D

.dat extension, 15

Data: evaluating quantitative. *see* quantitative data, evaluating: *.sas7bdat* extension, 15

DATA = *datasetname* option: PROC ANOVA, 204; PROC CORR, 219; PROC REG, 230

data entry, 19–49: SAS data sets, structure, 22–23; SAS data sets, using, 20–22; SAS variables, name rules, 23–24; SAS variables, types of, 24

data entry, methods: column input, 30–33; compact method, 28–30; CSV (comma separated values), 44; formatted input, 33–39; freeform list input, 25–28; importing Excel file. *see* Microsoft Excel files; INFILE technique, 39–41, 43–44; multiple-line data, 41–43

DATA option, PROC PRINT, 124

data preparation, for analysis: appending data sets, 97–99; creating new variables. *see* variables, creating new; creating subset of data set, 90–91; labeling variables with explanatory names, 74–76; merging data sets, 99–101; using DROP and KEEP to select variables, 89–90; using PROC FORMAT, 101–106; using PROC SORT, 94–97; using SET to read existing data set, 92–94

data sets, SAS. *see also* permanent SAS data sets: appending as mean of combining, 97–99; creating, 20–22; creating subset

- of, 90–91; discovering contents, 67–69; displaying contents, 58; merging, 99–101; methods for creating. *see* data entry, methods; SAS variable name rules, 23–24; sorting with PROC SORT, 94–97; structure, 22–23; transposing, 365–369; using SET to read existing, 92–94
 - data source: reading or importing data into SAS, 69–70; SAS DBMS, 66
 - DATA statement: creating SAS data set, 21–22, 51; entering data using freeform list input, 25–28
 - DATA step: appending data sets, 97–99; assigning missing values, 87–89; creating variables in, 77–80; labeling variables, 74–76; manipulating data within, 404; merging data sets, 99–101; overview of, 20; understanding, 9–10; using DROP and KEEP to select variables, 89–90; using PROC SORT, 94–97; using RETAIN, 356
 - data types, choosing SAS procedure, 397–401
 - DATA= datasetname option, PROC PTEST, 193
 - DATA= option: PROC FREQ, 160; PROC LOGISTIC, 263; PROC MEANS, 131; PROC statement syntax, 116; PROC UNIVARIATE, 138
 - DATALINES statement: defined, 39; freeform list input, 26; INFILE statement eliminating, 40
 - Date: DATE option, 111; date variables, 24; functions, 388–391; INFORMATs for reading, FORMATS for outputting, 38
 - DDE (Dynamic Data Exchange), 352–357, 372–373
 - dependent (paired) samples, 251–255
 - DESCENDING option, PROC LOGISTIC, 263, 270
 - descriptive statistics, decision table for, 398
 - destination, ODS output format and, 332–333
 - dialog box, creating SAS library, 56–59
 - DIFn function, 396
 - DLM option, INFILE statement, 43–44, 48–49
 - DO loops: ARRAY statement and, 360–362; DO UNTIL loop, 363–364; DO WHILE loop, 364; example, 362–363
 - DOCUMENT output format, ODS, 333
 - dot (.), INFORMATs and FORMATS, 34
 - DOUBLE option, PROC PRINT, 124
 - DROP statement, 89–90, 107–108
 - DSD option, INFILE statement, 43
 - Dynamic Data Exchange (DDE), 352–357, 372–373
- ## E
- Edit Current Graph, Graph window, 308
 - Edit menu, 15
 - Editor window: clearing, 8; menus, 15; as most-used SAS window, 14; purpose of, 3–4; running first analysis, 5–9; using Save As in, 8
 - Enhanced Editor window. *see* Editor window
 - equal variances, two-sample t-tests, 193
 - error bars, 313–315, 321–323
 - estimates, simple linear regression, 227
 - EXACT statement, PROC FREQ, 160
 - Exact Test p-values, 249
 - Excel. *see* Microsoft Excel files
 - Exit command, 9
 - Export, Graph window, 307–308
 - expressions. *see also* functions: date and time, 388–391; logical, 77; mathematical, 77, 386–387
 - external data, reading using INFILE, 39–41
 - extreme observations table, from PROC UNIVARIATE, 145–146
- ## F
- factors, fixed and random, 285
 - file extensions: common, 15–16; using Save As, 8
 - File menu, 15
 - FILENAME command, DDE, 353, 356
 - financial functions, SAS reference, 395–396
 - FIRSTOBS= option, INFILE statement, 43–44, 49
 - fixed factors, 285
 - \$FMTRACE format, PROC FREQ, 166
 - folders, creating, 2–3
 - fonts, graph reference guide, 377–378
 - FOOTNOTE statement: using, 112–114
 - FORMAT statement. *see also* PROC FORMAT: advantages and restrictions, 35–36; outputting or printing out data, 36–39; overview of, 33–35; reading data using, 403
 - formats: easy to read program structure, 11–12; specifying ODS output, 332–333

