

Cambridge University Press

978-0-521-13007-3 - Data Analysis Using SAS Enterprise Guide

Lawrence S. Meyers, Glenn Gamst and A. J. Guarino

Frontmatter

[More information](#)

Data Analysis Using SAS Enterprise Guide

This book presents the basic procedures for utilizing *SAS Enterprise Guide* to analyze statistical data. *SAS Enterprise Guide* is a graphical user (point-and-click) interface to the main SAS application. Each chapter contains a brief conceptual overview and then guides the reader through concrete step-by-step examples to complete the analyses.

The 11 sections of the book cover a wide range of statistical procedures, including descriptive statistics, correlation and simple regression, *t* tests, one-way chi-squares, data transformations, multiple regression, analysis of variance, analysis of covariance, multivariate analysis of variance, factor analysis, and canonical correlation analysis.

Designed to be used as either a stand-alone resource or an accompaniment to a statistics course, the book offers a detailed path to statistical analysis with *SAS Enterprise Guide* for advanced undergraduate and beginning graduate students, as well as professionals in psychology, education, business, health, social work, sociology, and many other fields.

Lawrence S. Meyers is Professor of Psychology at California State University, Sacramento. He teaches undergraduate and graduate courses in research design, data analysis, data interpretation, testing and measurement, and the history and systems of psychology. He was the coauthor of a textbook on research methods in the 1970s, has recently coauthored books on multivariate research design and analysis of variance, and has more than three dozen publications; some of his relatively recent work has been in areas such as measurement and testing and positive psychology. He received his doctorate from Adelphi University and worked on a National Science Foundation Postdoctoral Fellowship at the University of Texas, Austin and Purdue University.

Glenn Gamst is Professor and Chair of the Psychology Department at the University of La Verne, where he teaches the doctoral advanced statistics sequence. He received his doctorate from the University of Arkansas in experimental psychology. His research interests include the effects of multicultural variables, such as client–therapist ethnic match, client acculturation status and ethnic identity, and therapist cultural competence, on clinical outcomes. Additional research interests focus on conversation memory and discourse processing.

A. J. Guarino is on the faculty at Alabama State University, where he teaches graduate statistics courses in the Psychology Department. He received his bachelor's degree from the University of California, Berkeley, and he earned a doctorate in statistics and research methodologies from the University of Southern California through the Department of Educational Psychology.

Cambridge University Press

978-0-521-13007-3 - Data Analysis Using SAS Enterprise Guide

Lawrence S. Meyers, Glenn Gamst and A. J. Guarino

Frontmatter

[More information](#)

Data Analysis Using SAS *Enterprise Guide*

Lawrence S. Meyers

California State University, Sacramento

Glenn Gamst

University of La Verne

A. J. Guarino

Alabama State University



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press

978-0-521-13007-3 - Data Analysis Using SAS Enterprise Guide

Lawrence S. Meyers, Glenn Gamst and A. J. Guarino

Frontmatter

[More information](#)

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press

32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org

Information on this title: www.cambridge.org/9780521130073

© Lawrence S. Meyers, Glenn Gamst, and A. J. Guarino 2009

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2009

Printed in the United States of America

A catalog record for this publication is available from the British Library

Library of Congress Cataloging in Publication data

Meyers, Lawrence S.

Data analysis using *SAS Enterprise Guide* / Lawrence S. Meyers, Glenn
Gamst, A. J. Guarino.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-521-11268-0 (hbk.) – ISBN 978-0-521-13007-3 (pbk.)

1. SAS (Computer file) 2. Enterprise guide. 3. Social sciences – Statistical methods – Data processing.

4. Mathematical statistics – Data processing. I. Gamst, Glenn. II. Guarino, A. J. III. Title.

HA32.M499 2009

330.0285'555–dc22 2009012822

ISBN 978-0-521-11268-0 Hardback

ISBN 978-0-521-13007-3 Paperback

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party Internet Web sites referred to in
this publication and does not guarantee that any content on such Web sites is,
or will remain, accurate or appropriate.

