

Contents

1	Introduction	1
1.1	About SAS	1
1.1.1	Alternatives to SAS	2
1.1.2	Why SAS, Then?	3
1.2	About This Book	3
1.2.1	Goals	4
1.2.2	How to Use the Book	4

Part I Working with Data in SAS

2	The SAS Environment	9
2.1	The SAS Screen	9
2.2	The Program Editor	9
2.3	SAS Statements	11
2.4	Comments	11
2.5	Quick Demonstration of an SAS Program	12
2.6	Two Types of SAS Programs	13
2.7	Two Kinds of SAS Data	15
2.8	Two Parts to a SAS Data Set	15
2.9	Some Simple SAS Utilities	16
2.10	Getting Help	16
	Problems	17
3	Working with SAS Data	19
3.1	SAS Data Libraries	19
3.2	Two Special SAS Libraries	20
3.3	Three Ways to Browse SAS Data Libraries	21
3.4	Inputting Data into SAS	21
3.5	Reading in Data from the Editor Window	23
3.6	Two Basic INPUT Statements	24
3.6.1	Space-Delimited and Column Input	25

3.7	Reading in Data from External Files	26
3.7.1	The INFILE Statement	26
3.7.2	Formatted INPUT of External Data Sets	28
3.7.3	Informats	29
3.8	Behind the Scenes of a Data Step	30
3.8.1	Deciphering Error Statements.....	31
3.8.2	Error Messages	32
3.8.3	A Few Other Common Errors.....	35
3.9	Notes on Manipulating Data (or How to Tame an Annoying Data Set)	36
3.9.1	Illogically Arrayed Data.....	37
3.10	Data Input Miscellany.....	38
3.11	Importing Excel Spreadsheets	38
	Problems	39
4	Preliminary Procedures	41
4.1	PROC PRINT.....	41
4.2	PROC SORT.....	42
4.3	Enhancing Output: Titles and Footnotes	44
4.4	LABELS	46
4.5	PROC FORMAT and FORMAT.....	47
4.6	ODS	52
	Problems	54
5	Manipulating Data	57
5.1	The SET Statement.....	57
5.2	Using SET to Define and Create New Variables.....	58
5.2.1	Operations.....	58
5.2.2	Functions	59
5.2.3	Example: Deaths Following the Terrorist Attacks of September 11, 2001	60
5.3	Adding (Concatenating) Data Sets	62
5.3.1	Concatenating the September 11 Data Set	63
5.4	Merging Data Sets Using MERGE – BY	63
5.4.1	SORT Before You MERGE	64
5.4.2	Merging the 9/11 Data Set	65
5.5	Conditional Expressions Using IF-THEN-ELSE	67
5.6	Conditional Expressions Using a Restricting IF Statement	72
5.6.1	Restricting Variables Read into a New Data Set	73
5.7	Conditional Expressions with SAS Dates	73
5.7.1	Using Dates to Subset the 9/11 Data.....	73
	Problems	76

Part II Descriptive and Categorical Analysis

6	Descriptive Statistics	79
6.1	PROC MEANS	79
6.2	PROC FREQ	80
6.3	PROC TABULATE	81
6.3.1	Using TABULATE for Surveillance Data	84
	Problems	86
7	Histograms and Plots	91
7.1	Introduction	91
7.2	PROC GCHART for Histograms	91
7.3	PROC GPLOT to Plot Continuous Data	93
	Problems	97
8	Categorical Data Analysis I	99
8.1	Introduction to Categorical Outcomes	99
8.2	Associations	100
8.3	Examining Frequency Tables	101
8.4	Reordering Categorical Variables	103
8.5	Tests of Statistical Significance for Categorical Variables	105
8.5.1	Chi-Square	105
8.5.2	Exact Tests	108
8.5.3	The Mantel–Haenszel Chi-Square	112
8.5.4	The Spearman Correlation Coefficient	113
8.6	Significance vs. Strength	114
	Problems	116
9	Categorical Data Analysis II	119
9.1	Probabilities and Odds	119
9.2	The Odds Ratio	120
9.2.1	Why Epidemiologists Need the Odds Ratio	120
9.2.2	The Disease Odds Ratio	121
9.2.3	The Exposure Odds Ratio	123
9.3	Preterm Labor and Birth Weight Example 1	124
9.4	Confounding	126
9.4.1	Identifying and Controlling Confounding	126
9.5	Controlling for Confounding	128
9.5.1	Controlling Confounding in Study Design	128
9.5.2	Analytic Approaches to Confounding	128
9.6	Preterm Labor and Birth Weight Example 2	129
9.7	Adjusted Odds Ratios	129
9.7.1	Cochran–Mantel–Haenszel Statistic	131
9.7.2	The Mantel–Haenszel Odds Ratio	131
9.8	Summarizing Exploratory Contingency Table Analyses	135
	Problems	135

Part III Continuous Data and Regression

10	Cleaning and Assessing Continuous Data using MEANS, UNIVARIATE, and BOXPLOT	139
10.1	PROC MEANS (Redux)	139
10.2	Review of Some Basic Statistics for Continuous Variables	142
10.2.1	Confidence Intervals	147
10.3	PROC UNIVARIATE	148
10.4	PROC BOXPLOT	153
10.5	In Summary	156
	Problems	156
11	ANOVA	159
11.1	Review of ANOVA	159
11.1.1	Assumptions for ANOVA	161
11.2	Testing Assumptions with MEANS, UNIVARIATE, and BOXPLOT	162
11.3	ANOVA with PROC GLM	163
11.3.1	GLM ANOVA Output	166
11.3.2	Multiple Comparisons	167
11.4	Demonstration of One-Way ANOVA	169
11.5	Accounting for More than 1 Categorical Variable: n-Way ANOVA and Interaction Effects	175
11.6	Interaction and Effect Modification: An Epidemiological Perspective	180
11.6.1	The Conundrum of Interaction	180
11.6.2	Components and Causes	182
11.6.3	Usefulness of the Additive Model	183
11.6.4	A Final Thought on Interaction in Epidemiological Studies	184
	Problems	185
12	Correlation	187
12.1	Assessing Correlation	187
12.2	Assessing Correlation Using PROC CORR	188
	Problems	193
13	Linear Regression	197
13.1	Introduction to Regression	197
13.1.1	Variance Perspective of Regression	199
13.2	PROC REG	200
13.2.1	Regression Results	201
13.2.2	Predicted Values	202
13.2.3	Confidence and Prediction Intervals	202
13.3	Demonstration of PROC REG	202

13.4	Multiple Regression with PROC REG	206
13.5	Interpreting Coefficients	208
13.5.1	Categorical Predictor Variables	208
13.5.2	Demonstration of Multiple Linear Regression	211
	Problems	212
14	Regression Diagnostics	213
14.1	Introduction	213
14.2	Residuals Redux	214
14.2.1	Residual Plots	216
14.3	Outliers (Influential Observations)	217
14.4	Collinearity	217
14.5	Demonstration: Residual Diagnostics for the Fitness Data	219
14.6	A Word About Model Selection	225
14.6.1	SAS Model Selection Tools	225
14.6.2	Problems with Automated Selection Procedures	227
14.6.3	Some Advice on Model Selection	228
	Problems	229
	References	231
	Solutions	233
	Index	253



<http://www.springer.com/978-1-4614-4853-2>

SAS for Epidemiologists

Applications and Methods

DiMaggio, C.

2013, XVII, 258 p., Hardcover

ISBN: 978-1-4614-4853-2