formatted input: advantages and restrictions, 35–36; data entry using, 33–35; using `FORMATs` for output, 36–39

`FORWARD` procedure, `MODEL` statement: multiple linear regression, 231; multiple logistic analysis, 270; simple logistic model, 263–264

freeform list input, 25–27, 31

`FREQ` *variable* variable statement, `PROC MEANS`, 133–134

frequency tables, one-way: analyzing, 163–166; creating with summarized data, 167–168; testing goodness-of-fit, 168–171

Friedman's test, 253–255

F-test, 193, 278–281

function keys, 5, 12–14

functions, 384–396: arithmetic/mathematical, 386–387; character, 391–393; date and time, 388–391; defined, 385; financial, 395–396; miscellaneous, 396; special use, 394–395; trigonometric, 387–388; truncation, 394; using, 385; working with previous observations, 396

functions, creating new variables, 81–85

G

`GCHART`. *see* `PROC GCHART`

General Linear Model (GLM), `PROC GLM`, 203

getting started, 1–18: common file extensions, 15–16; getting Help on web, 16; tips and tricks, 9–12; understanding how it works, 9–10; using function keys, 12–14; using in Windows, 2–5; using menus, 14–15; using SAS Enhanced Editor, 12; your first analysis, 5–9

GLM (General Linear Model), `PROC GLM`, 203

goodness-of-fit analysis, 168–171, 184–185

`GOPTIONS` statement, 112, 306

`GPLOT`. *see* `PROC GPLOT`

graph options, SAS reference, 377–384: bar and block patterns for bar/pie charts, 380–381; color choices, 378; fonts, 377–378; line styles, 381–383; patterns for `GPLOT/PROC UNIVARIATE`, 379–380; plotting symbols, 383–384

Graph window, 4–5, 306–307

Graphics: bar and block patterns for, 380–381; ODS enhancing, 340–344

Graphing: analysis of covariance, 282; matrix of scatterplots, 220–221; means comparison, 213–215; normality using `PROC UNIVARIATE`, 147–148; one-sample t-test, 191–192; paired t-test, 199–200; repeated measures with grouping factor, 301–303; simple linear regression, 227–228; simple logistic results, 266–268; two-sample t-test, 196–197

graphs, creating, 305–330: bar charts, 315–319; boxplots, 323–326; means using `GCHART`, 321–323; ODS enhancing, 326–330, 340–344; scatterplots and line graphs. *see* `PROC GPLOT`; stacked bar charts, 319–321 /`GROUP` option, `PROC GCHART`, bar charts, 317–319

grouping factor, repeated measures with, 296–303

H

`H0`-value option, `PROC PTEST`, 194

`HAXIS-AXISk` option, `PLOT`, 311–312

`HBARs` (horizontal bar charts), 315–317

Help, 15–16

`HISTOGRAM` statement, `PROC UNIVARIATE`, 139–142, 154

Histograms: creating multiple, 151–152, 156–158; creating using `HISTOGRAM AGE/NORMAL`, 150; formatting options, 153–155; specifying patterns, 379–380

horizontal bar charts (`HBARs`), 315–317

Hosmer and Lemeshow test, 271–275

HTML (HyperText Markup Language), ODS output format, 332

hyperlinks, creating using ODS, 340–344

HyperText Markup Language (HTML), ODS output format, 332

I

`I=` statement, 313–315

`ID` statement, 120–121, 128

`ID` *variable(s)*, 117, 230

`IF` statement: assigning missing values, 87–89; creating new variable, 86–87; subsetting data sets, 90–91

