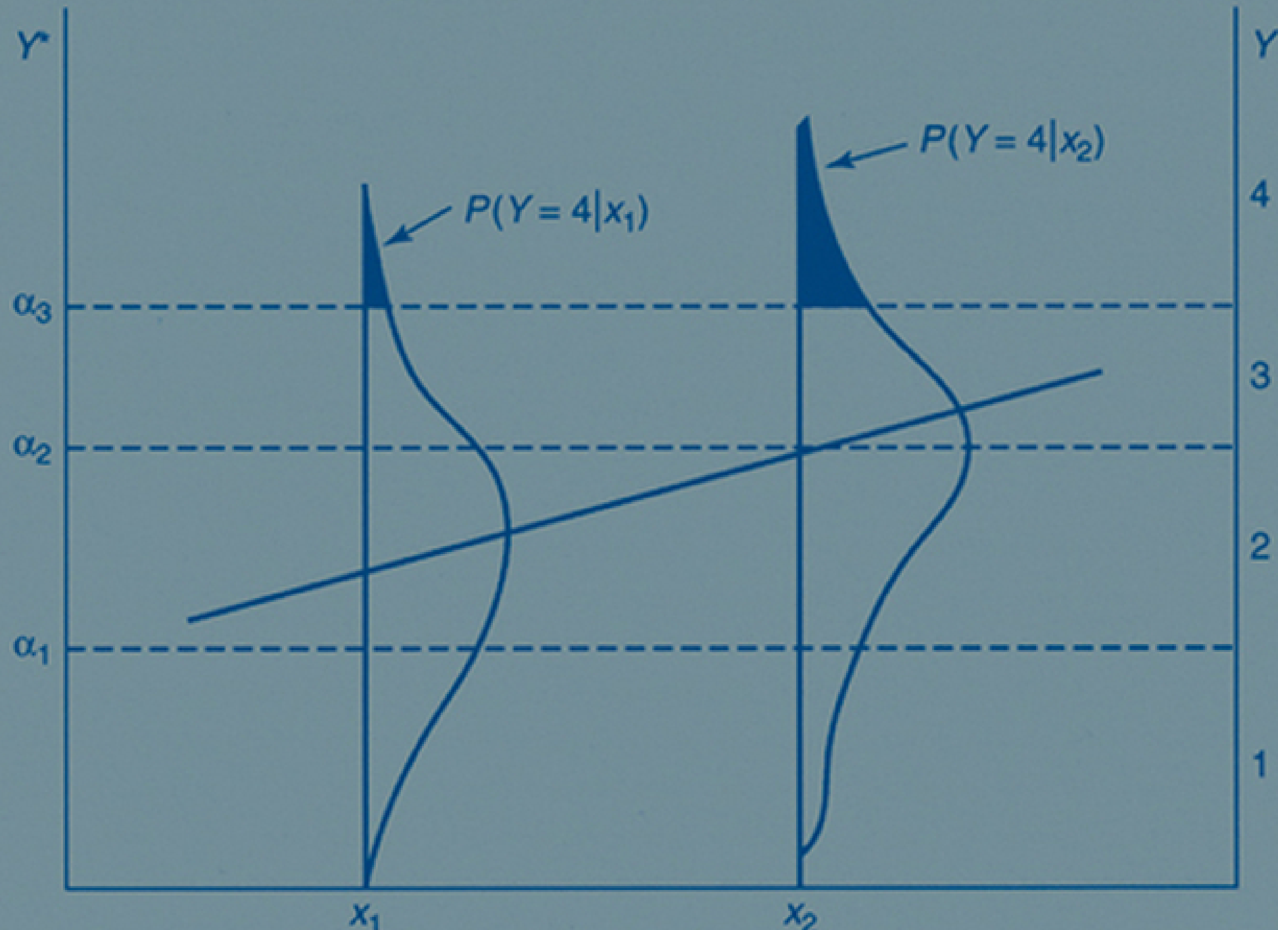
Wiley Series in Probability and Statistics

Analysis of Ordinal Categorical Data

Second Edition

ALAN AGRESTI



WILEY

Analysis of Ordinal
Categorical Data

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

A complete list of the titles in this series appears at the end of this volume.

