

Statistics for Anthropology

Second Edition

Anthropology as a discipline is rapidly becoming more quantitative, and anthropology students are now required to develop sophisticated statistical skills. This book provides students of anthropology with a clear, step-by-step guide to univariate statistical methods, demystifying the aspects that are often seen as difficult or impenetrable.

Explaining the central role of statistical methods in anthropology, and using only anthropological examples, the book provides a solid footing in statistical techniques. Beginning with basic descriptive statistics, this new edition also covers more advanced methods such as analyses of frequencies and variance, and simple and multiple regression analysis with dummy and continuous variables. It addresses commonly encountered problems such as small samples and non-normality. Each statistical technique is accompanied by clearly worked examples, and the chapters end with practice problem sets.

Many of the data sets are available for download at www.cambridge.org/9780521147088.

Lorena Madrigal is Professor of Anthropology at the University of South Florida, Tampa. A biological anthropologist, she is particularly interested in the evolution of Afro and Indo Costa Rican populations residing in the Atlantic coast of Costa Rica. She is currently President of the American Association of Physical Anthropologists. She lives in Tampa with her two daughters.

Cambridge University Press
978-0-521-14708-8 - Statistics for Anthropology: Second Edition
Lorena Madrigal
Frontmatter
[More information](#)

Cambridge University Press
978-0-521-14708-8 - Statistics for Anthropology: Second Edition
Lorena Madrigal
Frontmatter
[More information](#)

Statistics for Anthropology

Second Edition

LORENA MADRIGAL

University of South Florida, USA



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press
978-0-521-14708-8 - Statistics for Anthropology: Second Edition
Lorena Madrigal
Frontmatter
[More information](#)

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521147088

© L. Madrigal 1998, 2012

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 1998
Second edition 2012

Printed in the United Kingdom at the University Press, Cambridge

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

Library of Congress Cataloguing in Publication data

Madrigal, Lorena.

Statistics for anthropology / Lorena Madrigal. – Second edition.

pages cm

Includes bibliographical references and index.

ISBN 978-0-521-14708-8

1. Anthropology – Statistical methods. I. Title.

GN34.3.S7M33 2012

301.072'7 – dc23 2011044367

ISBN 978-0-521-14708-8 Paperback

Additional resources for this publication at
www.cambridge.org/9780521147088

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

Cambridge University Press
978-0-521-14708-8 - Statistics for Anthropology: Second Edition
Lorena Madrigal
Frontmatter
[More information](#)

This book is dedicated to all parents who never questioned their children's decision to pursue a career in anthropology. To partners who offered support and comfort. And to children who helped us grow to the limit. May all people be as lucky as I have been.

Cambridge University Press

978-0-521-14708-8 - Statistics for Anthropology: Second Edition

Lorena Madrigal

Frontmatter

[More information](#)

Contents

| | | |
|----------|--|-----------|
| | <i>List of partial statistical tables</i> | page xi |
| | <i>Preface</i> | xiii |
| 1 | Introduction to statistics and simple descriptive statistics | 1 |
| | 1.1 Statistics and scientific enquiry | 1 |
| | 1.2 Basic definitions | 3 |
| | 1.2.1 Variables and constants | 3 |
| | 1.2.2 Scales of measurement | 4 |
| | 1.2.3 Accuracy and precision | 6 |
| | 1.2.4 Independent and dependent variables | 6 |
| | 1.2.5 Control and experimental groups | 7 |
| | 1.2.6 Samples and statistics, populations and parameters. Descriptive and inferential statistics. A few words about sampling | 8 |
| | 1.3 Statistical notation | 9 |
| | 1.4 Chapter 1 key concepts | 12 |
| | 1.5 Chapter 1 exercises | 12 |
| 2 | The first step in data analysis: summarizing and displaying data. Computing descriptive statistics | 13 |
| | 2.1 Frequency distributions | 13 |
| | 2.1.1 Frequency distributions of discontinuous numeric and qualitative variables | 13 |
| | 2.1.2 Frequency distributions of continuous numeric variables | 15 |
| | 2.1.3 Stem-and-leaf displays of data | 17 |
| | 2.2 Graphing data | 18 |
| | 2.2.1 Bar graphs and pie charts | 19 |
| | 2.2.2 Histograms | 21 |
| | 2.2.3 Polygons | 21 |
| | 2.2.4 Box plots | 21 |
| | 2.3 Descriptive statistics. Measures of central tendency and dispersion | 25 |
| | 2.3.1 Measures of central tendency | 26 |
| | 2.3.2 Measures of variation | 29 |
| | 2.4 Chapter 2 key concepts | 39 |