`IF-THEN-ELSE` statement, 85–87

Import Wizard, 59–63

- Importing: data from another program, 59–63;
 Microsoft Excel data using SAS code, 65–67
- INCLUDE=*k* statement, PROC REG, 230
- independent groups: comparing *K*, 249–251;
 comparing using NPAR1WAY, 246–249
- INFILE statement: advanced options, 43–44;
 Dynamic Data Exchange (DDE), 353;
 entering data using, 39–41
- INFORMAT statement: advantages and
 restrictions, 35–36; entering data using
 formatted input, 33–35; examples, 37–39
- INPUT statement: @@ option, 29–30; column
 input, 30–33; compact method, 29–30;
 Dynamic Data Exchange (DDE), 353;
 formatted input, 34–38; freeform list input,
 25–27; multiple-line data, 41–43
- INSET option, PROC UNIVARIATE, 154
- interaction effects, two factor design, 284–288,
 291–293
- interface, SAS Windows, 3–5
- interpolation curves, scatterplots, 308–310
- interquartile range (IRQ), PROC
 UNIVARIATE, 144
- inter-rater reliability analysis, 178–181
- IRQ (interquartile range), PROC
 UNIVARIATE, 144
- K**
- K* dependent (paired) samples, comparing,
 253–255
- K* independent samples, comparing, 249–251
- Kappa statistic: calculating inter-rater
 reliability analysis, 178–181; calculating
 weighted Kappa, 181–184
- KEEP statement, 89–90, 107–108
- key identifier, merging data sets by, 97, 99–101
- Kruskal-Wallis (KW) test, 249–251, 256–258
- kurtosis, PROC UNIVARIATE, 143
- L**
- LABEL option, PROC PRINT, 124
- LABEL statement, variables, 74–76, 121–122,
 128
- LABEL var=*'label'*, 117
- LABEL= option, AXIS statement, 311–313
- LACKFIT option, MODEL statement, 271–275
- LAG*n* function, 396
- launching SAS, 3, 5–9
- libraries, SAS: creating using dialog box,
 56–59; creating using program code, 64–65;
 defined, 55; Explorer tab displaying, 3–4;
 permanent SAS data sets and, 51–52, 55–56;
 storing imported data from other programs
 as, 61–63
- Libraries icon, 57
- line graphs. *see* PROC GPLOT
- line styles, 381–383
- linear regression. *see also* simple linear
 regression: calculating predictions, 234–237;
 multiple, 229–234, 243–244; residual
 analysis, 237–241
- listing files, .lst extension, 15
- LISTING output format, ODS, 332, 333
 .log extension, 15
- Log window: clearing, 8; discovering table
 names, 338; purpose of, 3–4; using Save
 As in, 8
- logical expressions, 77
- logistic regression, 260–276; assessing model's
 fit and predictive ability, 271–275; graphing
 simple logistic results, 266–268; model
 (logit model), 261–262; multiple binary model,
 268–271; odds and odds ratios (OR), 262;
 performing using PROC LOGISTIC, 262–264;
 simple logistic model, 264–266
- logit model, 261
- Lotus 1-2-3 files, 16
- LS = # (LINESIZE) option, 111
- LSMEANS statement, 301, 303
 .lst extension, 15
- M**
- macros, SAS: calculating multiple comparisons,
 255–258; creating and using, 369–370;
 creating callable routine, 370–372; defined,
 369; example, 372
- main effects, two-factor design, 284, 286
- MARKUP output format, ODS, 333
- mathematical functions, SAS reference,
 386–387
- MAXDEC=*n* option, PROC MEANS, 131, 135
- Maximum Likelihood Estimates table, logistic
 regression, 265–267
 .mdb extension, 16