Analysis of Ordinal Categorical Data

Second Edition

Alan Agresti

University of Florida
Gainesville, Florida

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Agresti, Alan.
Analysis of ordinal categorical data / Alan Agresti.—2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-08289-8 (cloth)
1. Multivariate analysis. I. Title.
QA278.A35 2010
519.5′35—dc22

2009038760

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents

Preface

In recent years methods for analyzing categorical data have matured considerably in their development. There has been a tremendous increase in the publication of research articles on this topic. Several books on categorical data analysis have introduced the methods to audiences of nonstatisticians as well as to statisticians, and the methods are now used frequently by researchers in areas as diverse as sociology, public health, and wildlife ecology. Yet some types of methods are still in the process of development, such as methods for clustered data, Bayesian methods, and methods for sparse data sets with large numbers of variables.

What distinguishes this book from others on categorical data analysis is its emphasis on methods for response variables having ordered categories, that is, *ordinal* variables. Specialized models and descriptive measures are discussed that use the information on ordering efficiently. These ordinal methods make possible simpler description of the data and permit more powerful inferences about population characteristics than do models for *nominal* variables that ignore the ordering information.

This is the second edition of a book published originally in 1984. At that time many statisticians were unfamiliar with the relatively new modeling methods for categorical data analysis, so the early chapters of the first edition introduced generalized linear modeling topics such as logistic regression and loglinear models. Since many books now provide this information, this second edition takes a different approach, assuming that the reader already has some familiarity with the basic methods of categorical data analysis. These methods include descriptive summaries using odds ratios, inferential methods including chi-squared tests of the hypotheses of independence and conditional independence, and logistic regression modeling, such as presented in Chapters 1 to 6 of my books *An Introduction to Categorical Data Analysis* (2nd ed., Wiley, 2007) and *Categorical Data Analysis* (2nd ed., Wiley, 2002).

On an ordinal scale, the technical level of this book is intended to fall between that of the two books just mentioned. I intend the book to be accessible to a broad audience, particularly professional statisticians and methodologists in areas such as public health, the pharmaceutical industry, the social and behavioral sciences, and business and government. Although there is some discussion of the underlying theory, the main emphasis is on presenting various ordinal methodologies. Thus, the book has more discussion of interpretation and application of the methods than of the technical details. However, I also intend the book to be useful to specialists who may want to become aware of recent research advances, to supplement the background provided. For this purpose, the *Notes* section at the end of each chapter provides supplementary technical comments and embellishments, with emphasis on references to related research literature.

The text contains significant changes from and additions to the first edition, so it seemed as if I were writing a new book! As mentioned, the basic introductions to logistic regression and loglinear models have been removed. New material includes chapters on marginal models and random effects models for clustered data (Chapters 9 and 10) and Bayesian methods (Chapter 11), coverage of additional models such as the stereotype model, global odds ratio models, and generalizations of cumulative logit models, coverage of order-restricted inference, and more detail throughout on established methods.

Nearly all the methods presented can be implemented using standard statistical software packages such as R and S-Plus, SAS, SPSS, and Stata. The use of software for ordinal methods is discussed in the Appendix. The web site www.stat.ufl.edu/∼aa/cda/software.html gives further details about software for applying methods of categorical data analysis. The web site www.stat.ufl.edu/∼aa/ordinal/ord.html displays data sets not shown fully in the text (in the form of SAS programs), several examples of the use of a R function (mph.fit) that can conduct many of the nonstandard analyses in the text, and a list of known errata in the text.

The first edition was prepared mainly while I was visiting Imperial College, London, on sabbatical leave in 1981–1982. I would like to thank all who commented on the manuscript for that edition, especially Sir David Cox and Bent Jørgensen.