Contents

Preface	xv
Acknowledgments	xix
I Introducing SAS Enterprise Guide	
1 SAS Enterprise Guide Projects	3
1.1 A brief history of SAS	3
1.2 Opening a project	4
1.3 The contents of projects	5
1.4 Navigating tabs in the Process Flow screen	9
1.5 The main SAS Enterprise Guide menu	10
1.6 Additional resources	12
2 Placing Data into SAS Enterprise Guide Projects	13
2.1 Overview	13
2.2 Entering data directly into SAS Enterprise Guide	13
2.3 Saving a project	19
2.4 Importing data from Excel	19
II Performing Analyses and Viewing Output	
3 Performing Statistical Analyses in SAS Enterprise Guide	25
3.1 Overview	25
	v

vi Contents

3.2	Numerical example	25
3.3	Selecting the procedure	25
3.4	Assigning Task roles	26
3.5	The Variables to assign and Task roles panels	28
3.6	Other choices in the navigation panel	28
3.7	Performing the analysis	31
4	Managing and Viewing Output	32
4.1	Overview	32
4.2	Numerical example	32
4.3	Specifying the output format	32
4.4	Examining the statistical results	35
4.5	Saving the output as a PDF document	38
III	Manipulating Data	
5	Sorting Data and Selecting Cases	43
5.1	Overview	43
5.2	Numerical example	43
5.3	Sorting data	43
5.3	Selecting cases	46
6	Recoding Existing Variables	53
6.1	Overview	53
6.2	Numerical example	54
6.3	Performing the recoding	54
7	Computing New Variables	63
7.1	Overview	63
7.2	Numerical example	63
7.3	Computing a new variable from an existent variable	64
7.4	Computing a new variable by combining several variables	69

IV Describing Data

8 Descriptive Statistics	77
8.1 Overview	77
8.2 Categories of descriptive statistics	77
8.3 Numerical example	79
8.4 Obtaining basic descriptive statistics for the quantitative variables	79
8.5 Obtaining skewness and kurtosis statistics	84
8.6 Obtaining frequency counts for the categorical variables	88
9 Graphing Data	91
9.1 Overview	91
9.2 Numerical example	91
9.3 Constructing bar charts	92
9.4 Constructing line plots	97
10 Standardizing Variables Based on the Sample Data	104
10.1 Overview	104
10.2 Numerical example	105
10.3 Obtaining standardized scores: z scores	105
10.3 Obtaining standardized scores: linear <i>T</i> scores	108
11 Standardizing Variables Based on Existing Norms	111
11.1 Overview	111
11.2 Numerical example	111
11.3 Setting up the computing process	112
11.4 Obtaining the standardized values	115
V Score Distribution Assumptions	
12 Detecting Outliers	119
12.1 Overview	119

viii Contents

12.2	Specifying the boundary for an outlier	119
12.3	Numerical example	120
12.4	The box and whisker plot	121
12.5	Transforming values to z scores	123
12.6	Obtaining extreme values	124
13	Assessing Normality	130
13.1	Overview	130
13.2	The normality tests provided by SAS	130
13.3	Numerical example	131
13.4	Obtaining the normality assessments	131
14	Nonlinearly Transforming Variables in Order to Meet Underlying Assumptions	135
14.1	Overview	135
14.2	Notes on transformations	135
14.3	Examples of nonlinear transformations	136
14.4	Numerical example	137
14.5	Transformation strategy	138
14.6	Switch to Update mode	139
14.7	Setting up the computing process	140
14.8	Evaluating the effects of our transformations	148
VI	Correlation and Prediction	
15	Bivariate Correlation: Pearson Product–Moment and Spearman Rho Correlations	155
15.1	Overview	155
15.2	Some history	155
15.3	The two correlation coefficients of interest here	156
15.4	Numerical example	157
15.5	Setting up the correlation analysis	157
15.6	The correlation output	158
16	Simple Linear Regression	162
16.1	Overview	162