mean, PROC UNIVARIATE, 142, 144
means, displaying as error bars, 321–323
means comparison: graphing, 213–215; using ANOVA. *see* analysis of variance (ANOVA); using t-tests. *see* t-tests
MEANS statement, PROC ANOVA: example, 206–209; one-way ANOVA using, 204–206; repeated measures ANOVA, 210
median, PROC UNIVARIATE, 144
median test, 258
menus, 14–15
MERGE statement, data sets, 99–101
merging data sets, 97, 99–101
Microsoft Access, .mdb extension, 16
Microsoft Excel files: importing data, using DDE. *see* Dynamic Data Exchange (DDE); importing data, using Import Wizard, 59–63; importing data, using SAS code, 65–67; SAS data set structure vs., 23; .xls extension, 15
MIDPOINTS option, HISTOGRAM, 140
miscellaneous functions, SAS reference, 396
MISSOVER option, INFILE statement, 43
MLR. *see* multiple linear regression (MLR)
mode, PROC UNIVARIATE, 144
MODEL statement: analysis of covariance, 280–281; assessing model’s fit and predictive ability, 271–275; automated model selection, 233–234; logistic analysis, 262–264; multiple logistic analysis, 270; PROC ANOVA, 204, 210; PROC LOGISTIC, 263–264; PROC REG, 230–232; simple linear regression, 223–225
Moments output, PROC UNIVARIATE, 142–143
multiple comparisons: not using NPAR1WAY for, 251; one-way ANOVA with, 203–209; performing analysis of covariance, 281–284; repeated measures ANOVA with, 209–212
multiple linear regression (MLR): automated model selection, 233–234; overview of, 229; residual analysis, 237–241; using PROC REG, 229–231
multiple logistic regression, 268–271, 276
multiple-line data, entering, 41–43
MUO= option, PROC UNIVARIATE, 138
MYDATA, assigning to new data set, 21
MYFILE filename, DDE, 353

N

#n indicator, multiple-line data entry, 42–43
N option, PROC PRINT, 124
naming conventions: creating SAS folder storage, 2–3; library nicknames, 55–56; ODS output formats, 333; SAS data sets, 21–22, 55–56; SAS variables, 23–24, 74–76; temporary vs. permanent SAS data sets, 51
New Library dialog box, 56–59
nicknames, library, 55–59, 64–65
NOCENTER option, OPTIONS statement, 111
NODATE option, OPTIONS statement, 111–112
NOFRAME option, PLOT statement, 311–312
nonparametric analysis, 245–259; comparing *K* dependent samples, 253–255; comparing *K* independent samples, 249–251; comparing two dependent samples, 251–253; comparing two independent samples, 246–249; correlation estimates, 218; multiple comparison tests, 255–258; for non-normal data, 188
NONUMBER option, OPTIONS statement, 111
NOPRINT option: PROC MEANS, 131; PROC UNIVARIATE, 138
NOPROB option, PROC CORR, 219
NORMAL option: HISTOGRAM statement, 139; PROC UNIVARIATE, 138, 146–148
normal probability plot, 149–150, 156
normality of data: assessing using PROC UNIVARIATE, 146–148; choosing SAS procedure, 397–401; correlation analysis using Pearson test, 218; performing two-sample t-test, 193
NOSIMPLE option, PROC CORR, 219
Numbers: INFORMATs and FORMATs ending with, 34–35; option to include on printout, 111
numeric variables, 24, 77–80

O

OBS option, PROC PRINT, 124
OBS= option, INFILE statement, 44, 49
odds ratios (OR), 262, 265–266
ODS. *see* Output Delivery System (ODS)
ODS EXCLUDE command, 338
ODS GRAPHICS statements: comparing three or more repeated measures, 211; enhancing