| | | |
|----------|---|------------|
| 2.5 | Computer resources | 40 |
| 2.6 | Chapter 2 exercises | 40 |
| 3 | Probability and statistics | 42 |
| 3.1 | Random sampling and probability distributions | 43 |
| 3.2 | The probability distribution of qualitative and discontinuous numeric variables | 44 |
| 3.3 | The binomial distribution | 46 |
| 3.4 | The Poisson distribution | 48 |
| 3.5 | Bayes' theorem | 53 |
| 3.6 | The probability distribution of continuous variables | 57 |
| 3.6.1 | z scores and the standard normal distribution (SND) | 63 |
| 3.6.2 | Percentile ranks and percentiles | 71 |
| 3.6.3 | The probability distribution of sample means | 73 |
| 3.6.4 | Is my bell shape normal? | 77 |
| 3.7 | Chapter 3 key concepts | 78 |
| 3.8 | Computer resources | 79 |
| 3.9 | Chapter 3 exercises | 80 |
| 4 | Hypothesis testing and estimation | 83 |
| 4.1 | Different approaches to hypothesis testing and estimation | 83 |
| 4.1.1 | The classical significance testing approach | 83 |
| 4.1.2 | The maximum likelihood approach | 84 |
| 4.1.3 | The Bayesian approach | 84 |
| 4.2 | Estimation | 84 |
| 4.2.1 | Confidence limits and confidence interval | 85 |
| 4.2.2 | Point estimation | 89 |
| 4.3 | Hypothesis testing | 90 |
| 4.3.1 | The principles of hypothesis testing | 90 |
| 4.3.2 | Errors and power in hypothesis testing | 93 |
| 4.3.3 | Hypothesis tests using z scores | 98 |
| 4.3.4 | One- and two-tailed hypothesis tests | 100 |
| 4.3.5 | Assumptions of statistical tests | 101 |
| 4.3.6 | Hypothesis testing with the t distribution | 103 |
| 4.3.7 | Hypothesis tests using t scores | 104 |
| 4.3.8 | Reporting hypothesis tests | 105 |
| 4.3.9 | The classical significance testing approach. A conclusion | 106 |
| 4.4 | Chapter 4 key concepts | 106 |
| 4.5 | Chapter 4 exercises | 107 |
| 5 | The difference between two means | 108 |
| 5.1 | The un-paired t test | 108 |
| 5.1.1 | Assumptions of the un-paired t test | 112 |