For this edition, special thanks to Maria Kateri and Joseph Lang for reading a complete draft and making helpful suggestions and critical comments. Maria Kateri also very generously provided bibliographic checking and pointed out many relevant articles that I did not know about. Thanks to Euijung Ryu for computing help with a few examples, for help with improving a graphic and with my LaTeX code, and for many helpful suggestions on the text and the Bibliography. Bhramar Mukherjee very helpfully discussed Bayesian methods for ordinal data and case–control methods and provided many suggestions about Chapter 11. Also, Ivy Liu and Bernhard Klingenberg made helpful suggestions based on an early draft, Arne Bathke suggested relevant research on rank-based methods, Edgar Brunner provided several helpful comments about rank-based methods and elegant ways of constructing statistics, and Carla Rampichini suggested relevant research on ordinal multilevel models. Thanks to Stu Lipsitz for data for Example 9.2.3 and to John Williamson and Kyungmann Kim for data for Example 9.1.3. Thanks to Beka Steorts for WinBUGS help, Cyrus Mehta for the use of StatXact, Jill Rietema for arranging for the use of SPSS, and Oliver Schabenberger for arranging for the use of SAS. I would like to thank co-authors of mine on various articles for permission to use materials from those articles. Finally, thanks as always to my wife, Jacki Levine, for her unwavering support during the writing of this book.

A truly wonderful reward of my career as a university professor has been the opportunity to work on research projects with Ph.D. students in statistics and with statisticians around the world. It is to them that I would like to dedicate this book.

Alan Agresti

Gainesville, Florida and Brookline, Massachusetts
January 2010

CHAPTER 1

Introduction

1.1 ORDINAL CATEGORICAL SCALES

Until the early 1960s, statistical methods for the analysis of categorical data were at a relatively primitive stage of development. Since then, methods have been developed more fully, and the field of categorical data analysis is now quite mature. Since about 1980 there has been increasing emphasis on having data analyses distinguish between ordered and unordered scales for the categories. A variable with an ordered categorical scale is called *ordinal*. In this book we summarize the primary methods that can be used, and usually should be used, when response variables are ordinal.

Examples of ordinal variables and their ordered categorical scales (in parentheses) are opinion about government spending on the environment (too high, about right, too low), educational attainment (grammar school, high school, college, postgraduate), diagnostic rating based on a mammogram to detect breast cancer (definitely normal, probably normal, equivocal, probably abnormal, definitely abnormal), and quality of life in terms of the frequency of going out to have fun (never, rarely, occasionally, often). A variable with an unordered categorical scale is called *nominal*. Examples of nominal variables are religious affiliation (Protestant, Catholic, Jewish, Muslim, other), marital status (married, divorced, widowed, never married), favorite type of music (classical, folk, jazz, rock, other), and preferred place to shop (downtown, Internet, suburban mall). Distinct levels of such variables differ in quality, not in quantity. Therefore, the listing order of the categories of a nominal variable should not affect the statistical analysis.

Ordinal scales are pervasive in the social sciences for measuring attitudes and opinions. For example, each subject could be asked to respond to a statement such as "Same-sex marriage should be legal" using categories such as (strongly disagree, disagree, undecided, agree, strongly agree) or (oppose strongly, oppose

Analysis of Ordinal Categorical Data, Second Edition, By Alan Agresti

mildly, neutral, favor mildly, favor strongly). Such a scale with a neutral middle category is often called a *Likert scale*. Ordinal scales also occur commonly in medical and public health disciplines: for example, for variables describing pain (none, mild, discomforting, distressing, intense, excruciating), severity of an injury in an automobile crash (uninjured, mild injury, moderate injury, severe injury, death), illness after a period of treatment (much worse, a bit worse, the same, a bit better, much better), stages of a disease (I, II, III), and degree of exposure to a harmful substance, such as measuring cigarette smoking with the categories (nonsmoker, <1 pack a day, ≥ 1 pack a day) or measuring alcohol consumption of college students with the scale (abstainer, non-binge drinker, occasional binge drinker, frequent binge drinker). In all fields, ordinal scales result when inherently continuous variables are measured or summarized by researchers by collapsing the possible values into a set of categories. Examples are age measured in years (0–20, 21–40, 41–60, 61–80, above 80), body mass index (BMI) measured as (<18.5, 18.5–24.9, 25–29.9, ≥ 30) for (underweight, normal weight, overweight, obese), and systolic blood pressure measured as (<120, 120–139, 140–159, ≥ 160) for (normal, prehypertension, stage 1 hypertension, stage 2 hypertension).