- graphics with, 340–344; GPLOT and, 213–214; matrix of scatterplots, 220–221; overview of, 326–327; PROC ANOVA/PROC GLM and, 327–328; PROC BOXPLOT, 175; PROC CORR and, 329–330; PROC REG and, 240–241; running when doubling graphics type, 191; side-by-side boxplots, 249; two-sample t-test, 195
- ODS SELECT command, 349
- ODS TRACE command: capturing information from ODS tables, 344–347; discovering table names, 338–339
- OFFSET= option, AXIS statement, 311–313
- one-sample t-tests, 188–192
- one-way ANOVA, 203–209, 216
- one-way frequency tables. *see* frequency tables, one-way options: AXIS statement, 311–313; HISTOGRAM statement, PROC UNIVARIATE, 139–142; MODEL statement, PROC LOGISTIC, 263–264; PLOT statement AXIS, 311; PROC ANOVA, 204; PROC CORR, 219; PROC FREQ, 160; PROC LOGISTIC, 263; PROC MEANS, 131; PROC PTEST, 193–194; PROC REG, 230; PROC UNIVARIATE, 138; SYMBOL statement, 308–310
- OPTIONS statement, 111–112
- OR (odds ratios), 262, 265–266
- OR risk measure, PROC FREQ, 177–178
- ORDER= options: AXIS statement, 311–313; PROC ANOVA, 204; PROC FREQ, 160, 164–165
- ORIENTATION= option, OPTION statement, 112
- OUT= options, PROC statements, 116–117
- Output: logistic regression, 265–266; macro, 371; paired t-test, 198–199; PROC FREQ, 164–165; PROC PTEST, 195–196; PROC TRANSPOSE, 368; PROC UNIVARIATE, 142–146; simple linear regression, 226; using ODS to control. *see* Output Delivery System (ODS), controlling output
- Output Delivery System (ODS): overview of, 126–129; viewing data graphically, 289–290
- Output Delivery System (ODS), controlling output, 331–350: capturing information, 344–347; enhancing graphics, 340–344; extended features, 348; selecting specific tables, 338–340; specifying format and destination, 332–333; specifying style, 332–337
- OUTPUT statement: PROC MEANS, 133–134; PROC UNIVARIATE, 139; understanding, 9–10
- Output window, 3–4, 6–8
- OUTPUT= statement, PROC FREQ, 160
- ## P
- P statement, PROC REG, 230
- PAGE option, PROC FREQ, 160
- PAGENO = # option, 112
- paired (dependent) samples, 251–255
- paired t-tests, 197–200
- pairwise comparisons, ANCOVA, 281
- parallelism, ANCOVA, 278–281
- PATTERN statement, graphs, 380–381
- PCL (Print Command Language), ODS, 333
- PDF (Portable Document Format), ODS, 332–333, 348
- Pearson correlation coefficient (r), 218–220, 226
- permanent SAS data sets: creating SAS library, 64–65; naming conventions, 21; reading data from, 54; reading/creating using SAS library, 55–56; reading/creating using Windows file name technique, 52–54; SAS libraries and, 51–52; storing imported data from other programs as, 61–63; temporary vs., 51–52
- PFILL statement, PROC UNIVARIATE, 154
- PFILL=Mabcde option, histograms, 379–380
- pie charts, 315–319, 380–381
- Please Excuse My Dear Aunt Sally mnemonic, 80
- PLOT statement, AXIS statement with, 311–313
- plots: ANCOVA, 282; means comparison, 213–215; one-sample t-test with PROC TTEST, 191–192; paired t-test, 199–200; repeated measures with grouping factor, 301–303; simple linear regression, 227–228; simple logistic results, 266–268; symbols for, 383–384; two-factor analysis using PROC GLM, 289–290; two-factor analysis using PROC GPLOT, 290–293; two-sample t-test with PROC TTEST, 196–197
- PLOTS option, PROC UNIVARIATE, 138, 146–148

