

Jane E. Miller



THE CHICAGO GUIDE TO

Writing about
Multivariate
Analysis

A CHICAGO GUIDE
TO WRITING, EDITING,
AND PUBLISHING
FROM THE UNIVERSITY
OF CHICAGO PRESS

The
Chicago
Guide to
Writing
about
**Multivariate
Analysis**



On Writing, Editing, and Publishing
Jacques Barzun

Tricks of the Trade
Howard S. Becker

Writing for Social Scientists
Howard S. Becker

The Craft of Translation
John Biguenet and Rainer Schulte, editors

The Craft of Research
Wayne C. Booth, Gregory G. Colomb, and Joseph M. Williams

Glossary of Typesetting Terms
Richard Eckersley, Richard Angstadt, Charles M. Ellerston,
Richard Hendel, Naomi B. Pascal, and Anita Walker Scott

Writing Ethnographic Fieldnotes
Robert M. Emerson, Rachel I. Fretz, and Linda L. Shaw

Legal Writing in Plain English
Bryan A. Garner

From Dissertation to Book
William Germano

Getting It Published
William Germano

A Poet's Guide to Poetry
Mary Kinzie

Doing Honest Work in College

Charles Lipson

How to Write a B.A. Thesis

Charles Lipson

The Chicago Guide to Writing about Numbers

Jane E. Miller

Mapping It Out

Mark Monmonier

The Chicago Guide to Communicating Science

Scott L. Montgomery

Indexing Books

Nancy C. Mulvany

Getting into Print

Walter W. Powell

A Manual for Writers of Term Papers, Theses, and Dissertations

Kate L. Turabian

Tales of the Field

John Van Maanen

Style

Joseph M. Williams

A Handbook of Biological Illustration

Frances W. Zweifel

The University of Chicago Press, Chicago 60637
The University of Chicago Press, Ltd., London
© 2005 by The University of Chicago
All rights reserved. Published 2005
Printed in the United States of America

14 13 12 11 10 09 08 07 06 2 3 4 5

ISBN: 0-226-52782-4 (cloth)

ISBN: 0-226-52783-2 (paper)

Library of Congress Cataloging-in-Publication Data

Miller, Jane E. (Jane Elizabeth).

The Chicago guide to writing about multivariate
analysis / Jane E. Miller.

p. cm.

Includes bibliographical references and index.

ISBN 0-226-52782-4 (alk. paper)—ISBN 0-226-52783-2
(pbk. : alk. paper)

1. Technical writing. 2. Multivariate analysis.

I. Title.

T11.M484 2005

808'.066519535—dc22 2004026821

⊗ The paper used in this publication meets the minimum
requirements of the American National Standard for
Information Sciences—Permanence of Paper for Printed
Library Materials, ANSI Z39.48–1992.

JANE E. MILLER
The University of Chicago Press
Chicago and London

The
Chicago
Guide to
Writing
about
**Multivariate
Analysis**

Jane E. Miller is on the faculty at the Institute for Health, Health Care Policy, and Aging Research and the Edward J. Bloustein School of Planning and Public Policy at Rutgers University. Trained as a demographer at the University of Pennsylvania, she has taught research methods and statistics for more than a decade.

The University of Chicago Press, Chicago 60637
The University of Chicago Press, Ltd., London
© 2005 by The University of Chicago
All rights reserved. Published 2005
Printed in the United States of America

14 13 12 11 10 09 08 07 06 05 1 2 3 4 5

ISBN: 0-226-52782-4 (cloth)
ISBN: 0-226-52783-2 (paper)

Library of Congress Cataloging-in-Publication Data
Miller, Jane E. (Jane Elizabeth).

The Chicago guide to writing about multivariate analysis / Jane E. Miller.

p. cm.

Includes bibliographical references and index.

ISBN 0-226-52782-4 (alk. paper)—ISBN 0-226-52783-2 (pbk. : alk. paper)

1. Technical writing. 2. Multivariate analysis.

I. Title.

T11.M484 2005

808'.066519535—dc22 2004026821

⊗ The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48–1992.

*To my sons,
Ian and Colin,
with whom I share
a love of numbers*

CONTENTS

List of Tables xi

List of Figures xiii

List of Boxes xvii

Preface xix

Acknowledgments xxi

1. Introduction 1

PART I. PRINCIPLES

2. Seven Basic Principles 13

3. Causality, Statistical Significance, and Substantive Significance 34

4. Five More Technical Principles 50

PART II. TOOLS

5. Creating Effective Tables 81

6. Creating Effective Charts 120

7. Choosing Effective Examples and Analogies 167

8. Basic Types of Quantitative Comparisons 184

9. Quantitative Comparisons for Multivariate Models 207

10. Choosing How to Present Statistical Test Results 231

PART III. PULLING IT ALL TOGETHER

11. Writing Introductions, Conclusions, and Abstracts 257

12. Writing about Data and Methods 272

13. Writing about Distributions and Associations 301

14. Writing about Multivariate Models 317

15. Speaking about Multivariate Analyses 349

16. Writing for Applied Audiences 380

Appendix A. Implementing “Generalization, Example, Exceptions”
(GEE) 407

Appendix B. Translating Statistical Output into Table and Text 417

x : CONTENTS

Appendix C. Terminology for Ordinary Least Squares (OLS) and Logistic Models	423
Appendix D. Using a Spreadsheet for Calculations	433
Notes	439
Reference List	445
Index	457