Often, for each observation the choice of a category is subjective, such as in a subject's report of pain or in a physician's evaluation regarding a patient's stage of a disease. (An early example of such subjectivity was U.S. President Thomas Jefferson's suggestion during his second term that newspaper articles could be classified as truths, probabilities, possibilities, or lies.) To lessen the subjectivity, it is helpful to provide guidance about what the categories represent. For example, the College Alcohol Study conducted at the Harvard School of Public Health defines "binge drinking" to mean at least five drinks for a man or four drinks for a woman within a two-hour period (corresponding to a blood alcohol concentration of about 0.08%); "occasional binge drinking" is defined as binge drinking once or twice in the past two weeks; and "frequent binge drinking" is binge drinking at least three times in the past two weeks.

For ordinal scales, unlike *interval* scales, there is a clear ordering of the levels, but the absolute distances among them are unknown. Pain measured with categories (none, mild, discomforting, distressing, intense, excruciating) is ordinal, because a person who chooses "mild" feels more pain than if he or she chose "none," but no numerical measure is given of the difference between those levels. An ordinal variable is *quantitative*, however, in the sense that each level on its scale refers to a greater or smaller magnitude of a certain characteristic than another level. Such variables are of quite a different nature than qualitative variables, which are measured on a nominal scale and have categories that do not relate to different magnitudes of a characteristic.

1.2 ADVANTAGES OF USING ORDINAL METHODS

Many well-known statistical methods for categorical data treat all response variables as nominal. That is, the results are invariant to permutations of the categories of those variables, so they do not utilize the ordering if there is one. Examples are the Pearson chi-squared test of independence and multinomial response modeling using baseline-category logits. Test statistics and P-values take the same values regardless of the order in which categories are listed. Some researchers routinely apply such methods to nominal and ordinal variables alike because they are both categorical.

Recognizing the discrete nature of categorical data is useful for formulating sampling models, such as in assuming that the response variable has a multinomial distribution rather than a normal distribution. However, the distinction regarding whether data are continuous or discrete is often less crucial to substantive conclusions than whether the data are qualitative (nominal) or quantitative (ordinal or interval). Since ordinal variables are inherently quantitative, many of their descriptive measures are more like those for interval variables than those for nominal variables. The models and measures of association for ordinal data presented in this book bear many resemblances to those for continuous variables.

A major theme of this book is how to analyze ordinal data by utilizing their quantitative nature. Several examples show that the type of ordinal method used is not that crucial, in the sense that we obtain similar substantive results with ordinal logistic regression models, loglinear models, models with other types of response functions, or measures of association and nonparametric procedures. These results may be quite different, however, from those obtained using methods that treat all the variables as nominal.

Many advantages can be gained from treating an ordered categorical variable as ordinal rather than nominal. They include:

- Ordinal data description can use measures that are similar to those used in ordinary regression and analysis of variance for quantitative variables, such as correlations, slopes, and means.
- Ordinal analyses can use a greater variety of models, and those models are more parsimonious and have simpler interpretations than the standard models for nominal variables, such as baseline-category logit models.
- Ordinal methods have greater power for detecting relevant trend or location alternatives to the null hypothesis of "no effect" of an explanatory variable on the response variable.
- Interesting ordinal models apply in settings for which standard nominal models are trivial or else have too many parameters to be tested for goodness of fit.