- PLOTS=MATRIX option, ODS GRAPHICS, 329–330
- PLOTS=SCATTER option, ODS GRAPHICS, 329–330
- Portable Document Format (PDF), ODS, 332–333, 348
- Postscript (PS), ODS, 333
- POSTTEST variable, ANCOVA, 278–283
- Predictions: calculating, 234–237; logistic regression model, 271–275; running regression analysis, 242–243; simple logistic model variables, 263–264
- PRETEST variable, ANCOVA, 278, 280–281
- previous observations functions, 396
- Print, Graph window, 308
- Print Command Language (PCL), ODS, 333
- PROC ANOVA: comparing three or more means, 203–209; defined, 203; ODS GRAPHICS for, 327–328
- PROC BOXPLOT: changing appearance of, 325–326; creating simple boxplot, 324–325; means comparison, 197
- PROC CORR: calculating correlations, 218, 222; creating matrix of scatterplots, 220–221; example, 219–220; ODS GRAPHICS for, 329–330; options and statements, 218
- PROC DATASETS, 67–69
- PROC DOCUMENT, 348
- PROC FORMAT: creating custom format, 101–106; formatting histogram, 153
- PROC FREQ: analyzing count data, 160–163; analyzing one-way frequency tables, 163–166; analyzing two-way tables, 171–176; calculating relative risk measures, 177–178; calculating weighted Kappa, 181–184; comparing K dependent samples, 253–255; creating contingency table from raw data, 175–176; creating one-way frequency tables from summarized data, 167–168; goodness-of-fit test using, 168–171
- PROC GCHART: creating bar and pie charts, 315–319; creating bar chart with error bars, 321–323; creating means, 321–323; creating stacked bar charts, 319–321
- PROC GLM: defined, 203; ODS GRAPHICS for, 327–328; one-way ANOVA using, 209; performing ANCOVA using, 280–281; repeated measures ANOVA using, 210–212; two-way ANOVA using, 284–293
- PROC GPLOT, 306–315; creating simple scatterplot, 306–307; customizing axes in plot, 310–313; displaying error bars, 313–315; displaying two-factor analysis, 290–293; enhancing using SYMBOL statement, 308–310; graphing means comparison, 213–215; managing Graph window, 306–307; selecting line styles, 381–383; specifying patterns, 379–380
- PROC LOGISTIC: assessing model's fit and predictive ability, 271–275; graphing simple logistic results, 266–268; multiple binary logistic analysis, 268–271; performing logistic analysis, 262–264; simple logistic model, 264–266
- PROC MEANS, 131–137; common statements for, 133–134; example of, 134–135; BY group and CLASS statements using, 136–137; overview of, 131; syntax and options, 131–133; using PROC UNIVARIATE vs., 138, 141
- PROC MIXED: defined, 203; repeated measures with grouping factor using, 296–303; two-way ANOVA using, 293–295
- PROC NPAR1WAY: comparing K independent samples, 249–251; comparing two independent samples, 246–249
- PROC PRINT: entering data using freeform list input, 26–27; outputting labels for variables, 76; using in SAS procedures, 123–125
- PROC REG: examples, 224–227, 231–232; multiple linear regression, 229–231; residual analysis, 237–241; simple linear regression, 224
- PROC SORT: Boxplot; sorting SAS data sets, 94–97
- PROC statements: defined, 110; getting Help on Web, 16; syntax, 116–120; understanding, 9–10
- PROC TEMPLATE, 348
- PROC TRANSPOSE, 365–369; dealing with multiple records per subject, 366–367; overview of, 365
- PROC TTEST: one-sample t-test, 189–192; paired t-test, 197–200; two-sample t-test, 192–197

PROC UNIVARIATE, 138–155; assessing normality of data, 146–150; commonly used statements, 139; creating histograms, 150; creating histograms for comparing subgroups of data, 150–155; HISTOGRAM statement and, 139–142; nonparametric tests for paired data in, 251–253; performing one-sample t-test, 188–192; specifying patterns, 379–380; syntax and options, 138; understanding output, 142–146

procedures, SAS, 109–129; choosing, 397–401; including comments in code, 115; Output Delivery System and, 126–127; PROC statement syntax, 116–120; as PROCs, 110; using ID statement in, 120–121; using LABEL statement, 121–122; using ODS to select output tables for, 338–340; using OPTIONS statement, 111–112; using PROC PRINT statement, 123–125; using RUN to execute statements, 115–116; using TITLE and FOOTNOTE statements, 112–114; using WHERE statement, 123

program code. *see* code files, SAS

PS (Postscript), ODS, 333

PS = # (PAGESIZE) option, 112

p-values: graphing means comparison, 213–215; one-sample t-test, 192; two-sample t-test, 193

Q

Q-Q plot, 149–150, 156

quantiles, from PROC UNIVARIATE, 145

quantitative data, evaluating: using PROC MEANS, 131–137; using PROC UNIVARIATE. *see* PROC UNIVARIATE

quotation marks: avoiding unmatched, 11; subsetting data sets, 90; using in LABEL statement, 74