LIST OF TABLES

- 4.1. Three tables based on the same cross-tabulation: (a) joint distribution, (b) composition within subgroups, (c) rates of occurrence within subgroups 61
- 4.2. Guidelines on number of digits and decimal places for text, charts, and tables, by type of statistic 74
- 5.1. Anatomy of a table 84
- 5.2. Use of panels to organize conceptually related sets of variables 90
- 5.3. Univariate table: Sample composition 93
- 5.4. Univariate table: Descriptive statistics 94
- 5.5. Comparison of sample with target population 95
- 5.6. Bivariate table: Rates of occurrence based on a cross-tabulation 96
- 5.7. Bivariate table: Pairwise correlations 98
- 5.8. Three-way table with nested column spanners 99
- 5.9a. Table of multivariate model results: Poor version 101
- 5.9b. Table of multivariate model results: Better version 101
- 6.1. Choice of chart type for specific tasks and types of variables 150
- 7.1. Tabular presentation of a sensitivity analysis 176
- 7.2. Calculated effects of substantively meaningful contrasts based on regression coefficients 180
- 8.1. Formulas and case examples for different types of quantitative comparisons 188
- 8.2. Application of absolute difference, relative difference, and percentage change to United States population data 192
- 8.3. Phrases for describing ratios and percentage difference 194
- 8.4. Examples of raw scores, percentage, percentile, and percentage change 200
- 8.5. Relationship between relative risk, prevalence of a risk factor, and attributable risk 202
- 9.1. Standardized and unstandardized coefficients from two ordinary least squares regressions 210
- 9.2. Net effects of an interaction 216

- 9.3. Results of an ordinary least squares model with an interaction between continuous and categorical independent variables 218
- 9.4. Results of a logistic regression 223
- 9.5. Relationship between relative risk, disease prevalence, and odds ratio 226
- 9.6. Relations among relative risk, excess risk, and percentage change in excess risk 228
- 10.1. Alternative ways of presenting statistical significance for an ordinary least squares regression 240
- 10.2. Alternative ways of presenting effect size and statistical significance for a logistic regression 242
- 10.3. Strengths and weaknesses of alternative approaches to presenting statistical significance test results 252
- 13.1. Cross-tabulations and differences in means for a dependent variable 308
- 14.1. Bivariate associations among independent and dependent variables 324
- 14.2. Three-way association among a dependent variable and two independent variables 330
- 14.3. Estimated coefficients from a series of nested models 336
- 14.4. Net effects of main effects and interaction terms 340
- A.1. Generalizing one pattern within a three-way table 408
- B.1. Table of logistic regression results created from computerized output 418
- C.1. Terminology for components of ordinary least squares and logistic regression models 424
- C.2. Names of common multivariate ordinary least squares and categorical dependent variable models 428
- C.3. Terminology for statistical output from ordinary least squares and logistic regression models 430

LIST OF FIGURES

- 2.1. Generalizing patterns from a multiple-line trend chart 31
- 4.1. Effect of a change in scale of reported data on apparent trends 56
- 4.2. Effects of different referent group definitions on percentage calculations 59
- 4.3. Different distributions, each with a mean value of 6.0.
(a) Normal distribution; (b) Normal distribution, same mean, higher SD; (c) Uniform distribution; (d) Polarized bimodal distribution; (e) Skewed distribution. 64
- 6.1. Anatomy of a chart: Multiple-line chart 122
- 6.2. Pie charts to illustrate composition, (a) without data labels, (b) with data (numeric value) labels, (c) with value labels 125
- 6.3. Histogram to illustrate distribution of an ordinal variable 127
- 6.4. Simple bar chart 128
- 6.5. Clustered bar chart: Series of related outcomes by a nominal variable 129
- 6.6. Clustered bar chart to illustrate change or difference in several outcomes by a nominal variable 130
- 6.7. Two versions of clustered bar chart to illustrate an interaction: Patterns by two nominal independent variables, (a) by income group, and (b) by race 131
- 6.8. Clustered bar chart to portray results of competing risks model 132
- 6.9. Two versions of stacked bar charts, illustrating (a) variation in level, (b) percentage distribution 134–35
- 6.10. Single-line chart with a logarithmic scale 136
- 6.11. Single-line chart to illustrate projected pattern for continuous independent variable 137
- 6.12. Multiple-line chart to illustrate an interaction between a continuous and a categorical independent variable 138
- 6.13. Line chart with two y scales 139
- 6.14. High/low/close chart to illustrate median and interquartile range 140