An ordinal analysis can give quite different and much more powerful results than an analysis that ignores the ordinality. For a preview of this, consider Table 1.1, with artificial counts in a contingency table designed to show somewhat of a trend from the top left corner to the bottom right corner. For two-way contingency tables, the first analysis many methodologists apply is the chi-squared test of independence. The Pearson statistic equals 10.6 with df $= 9$, yielding an unimpressive P-value of 0.30. By contrast, various possible ordinal analyses for testing this hypothesis have

TABLE 1.1. Data Set for Which Ordinal Analyses Give Very Different Results from Unordered Categorical Analyses

	Column 1	Column 2	Column 3	Column 4
Row 1	8	6	4	2
Row 2	6	8	6	4
Row 3	4	6	8	6
Row 4	2	4	6	8

chi-squared statistics on the order of 9 or 10, but with df $= 1$, and have P-values on the order of 0.002 and 0.001.

1.3 ORDINAL MODELING VERSUS ORDINARY REGESSION ANALYSIS

There are two relatively extreme ways to analyze ordered categorical response variables. One way, still common in practice, ignores the categorical nature of the response variable and uses standard parametric methods for continuous response variables. This approach assigns numerical scores to the ordered categories and then uses ordinary least squares (OLS) methods such as linear regression and analysis of variance (ANOVA). The second way restricts analyses solely to methods that use only the ordering information about the categories. Examples of this approach are nonparametric methods based on ranks and models for cumulative response probabilities.

1.3.1 Latent Variable Models for Ordinal Data

Many other methods fall between the two extremes described above, using ordinal information but having some parametric structure as well. For example, often it is natural to assume that an unobserved continuous variable underlies the ordinal response variable. Such a variable is called a *latent variable*.

In a study of political ideology, for example, one survey might use the categories liberal, moderate, and conservative, whereas another might use very liberal, slightly liberal, moderate, slightly conservative, and very conservative or an even finer categorization. We could regard such scales as categorizations of an inherently continuous scale that we are unable to observe. Then, rather than assigning scores to the categories and using ordinary regression, it is often more sensible to base description and inference on parametric models for the latent variable. In fact, we present connections between this approach and a popular modeling approach that has strict ordinal treatment of the response variable: In Chapters 3 and 5 we show that a logistic model and a probit model for cumulative probabilities of an ordinal response variable can be motivated by a latent variable model for an underlying quantitative response variable that has a parametric distribution such as the normal.

1.3.2 Using OLS Regression with an Ordinal Response Variable

In this book we do present methods that use only the ordering information. It is often attractive to begin a statistical analysis by making as few assumptions as possible, and a strictly ordinal approach does this. However, in this book we also present methods that have some parametric structure or that require assigning scores to categories. We believe that strict adherence to operations that utilize only the ordering in ordinal scales limits the scope of useful methodology too severely. For example, to utilize the ordering of categories of an ordinal explanatory variable, nearly all models assign scores to the categories and regard the variable as quantitative—the alternative being to ignore the ordering and treat the variable as nominal, with indicator variables. Therefore, we do not take a rigid view about permissible methodology for ordinal variables.

That being said, we recommend against the simplistic approach of posing linear regression models for ordinal response scores and fitting them using OLS methods. Although that approach can be useful for identifying variables that clearly affect a response variable, and for simple descriptions, limitations occur. First, there is usually not a clear-cut choice for the scores. Second, a particular response outcome is likely to be consistent with a range of values for some underlying latent variable, and an ordinary regression analysis does not allow for the measurement error that results from replacing such a range by a single numerical value. Third, unlike the methods presented in this book, that approach does not yield estimated probabilities for the response categories at fixed settings of the explanatory variables. Fourth, that approach can yield predicted values above the highest category score or below the lowest. Fifth, that approach ignores the fact that the variability of the responses is naturally nonconstant for categorical data: For an ordinal response variable, there is little variability at predictor values for which observations fall mainly in the highest category (or mainly in the lowest category), but there is considerable variability at predictor values for which observations tend to be spread among the categories.

Related to the second, fourth, and fifth limitations, the ordinary regression approach does not account for "ceiling effects" and "floor effects," which occur because of the upper and lower limits for the ordinal response variable. Such effects can cause ordinary regression modeling to give misleading results. These effects also result in substantial correlation between values of residuals and values of quantitative explanatory variables.