Contents	ix
16.2 Naming the classes of variables	162
16.3 Numerical example	163
16.4 Setting up the regression solution	164
16.5 The regression output	166
17 Multiple Linear Regression	170
17.1 Overview	170
17.2 Numerical example	170
17.3 Viewing the correlations	171
17.4 Setting up the regression solution	172
17.5 The regression output	175
18 Simple Logistic Regression	177
18.1 Overview	177
18.2 Some differences between linear and logistic regression	177
18.3 Two notable features of logistic regression	178
18.4 Numerical example	178
18.5 Setting up the logistic regression solution	179
18.6 The logistic regression output	181
19 Multiple Logistic Regression	185
19.1 Overview	185
19.2 Coding of binary predictor variables	185
19.3 Numerical example	186
19.4 Setting up the logistic regression solution	186
19.5 The logistic regression output	188
VII Comparing Means: The t Test	
20 Independent-Groups t Test	195
20.1 Overview	195
20.2 Some history	195
20.3 Numerical example	196

x Contents

20.4	Setting up the analysis	197
20.5	The t-test output	197
20.6	Magnitude of the effect	199
21	Correlated-Samples t Test	201
21.1	Overview	201
21.2	Relation to bivariate correlation	201
21.3	Numerical example	202
21.4	Setting up the analysis	203
21.5	The t-test output	203
21.6	Magnitude of the effect	205
22	Single-Sample t Test	206
22.1	Overview	206
22.2	The general approach	206
22.3	Numerical example	207
22.4	Setting up the analysis	207
22.5	The t-test output	207
VIII Comparing Means: ANOVA		
23	One-Way Between-Subjects ANOVA	213
23.1	Overview	213
23.2	Naming of ANOVA designs	213
23.3	Some history	214
23.4	Numerical example	215
23.5	Setting up the analysis	216
23.6	The ANOVA output	219
24	Two-Way Between-Subjects Design	223
24.1	Overview	223
24.2	Omnibus and simple effects analysis	224
24.3	Numerical example	224
24.4	Setting up the analysis	225
24.5	The ANOVA output	233

Contents	xi
<hr/>	
25 One-Way Within-Subjects ANOVA	238
25.1 Overview	238
25.2 Numerical example	238
25.3 The structure of the data set	239
25.4 Setting up the analysis	240
25.5 Output for the analysis	250
26 Two-Way Mixed ANOVA Design	253
26.1 Overview	253
26.2 The partitioning of the variance in a mixed design	253
26.3 Numerical example	254
26.4 Setting up the analysis	254
26.5 The ANOVA output	263
 IX Nonparametric Procedures	
27 One-Way Chi-Square	269
27.1 Overview	269
27.2 Numerical example	270
27.3 Setting up the analysis	271
27.4 The chi-square output	272
27.5 Comparing the two most preferred categories: analysis setup	274
27.6 Comparing the two most preferred categories: chi-square output	277
28 Two-Way Chi-Square	279
28.1 Overview	279
28.2 The issue of small frequency counts	280
28.3 Numerical example	282
28.4 Setting up the analysis	282
28.5 The chi-square output	284
29 Nonparametric Between-Subjects One-Way ANOVA	291
29.1 Overview	291

xii Contents

29.2	The nonparametric analogues to One-Way ANOVA	291
29.3	Numerical example	292
29.4	Setting up the analysis	292
29.5	Output of the analyses	293
X	Advanced ANOVA Techniques	
30	One-Way Between-Subjects Analysis of Covariance	299
30.1	Overview	299
30.2	Assumptions of ANCOVA	300
30.3	Numerical example	300
30.4	Evaluating the assumptions of ANCOVA	301
30.5	Setting up the ANCOVA	308
30.6	The ANCOVA output	310
31	One-Way Between-Subjects Multivariate Analysis of Variance	313
31.1	Overview	313
31.2	Univariate and multivariate ANOVA	313
31.3	Numerical example	314
31.4	Setting up the MANOVA	315
31.5	The MANOVA output	316
31.6	Follow-up analyses: setup	319
31.7	Follow-up analyses: output	322
XI	Analysis of Structure	
32	Factor Analysis	327
32.1	Overview	327
32.2	Some history	327
32.3	The basis of factor analysis	328
32.4	The extraction phase	328
32.5	The rotation phase	330
32.6	Numerical example	331