R

r (Pearson correlation coefficient), 218–220, 226

R statement, PROC REG, 230

random factors, 285

ranges, creating formats, 104–106

raw data files: creating contingency table from, 175–176; file extensions for, 15

reading, writing and importing data, 50–72; creating SAS library using dialog box, 56–59; creating SAS library using program code, 64–65; discovering contents of SAS data set, 67–69; importing data from another program, 59–63; importing Excel data using SAS code, 65–67; reading data from permanent SAS data sets, 54; reading Excel data using DDE, 352–355; reading using format specification, 403; reading/creating permanent SAS data sets, 52–56; summary techniques, 69–70; working with SAS libraries and permanent data sets, 51–52

records, appending data sets to combine, 97–99

regression analysis: ANCOVA using, 278–284; automated model selection, 233–234; *elogistic regression. see* logistic regression; multiple linear regression, 229–234; predictions, 234–237; residual analysis, 237–241; simple linear regression, 222–228

relational analyses, decision table for, 401

relative risk, calculating, 177–178, 185–186

REL RISK risk measure, PROC FREQ, 177–178

repeated measures ANOVA, with multiple comparisons, 209–212

residual analysis, 237–241

Response Profile table, logistic regression, 265–266

Results window, 6–7

RETAIN statement, 357–359

Rich Text Format (RTF), ODS, 127, 129, 333

risk, calculating, 177–178, 185–186

RISKLIMITS option, MODEL statement, 263–264, 270

ROC curve, multiple logistic regression, 272–276

ROUND option, PROC PRINT, 124

rows, SAS data set, 23

R-square, linear regression, 226, 229

RTF (Rich Text Format), ODS, 127, 129, 333

rules, SAS variable names, 23–24

Run menu, 15

RUN statement: ending every program with, 11, 27; entering data using freeform list input, 27; executing statement with, 115; failing to include, 116

“running man” icon, 6

S

SAS Enhanced Editor, 12
 SAS Explorer/Results window, 4
 .sas extension, 15
 SAS icon, 3
 .sas7bdat extension: defined, 15;
 discovering contents of SAS data set, 67–69;
 making available in library, 59; reading/
 writing data, 54
 Save As command, 7–8
 Scatterplots: automated model selection, 237;
 creating, 227–228; creating matrix of,
 220–221; creating simple linear regression
 plot, 227–228
 scatterplots, creating using PROC GPLOT:
 customizing axes, 310–313; displaying error
 bars, 313–315; enhancing using SYMBOL
 statement, 308–310; managing Graph
 window, 306–307
 Scheffé multiple comparisons, 281–284
 SELECTION= options, MODEL statement,
 263–264
 SELECTION= statement, PROC REG, 230
 Semicolons: avoiding misplaced or missing, 11;
 in comments, 115; in LABEL statement, 75
 SET statement, 92–94, 97–99
 sign test: for non-normal data, 188; overview
 of, 251–252; PROC UNIVARIATE, 144–145
 signed rank test, 144–145, 188
 simple linear regression: creating scatterplot,
 227–228; example of, 224–227; model of,
 223–224; overview of, 222–223; residual
 analysis, 237–241; using PROC REG for, 224
 SIMPLE option, PROC REG, 230
 single-factor experiment, 210
 skewness, PROC UNIVARIATE, 143
 SLENTY statement, PROC REG, 230–231
 SLENTY= option, MODEL statement: multiple
 logistic analysis, 270; simple logistic model,
 263–264
 slope, simple linear regression, 227
 SLSTAY statement, PROC REG, 230
 SLSTAY= option, MODEL statement, 263–264
 Solutions menu, 15
 sorting data, 94–97, 106–107
 Spearman correlations, 218–220
 special use functions, SAS reference, 394–395
 stacked bar charts, 319–321

starting SAS program, 3
 statements: SAS procedures, 117–118; SAS tips
 and tricks, 10–11
 statistical measures, PROC UNIVARIATE,
 144–146
 std deviation, PROC UNIVARIATE, 143
 std err mean, PROC UNIVARIATE, 143
 stem-and-leaf plot, 149
 STEPWISE procedure, MODEL statement,
 263–264, 270
 storage, creating folders, 2–3
 student's t-test, PROC UNIVARIATE, 144–145
 styles: ODS output, 332–337; SAS line,
 381–383
 subsetting data sets, 90–91, 107
 sum observations, PROC UNIVARIATE, 142
 SUM option, PROC PRINT, 124–125
 sum weights, PROC UNIVARIATE, 142
 summarized counts, one-way frequency tables
 from, 167–168
 sums-of-squares: comparing three or more
 means using one-way ANOVA, 209;
 comparing three or more repeated measures,
 210; performing ANCOVA using, 280–281
 SYMBOL statement: enhancing scatterplot,
 308–310; plotting symbols, 383–384
 SYMBOLn statement, SAS line styles, 381–383