1.3.3 Example: Floor Effect Causes Misleading OLS Regression

How can ordinary regression give misleading results when used with ordered categorical response variables? To illustrate, we apply the standard linear regression model to simulated data with an ordered categorical response variable y based on an underlying continuous latent variable y^*. The explanatory variables are a continuous variable x and a binary variable z. The data set of 100 observations was generated as follows: The x values were independently uniformly generated between 0 and 100, and the z values were independently generated with $P(z = 0) = P(z = 1) = 0.50$. At a given x, the latent response outcome y^* was generated according to a normal distribution with mean

$$E(y^*) = 20.0 + 0.6x - 40.0z$$

and standard deviation 10. The first scatterplot in Figure 1.1 shows the 100 observations on y^* and x, each data point labeled by the category for z. The plot also shows the OLS fit that estimates this model.

We then categorized the 100 generated values on y^* into five categories to create observations for an ordinal variable y, as follows:

$$y = 1 \text{ if } y^* \leq 20, \quad y = 2 \text{ if } 20 < y^* \leq 40, \quad y = 3 \text{ if } 40 < y^* \leq 60,$$
$$y = 4 \text{ if } 60 < y^* \leq 80, \quad y = 5 \text{ if } y^* > 80.$$

The second scatterplot in Figure 1.1 shows 100 observations on y and x. At low x levels, there is a floor effect for the observations with $z = 1$. When $x < 50$ with $z = 1$, there is a very high probability that observations fall in the lowest category of y.

Using OLS with scores 1, 2, 3, 4, and 5 for the categories of y suggests either (a) a model with an interaction term, allowing different slopes relating $E(y)$ to x when

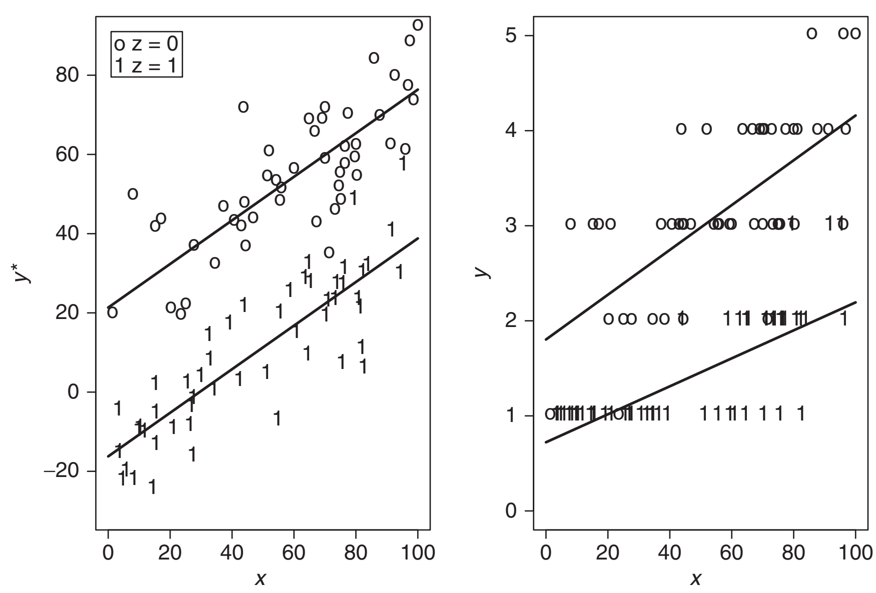


Figure 1.1. Ordered categorical data (in second panel) for which ordinary regression suggests interaction, because of a floor effect, but ordinal modeling does not. The data were generated (in first panel) from a normal main-effects regression model with continuous (x) and binary (z) explanatory variables. When the continuous response y^* is categorized and y is measured as (1, 2, 3, 4, 5), the observations labeled "1" for the category of z have a linear x effect with only half the slope of the observations labeled "0" for the category of z.

$z = 0$ and when $z = 1$, or (b) a model with a quadratic effect of x on $E(y)$ when $z = 1$. The second scatterplot in Figure 1.1 shows the fit of the linear interaction model, that is, using OLS to fit the model $E(y) = \alpha + \beta_1 x + \beta_2 z + \beta_3(x \times z)$ to the ordered categorical response. The slope of the line is about twice as high when $z = 0$ as when $z = 1$. This interaction effect is caused by the observations when $z = 1$ tending to fall in category $y = 1$ whenever x takes a relatively low value. As x gets lower, the underlying value y^* can continue to tend to get lower, but the observed ordinal response cannot fall below 1.