| | | |
|----------|---|------------|
| | | |
| | 5.2 The comparison of a single observation with the mean of a sample | 116 |
| | 5.3 The paired <i>t</i> test | 117 |
| | 5.3.1 Assumptions of the paired <i>t</i> test | 119 |
| | 5.4 Chapter 5 key concepts | 120 |
| | 5.5 Computer resources | 120 |
| | 5.6 Chapter 5 exercises | 121 |
| 6 | The analysis of variance (ANOVA) | 122 |
| | 6.1 Model I and model II ANOVA | 122 |
| | 6.2 Model I, one-way ANOVA. Introduction and nomenclature | 123 |
| | 6.3 ANOVA assumptions | 131 |
| | 6.4 Post-hoc tests | 132 |
| | 6.4.1 The Scheffé test | 133 |
| | 6.5 Model I, two-way ANOVA | 135 |
| | 6.6 Other ANOVA designs | 143 |
| | 6.7 Chapter 6 key concepts | 144 |
| | 6.8 Computer resources | 145 |
| | 6.9 Chapter 6 exercises | 145 |
| 7 | Non-parametric tests for the comparison of samples | 146 |
| | 7.1 Ranking data | 147 |
| | 7.2 The Mann–Whitney U test for a two-sample un-matched design | 148 |
| | 7.3 The Kruskal–Wallis for a one-way, model I ANOVA design | 153 |
| | 7.4 The Wilcoxon signed-ranks test for a two-sample paired design | 159 |
| | 7.5 Chapter 7 key concepts | 164 |
| | 7.6 Computer resources | 164 |
| | 7.7 Chapter 7 exercises | 164 |
| 8 | The analysis of frequencies | 166 |
| | 8.1 The X^2 test for goodness-of-fit | 166 |
| | 8.2 The Kolmogorov–Smirnov one sample test | 170 |
| | 8.3 The X^2 test for independence of variables | 172 |
| | 8.4 Yates’ correction for continuity | 175 |
| | 8.5 The likelihood ratio test (the G test) | 176 |
| | 8.6 Fisher’s exact test | 178 |
| | 8.7 The McNemar test for a matched design | 183 |
| | 8.8 Tests of goodness-of-fit and independence of variables. Conclusion | 184 |
| | 8.9 The odds ratio (OR): measuring the degree of the association between two discrete variables | 185 |
| | 8.10 The relative risk (RR): measuring the degree of the association between two discrete variables | 188 |
| | 8.11 Chapter 8 key concepts | 190 |
| | 8.12 Computer resources | 190 |
| | 8.13 Chapter 8 exercises | 191 |

| | | |
|-----------|--|-----|
| 9 | Correlation analysis | 193 |
| 9.1 | The Pearson product-moment correlation | 193 |
| 9.2 | Non-parametric tests of correlation | 199 |
| 9.2.1 | The Spearman correlation coefficient r_s | 199 |
| 9.2.2 | Kendall's coefficient of rank correlation – tau (τ) | 202 |
| 9.3 | Chapter 9 key concepts | 208 |
| 9.4 | Chapter 9 exercises | 208 |
| 10 | Simple linear regression | 209 |
| 10.1 | An overview of regression analysis | 210 |
| 10.2 | Regression analysis step-by-step | 214 |
| 10.2.1 | The data are plotted and inspected to detect violations of the linearity and homoscedasticity assumptions | 214 |
| 10.2.2 | The relation between the X and the Y is described mathematically with an equation | 215 |
| 10.2.3 | The regression analysis is expressed as an analysis of the variance of Y | 215 |
| 10.2.4 | The null hypothesis that the parametric value of the slope is not statistically different from 0 is tested | 217 |
| 10.2.5 | The regression equation is used to predict values of Y | 217 |
| 10.2.6 | Lack of fit is assessed | 219 |
| 10.2.7 | The residuals are analyzed | 221 |
| 10.3 | Transformations in regression analysis | 225 |
| 10.4 | Chapter 10 key concepts | 232 |
| 10.5 | Computer resources | 232 |
| 10.6 | Chapter 10 exercises | 232 |
| 11 | Advanced topics in regression analysis | 234 |
| 11.1 | The multiple regression model | 234 |
| 11.1.1 | The problem of multicollinearity/collinearity | 235 |
| 11.1.2 | The algebraic computation of the multiple regression equation | 236 |
| 11.1.3 | An overview of multiple-regression-model building | 240 |
| 11.1.4 | Dummy independent variables | 247 |
| 11.2 | An overview of logistic regression | 251 |
| 11.3 | Writing up your results | 255 |
| 11.4 | Chapter 11 key concepts | 255 |
| 11.5 | Computer resources | 256 |
| 11.6 | Chapter 11 exercises | 256 |
| | <i>References</i> | 257 |
| | <i>Index</i> | 260 |

List of partial statistical tables

| | |
|--|----------------|
| Areas under the normal curve. Table 3.5 | <i>page</i> 65 |
| Selected critical values of the t distribution. Table 4.1 | 87 |
| Selected critical values of the t distribution. Table 5.1 | 111 |
| Selected critical values of the maximum F ratio. Table 5.2 | 115 |
| Selected critical values of the F distribution. Table 6.1 | 128 |
| Selected critical values of U , the Mann–Whitney statistic for a two-tailed test. Table 7.1 | 149 |
| Selected critical values of the chi-square distribution. Table 7.2 | 154 |
| Selected critical values of T for the Wilcoxon signed-ranks test. Table 7.3 | 160 |
| Selected critical values of the chi-square distribution. Table 8.1 | 168 |
| Selected critical values for d_{max} for the Kolmogorov–Smirnov goodness-of-fit test for discrete data. Table 8.2 | 171 |
| Selected critical values of the Pearson correlation coefficient. Table 9.1 | 196 |