T

tables, ODS: capturing information from,
 344–347, 349–350; selecting specific output
 for procedures, 338–340
 TABLES statement: analyzing one-way
 frequency tables, 163–166; analyzing
 two-way tables, 171–174; creating
 contingency table from raw data, 175–176;
 creating one-way frequency tables from
 summarized data, 167–171; options,
 161–162; PROC FREQ, 160; sample, 161
 tabs, SAS windows interface, 3–5
 temporary SAS data sets, 51–52, 61–63
 TEST KAPPA option, 179
 tests. *see also* t-tests: goodness-of-fit
 analysis, 168–171, 184–185; homogeneity,
 cross-tabulation tables, 171–172;
 independence, cross-tabulation tables,
 171–172; Kruskal-Wallis (KW) test,

- 249–251, 256–258; location, PROC UNIVARIATE, 144–145; normality, PROC UNIVARIATE, 146–149; Tukey multiple comparison, 208–209, 282–293; Wilcoxon-Mann-Whitney, 246–251
- theoretical line, simple linear regression model, 223
- time and date functions, SAS reference, 388–391
- TITLE statement: using, 112–114
- Tools menu, 15
- totals. *see* sums
- transposing data sets: dealing with multiple records per subject, 366–367; example of, 365–366; overview of, 365
- trigonometric functions, SAS reference, 387–388
- TRIMMED= option, PROC UNIVARIATE, 138
- triplet method, DDE, 354–355
- truncation functions, SAS reference, 394
- t-tests, 187–201: one-sample, 188–192; paired, 197–200; two-sample, 192–197
- Tukey multiple comparisons results, 208–209, 282–293
- 2 x 2 analysis, contingency tables, 174–176
- two-sample t-tests, 192–197
- .txt extension, 15
- ## U
- uncorrected SS, PROC UNIVARIATE, 143
- ## V
- VAR *variable(s)* statement: defined, 117; PROC CORR, 219; PROC MEANS, 133–135; PROC PTEST, 194–195; PROC UNIVARIATE, 139; in SAS procedures, 118
- variables: creating matrix of scatterplots between, 220–221; as data set columns, 23; discovering contents of SAS data set, 67–69; labeling, 74–76; merging data sets by key identifier, 99–101; multiple logistic analysis, 269–271; retaining values across iterations, 357–359; rules for names, 23–24; types of, 24; using DROP and KEEP to select, 89–90
- variables, creating new: as constant values, 80; in DATA step, 77–80; using IF to assign missing values, 87–89; using IF-THEN-ELSE, 85–87; using SAS functions, 81–85
- variance. *see also* analysis of variance (ANOVA): analysis of covariance, 278–284; PROC UNIVARIATE. *see* PROC UNIVARIATE
- VAXIS-AXISk option, PLOT statement, 311–312
- View menu, 15
- Viewtable, 58
- VSCALE option, HISTOGRAM, 140
- ## W
- WAXIS option, HISTOGRAM, 140
- WBARLINE option, HISTOGRAM, 140
- WEIGHT *variable* statement, PROC FREQ, 160
- weighted kappa, calculating, 181–184
- WHERE (expression), 118
- WHERE statement, in SAS procedures, 123
- WIDTH= option, AXIS, 311–313
- Wilcoxon-Mann-Whitney test, 246–251
- Window menu, 15
- windows, SAS interface, 2–5
- Windows file name technique, 52–54
- WITH statement, 222, 244
- WITH *variable(s)* statement, 219, 222
- *.wks extension, Lotus 1-2-3 file, 16
- WML (Wireless Markup Language), ODS, 333
- WPGM (Windows Programming Editor). *see* Editor window
- ## X
- .xls extension, 15
- ## Y
- Yate's Chi-Square, 176
- ## Z
- ZIPCITY function, 396
- ZIPCITYDISTANCE function, 396
- ZIPSTATE function, 396