- 6.15. High/low/close chart to illustrate confidence intervals around estimated odds ratios 141
- 6.16. Scatter chart to show association of two continuous variables 143
- 6.17. Map to illustrate numeric pattern by geographic unit 144
- 6.18. Line chart with reference regions 147
- 6.19. Portrait layout of a clustered bar chart 156
- 6.20. Line chart of widely ranging y variable with linear scale 157
- 6.21. Odds ratios with (a) a linear y scale, (b) a logarithmic y scale 159
- 6.22. Line charts of same pattern with (a) truncated y scale, (b) full y scale 161
- 6.23. Chart illustrating inconsistent (figures a and b) and consistent (figures b and c) y scales 162
- 6.24. Inappropriate use of line chart with nominal data 164
- 6.25. Line chart with unequally spaced ordinal variable. (a) Incorrect x scale, (b) Correct x scale 165
- 7.1. Graphical depiction of data heaping 181
- 10.1 F -distributions for selected degrees of freedom 233
- 10.2. Critical values for a chi-square distribution 235
- 12.1. Line chart to show sample size in a longitudinal study 279
- 13.1. Age distribution illustrated with a histogram 305
- 13.2. Interaction: Exception in direction of association 312
- 13.3. Interaction: Exception in magnitude of association 313
- 15.1. Introductory slide: Bulleted text on consequences of issue under study 353
- 15.2. Introductory slide: Chart and text on incidence of issue under study 354
- 15.3. Slide outlining contents of speech 354
- 15.4. Slide with tabular presentation of literature review 355
- 15.5. Slide describing data source using text and a pie chart 355
- 15.6. Slide describing major variables in the analysis 356
- 15.7. Slide with schematic diagram of hypothesized relationship among major variables in the analysis 357
- 15.8. Slide describing model specification for linear and logistic regressions and series of nested models 357
- 15.9. Slide of bivariate association between key independent variable and dependent variable 358

- 15.10. Slide of bivariate associations between key independent variable and potential mediators 359
- 15.11. Text slide summarizing major study conclusions 360
- 15.12. Example of a poor introductory slide 364
- 15.13. Example of a poor data slide 365
- 15.14. Example of a poor results slide using a table directly from the paper 366
- 15.15. Results slide: Chart to illustrate change in estimated coefficients from nested models 367
- 15.16. Results slide: High/low/close chart to present coefficients from categorical independent variables 368
- 15.17. Results slide: Chart to show net effect of nonlinear specification of a continuous independent variable 369
- 16.1. Line chart to display a nonlinear pattern to a nonstatistical audience 386
- 16.2. Example layout for a research poster 392
- 16.3. Example layout for a two-page issue brief 399
- 16.4. Example data and methods summary for a chartbook 400
- 16.5. Bar chart to accompany a written comparison 403
- A.1. Generalizing patterns from a multiple-line trend chart 408
- A.2a. Generalizing one pattern within a three-way chart:
Within clusters 413
- A.2b. Generalizing a second pattern within a three-way chart:
Across clusters 415
- B.1 SAS computer output from a logistic regression model 419
- D.1. Spreadsheet to calculate predicted pattern from a polynomial 434
- D.2. Spreadsheet to calculate net effects of an interaction 436

LIST OF BOXES

- 7.1. Analogy for seasonal adjustment 170
- 8.1. Relations among percentage, percentile, and percentage change 200
- 11.1. Using numbers in an introduction to a scientific paper or report 261
- 11.2. Using numbers in a discussion and conclusion to a scientific paper 266
- 11.3. Structured abstract and keywords for a scientific paper 270
- 12.1. Data section for a scientific paper 288
- 12.2. Methods section for a scientific paper 295
- 12.3. Data and methods in the discussion of a scientific paper 298
- 14.1a. Description of a bivariate table preceding a multivariate analysis: Poor version 326
- 14.1b. Description of a bivariate table preceding a multivariate analysis: Better version 327
- 14.2a. Describing multivariate model results: Poor version 338
- 14.2b. Describing multivariate model results: Better version 339
- 16.1. Excerpt from an issue brief 396
- 16.2. Using numbers in a general-interest article 402
- 16.3. Executive summary 405

PREFACE

The Chicago Guide to Writing about Multivariate Analysis is intended for people who estimate and present multivariate models, focusing on writing about ordinary least squares (OLS), logistic regression, and related methods such as survival analysis or multilevel models. Although measures of association are interpreted differently for probit models, log-linear models, principal components analysis, and other multivariate statistical methods, many of the other principles and tools described here also apply to these types of analyses. In addition to covering basic aspects of writing about numbers, this book shows how to explain why multivariate regression is needed for your research question and data, and how to present the results effectively to both statistical and nonstatistical (“applied”) audiences.

Although I review some basic concepts about OLS and logistic regression, this book is not a substitute for a statistics textbook or a course on regression analysis. To take full advantage of this book, you should have a solid grounding in those methods, ideally including hands-on experience estimating models and interpreting their output. *The Chicago Guide to Writing about Multivariate Analysis* can be used as a companion volume in courses on regression analysis, research methods, or research writing, or as a shelf reference for experienced multivariate analysts who seek to improve their communication about these models and their application. For a study guide with problem sets and suggested course applications, see <http://www.press.uchicago.edu/books/miller/>.

If you write about numbers but do not work with multivariate analyses, see Miller (2004) for guidance on writing about elementary statistics and mathematical concepts.