Standard ordinal models such as those introduced in Chapters 3 to 5 fit the data well without the need for an interaction term. Such models can be motivated by a latent variable model. They allow for underlying values of y^* when $z = 1$ to be below those when $z = 0$, even if x is so low that y is very likely to be in the first category at both levels of z. (The data in Figure 1.1 are revisited with such a model in Exercise 5.2.)

Hastie et al. (1989) showed a real-data example of the type we presented here with simulated data. They described a study of women in South Africa that modeled an ordinal measurement y of osteoporosis in terms of x = age and an indicator variable z for whether the woman had osteoarthritis. At low age levels, a high proportion of women clustered in the lowest category of osteoporosis, regardless of osteoarthritis status. Using OLS, for each osteoarthritis group the line relating age to the predicted osteoporosis score took value at the lowest ordinal level near a relatively low age level, but the line for the group positive for osteoarthritis had a significantly greater slope as age increased. In fact, there was also a significant quadratic effect for that group. When the authors used an ordinal model instead, they found no evidence of interaction. For other such examples, see McKelvey and Zavoina (1975, Sec. 4) and Winship and Mare (1984).

1.3.4 Ordinal Methods with Truly Quantitative Data

Even when the response variable is interval scale rather than ordered categorical, ordinal models can still be useful. One such case occurs when the response outcome is a count but when standard sampling models for counts, such as the Poisson, do not apply. For example, each year the British Social Attitudes Survey asks a sample of people their opinions on a wide range of issues. In several years the survey asked whether abortion should be legal in each of seven situations, such as when a woman is pregnant as a result of rape. The number of cases to which a person responds "yes" is a summary measure of support for legalized abortion. This response variable takes values between 0 and 7. It is inappropriate to treat it as a binomial variate because the separate situations would not have the same probability of a "yes" response or have independent responses. It is inappropriate to treat it as a Poisson or negative binomial variate, because there is an upper bound for the possible outcome, and at some settings of explanatory variables most observations could cluster at the upper limit of 7. Methods for ordinal data *are* valid, treating each observation as a single multinomial trial with eight ordered categories.

For historical purposes it is interesting to read the extensive literature of about 40 years ago, much of it in the social sciences, regarding whether it is permissible to assign scores to ordered categories and use ordinary regression methods. See, for example, Borgatta (1968), Labovitz (1970), and Kim (1975) for arguments in favor and Hawkes (1971), Mayer (1971), and Mayer and Robinson (1978) for arguments against.

1.4 ORGANIZATION OF THIS BOOK

The primary methodological emphasis in this book is on models that describe associations and interactions and provide a framework for making inferences. In Chapter 2 we introduce ordinal odds ratios that are natural parameters for describing most of these models. In Chapter 3 we introduce the book's main focus, presenting logistic regression models for the cumulative probabilities of an ordinal response. In Chapter 4 we summarize other types of models that apply a logit link function to ordinal response variables, and in Chapter 5 we present other types of link functions for such models.

The remainder of the book deals with multivariate ordinal responses. In Chapter 6 we present loglinear and other models for describing association and interaction structure among a set of ordinal response variables, and in Chapter 7 present bivariate ordinal measures of association that summarize the entire structure by a single number. The following three chapters deal with multivariate ordinal responses in which each response has the same categories, such as happens in longitudinal studies and other studies with repeated measurement. This topic begins in Chapter 8 with methods for square contingency tables having ordered rows and the same ordered columns and considers applications in which such tables arise. Chapters 9 and 10 extend this to an analysis of more general forms of correlated, clustered ordinal responses. Primary attention focuses on models for the marginal components of a multivariate response and on models with random effects for the clusters.