Contents	xiii
<hr/>	
32.7 Setting up the factor analysis	333
32.8 The factor analysis output	337
33 Canonical Correlation Analysis	345
33.1 Overview	345
33.2 Canonical and linear regression	346
33.3 Number of canonical functions	346
33.4 Canonical and factor analysis	347
33.5 Numerical example	347
33.6 Setting up the Canonical Correlation Analysis	349
33.7 Output for Canonical Correlation Analysis	352
References	365
Author Index	371
Subject Index	373

Preface

The present book, *Data Analysis Using SAS Enterprise Guide*, provides readers with an overview of *Enterprise Guide*, the newest point-and-click interface from SAS. *SAS Enterprise Guide* is a graphical user (point-and-click) interface to the main SAS application, having relatively recently replaced the Analyst interface, which itself had replaced the original Assist interface. *Enterprise Guide* makes it easier than ever to access many SAS statistical analyses without learning to write the SAS code underlying its procedures.

We have written this book for readers who have little or no knowledge of *SAS Enterprise Guide* but who may wish to employ it for statistical analysis. Some of these readers will be students in an introductory statistics or data-analysis course; other readers will have taken an introductory statistics course and possibly a research methods course at some time in their past; and still other readers may have had several statistics and research design courses as a part of their background. We have therefore included in this book a relatively wide range of statistical procedures to meet the needs of various readers. There are chapters devoted to the more basic procedures such as descriptive statistics, correlation and simple linear regression, t tests, and one-way chi-square analysis. In addition, we have also included statistical procedures at a somewhat higher level; these include data transformations and other types of computations, multiple linear regression, logistic regression, and some analysis of variance designs. Finally, we have incorporated topics that are more advanced for those readers who might have the need to use such techniques as analysis of covariance, multivariate analysis of variance, factor analysis, and canonical correlation analysis.

Given the wide range and level of topics that we cover, it may not be surprising that the present book is intended to be neither a stand-alone statistics text nor a SAS “cookbook.” Rather, our intent is to instruct readers on how to use SAS

xvi Preface

Enterprise Guide to perform the statistical data analyses covered in the book as well as to understand the concepts underlying those procedures. That is, it is our belief that an exclusive and isolated “select this, then select that” robotic or cookbook synopsis of the steps involved in a given statistical analysis does not serve the needs of most readers. For this reason, we supply for each chapter some analytic and methodological context for the particular statistical procedure that we are describing, enabling readers to gain a sense of the research and statistical framework within which the particular procedure can be used. We also provide interpretations of the statistical results rather than just discussing how to read the output tables that were obtained from *SAS Enterprise Guide*.

There are 33 chapters in this book. They are organized into the following 11 sections.

Section I, “Introducing *SAS Enterprise Guide*,” consists of two chapters presenting the basics of *SAS Enterprise Guide*. The software is designed to work on “projects.” Chapter 1 describes what projects are and focuses on creating projects and navigating within them. Chapter 2 describes how to import data into projects, how to enter data directly into projects, and how to save projects.

Section II, “Performing Analyses and Viewing Output,” consists of two chapters describing how to use *SAS Enterprise Guide*. Chapter 3 informs readers about how to select the statistical procedure they intend to use and how to interact with the dialog screens presented by *SAS Enterprise Guide* in the process of structuring the analysis. Chapter 4 addresses the management and viewing of output.

Section III, “Manipulating Data,” contains three chapters focusing on some ways to organize existing data and generate new variables. Chapter 5 deals with sorting data and selecting a subset of the cases in the data set. Chapter 6 discusses how to recode variables into new or existing variables. Chapter 7 shows how to compute new variables.