Cambridge University Press

978-0-521-14708-8 - Statistics for Anthropology: Second Edition

Lorena Madrigal

Frontmatter

[More information](#)

Preface

The first edition of this book was published 13 years ago, and much has changed since that time. To put it directly, the first edition of *Statistics for Anthropology* became terribly old-fashioned. When I wrote the first edition it made sense to discuss rounding-off rules, and to dedicate an entire chapter to the computation of frequency distributions. In the year 2012 it does not make any sense to devote time to these topics, as most students will be using a computer package which will easily round the numbers and compute the frequency distributions for them. I have also decided to dedicate very little time to the production of graphs because this is something that college students learned in their pre-college education, and which is easily done with many computer programs. At the same time, I augmented the book by including several issues that were missing in the first edition: the Poisson distribution, two-way ANOVA, the odds ratio, Fisher's exact test, Kendall's tau, and an entire chapter on advanced regression topics. In addition, and to reflect current practice, I have expanded my discussion on probability as to facilitate an understanding of maximum likelihood and Bayesian approaches to statistical inference, as well as to classical hypothesis testing. However, most of the book focuses on the latter, which is still the most frequent approach used in anthropological quantitative analysis. There are specialized books on maximum likelihood and Bayesian approaches which should be consulted for more in-depth discussion on maximum likelihood and Bayesian statistics.

Although this edition is going to be more advanced than the first one, it does not intend to give students their entire statistical education! Instead, this textbook will cover important univariate techniques. Students wishing to learn multivariate statistics need to take a second-level course.

There are two other changes in the book which reflect the passing of time. The first is that no statistical table is included at the end of the book. These tables are now part of the public domain, and they are posted in numerous websites, including this book's website. In my opinion, computer packages have made these tables virtually obsolete because computer packages give analysts the information they used to get in the statistical tables. Therefore, I have only reproduced selected values of statistical tables within the book, enough to illustrate their use as needed in the book's practices and exercises. Another obvious change to this edition is that the data sets used for examples throughout the book are also available in Excel format at the book's website. A description of most of the data sets is included in the website as well. If the data are described within the book, no description is found in the website. When doing the problems by hand I tried to keep as

many digits as possible so as to replicate as close as possible the computer output I was getting with SAS or with PASW. However, this does not mean that the answer you will see in the book will be exactly the same (to the very last digit) that you will get with any and all computer packages, many of which use different algorithms. Slight differences in rounding are certainly to be expected.

Most chapters have the same format: (almost) every statistical technique is explained with a mathematical formula and illustrated with an example for which the reader will need a calculator. Afterwards another example will be discussed. Since every instructor is going to choose his or her own computer package, I will not dedicate much time explaining how to work with any particular computer program. Sometimes I will briefly note at the end of the chapter (under computer resources) how to perform the statistical analysis using PASW and SAS, two popular computer packages.

What has not changed since the publication of the first edition is the need for anthropology students to have a solid foundation in statistical analysis. A cursory reading in the major journals in the discipline will show that quantitative methods of data analysis are crucial to anthropological research. When appropriate I will discuss research articles which use the statistical technique I am discussing. I would like to note that this book covers statistical techniques to analyze quantitative data. Therefore, I do not cover any kind of qualitative data analysis.

I wish to thank the instructors who adopted the first edition, and the students who told me that they enjoyed learning statistics with the first edition. I also want to thank everyone who gave me suggestions for improving the book, whether personally or in writing. In this second edition I endeavored to remove all typos and errors which did exist in the first one. However, as a typical member of the *Homo sapiens* species I may have made a few mistakes. Please contact me at madrigal@usf.edu to let me know of any.

Happy computing!