Section IV, “Describing Data,” consists of four chapters focused on descriptive statistical and graphical summary procedures. Chapter 8 focuses on computing measures of central tendency and variability. Chapter 9 shows how to graph data in different ways. Chapters 10 and 11 demonstrate how to generate standardized scores based on the sample mean and standard deviation (Chapter 10) and based on existing norms (Chapter 11).

Section V, “Score Distribution Assumptions,” contains three chapters concerning some of the assumptions underlying most of the statistical procedures covered in this book. Chapter 12 explains what statistical outliers are and how to detect them. Chapter 13 focuses on the assessment of normality. Chapter 14 demonstrates how to perform data transformations in order to drive skewed distributions toward normality.

Section VI, “Correlation and Prediction,” contains five chapters dealing with correlation as well as linear and nonlinear regression. Chapter 15 demonstrates how to perform a bivariate correlation analysis by using the Pearson product-moment correlation (r) and Spearman rho. Chapters 16 and 17 cover simple and multiple linear (ordinary least squares) regression, respectively. Chapters 18 and 19 describe the procedures involved in performing simple and multiple logistic regression, respectively.

Section VII, “Comparing Means: The t Test,” contains three chapters encompassing different types of t tests. Chapters 20, 21, and 22 demonstrate how to conduct independent-groups t tests, correlated-samples t tests, and single-sample t tests, respectively.

Section VIII, “Comparing Means: ANOVA,” contains four chapters. Chapters 23, 24, 25, and 26 describe the steps involved in computing analysis of variance (ANOVA) designs for a one-way between-subjects design, a two-way between-subjects design, a one-way within-subjects design, and a two-way mixed design ANOVA, respectively.

Section IX, “Nonparametric Procedures,” consists of three chapters presenting some ways of analyzing frequency and rank-ordered data. Chapters 27 and 28 cover one-way and two-way contingency (chi-square) tables, respectively. Chapter 29 examines nonparametric one-way comparisons of means based on ranked data.

Section X, “Advanced ANOVA Techniques,” is the first section focusing on advanced topics. It contains two chapters extending our treatment of ANOVA to more complex designs. Chapter 30 describes how to perform an analysis of covariance (ANCOVA). Chapter 31 demonstrates how to conduct a one-way multivariate analysis of variance (MANOVA).

Section XI, “Analysis of Structure,” completes our book with two additional chapters on advanced topics, this time covering structural analysis. Chapter 32 describes how to perform and interpret an exploratory factor analysis. Chapter 33 focuses on canonical correlation analysis.

With the exception of those chapters in the first section in which we introduce the software and its interface, the chapters are generally structured in the following manner. We begin with an overview of the topic.

We then present some historical information on the statistical procedure where it is appropriate. We follow this by a numerical example – a data set that we subject to the statistical procedure that is the topic of the chapter. Most of the examples are based on data sets that we have created for this book, but a few draw on real data sets that we or our students have collected in the past; we make clear which is which when we present the data. We also very briefly describe the research design elements involved in the data collection to provide the context for the data sets. For

xviii Preface

each numerical example, we also include a description of the *SAS Enterprise Guide* data set structure and a screen shot showing at least a portion of the data set.

We follow the numerical example by presenting step-by-step guidelines for setting up the analysis in *SAS Enterprise Guide*. Our presentation includes a narration of what has to be done and why it has to be done. This is accompanied by screen shots of the various dialog windows. Finally, we offer step-by-step guidelines for reading and interpreting the output (the printed results) of the analysis. These, too, are accompanied by screen shots of the output.

Acknowledgments

We wish to acknowledge and thank the following individuals for their efforts in maximizing the quality of this book. Lauren Cowles, our editor, has been most helpful and supportive to us during the entire writing process, and her assistant, David Jou, has been very responsive to our inquiries and requests. Peter Katsirubas, our Project Manager at Aptara, kept the production process moving and helped us through that stage of preparing the book. Finally, we are extremely grateful to Susan Zinninger for her marvellous copyediting skills; her time and effort have made the narrative smoother, more readable, and more consistent than what it was when we mailed it to Lauren.