In Chapters 2 to 10 we take a frequentist approach to statistical inference, focusing on methods that use only the likelihood function. In the final chapter we show ways of implementing Bayesian methods with ordinal response variables, combining prior information about the parameters with the likelihood function to obtain a posterior distribution of the parameters for inference. The book concludes with an overview of software for the analysis of ordered categorical data, emphasizing R and SAS.

For other surveys of methods for ordinal data, see Hildebrand et al. (1977), Agresti (1983a, 1999), Winship and Mare (1984), Armstrong and Sloan (1989), Barnhart and Sampson (1994), Clogg and Shihadeh (1994), Ishii-Kuntz (1994), Ananth and Kleinbaum (1997), Scott et al. (1997), Johnson and Albert (1999), Bender and Benner (2000), Guisan and Harrell (2000), Agresti and Natarajan (2001), Borooah (2002), Cliff and Keats (2002), Lall et al. (2002), Liu and Agresti (2005), and O'Connell (2006).

CHAPTER 2

Ordinal Probabilities, Scores, and Odds Ratios

In this chapter we introduce ways of using odds ratios and other summary measures to describe the association between two ordinal categorical variables. The measures apply to sample data or to a population. We also present confidence intervals for these measures. First, though, we introduce some probabilities and scores that are a basis of ways of describing marginal and conditional distributions of ordinal response variables.

2.1 PROBABILITIES AND SCORES FOR AN ORDERED CATEGORICAL SCALE

For an ordinal response variable Y, let c denote the number of categories. For n observations in a sample, $n_1, n_2, \ldots, n_c$ denote the frequencies in the categories, with $n = \sum_j n_j$, and $\{p_j = n_j/n\}$ denote the sample proportions.

For an observation randomly selected from the corresponding population, let π_j denote the probability of response in category j. Some measures and some models utilize the *cumulative probabilities*

$$F_j = P(Y \leq j) = \pi_1 + \cdots + \pi_j, \qquad j = 1, 2, \ldots, c.$$

These reflect the ordering of the categories, with

$$0 < F_1 < F_2 < \cdots < F_c = 1.$$

2.1.1 Types of Scores for Ordered Categories

How can summary measures utilize the ordinal nature of the categorical scale? One simple way uses the cumulative probabilities to identify the *median* response:

namely, the minimum j such that $F_j \geq 0.50$. With a categorical response, an unappealing aspect of this measure for making comparisons of groups is its discontinuous nature: Changing a tiny bit of probability can have the effect of moving the median from one category to the next. Also, two groups can have the same median even when an underlying latent variable has distribution shifted upward for one group relative to the other.

Alternatively, we could assign ordered scores

$$v_1 < v_2 < \cdots < v_c$$

to the categories and summarize the observations with ordinary measures for quantitative data such as the *mean*. Doing this treats the ordinal scale as an interval scale. There is no unique way to select scores, and the key aspect is the choice for the relative distances between pairs of adjacent categories. For example, with $c = 3$, comparisons of means for two groups using the scores (1, 2, 3) yields the same substantive conclusions as using the scores (0, 5, 10) or any set of linearly transformed scores but possibly different conclusions from using scores such as (1, 2, 5) or (0, 3, 10). Often, an appropriate choice of scores is unclear. In that case it is advisable to perform a sensitivity analysis: Choose scores in a few sensible ways that are not linear translations, and check whether conclusions for the method that uses those scores depend on the choice.

An alternative approach to selecting scores uses the data themselves to determine the scores. One such set uses the average cumulative proportions for the ordinal response variable. For sample proportions $\{p_j\}$, the average cumulative proportion in category j is

$$a_j = \sum_{k=1}^{j-1} p_k + \frac{1}{2} p_j, \qquad j = 1, 2, \ldots, c,$$

that is, the proportion of subjects below category j plus half the proportion in category j. In terms of the sample cumulative proportions $\hat{F}_j = p_1 + \cdots + p_j$,

$$a_j = \frac{\hat{F}_{j-1} + \hat{F}_j}{2},$$

with $\hat{F}_0 = 0$. Bross (1958) introduced the term *ridits* for the average cumulative proportion scores.

The ridits have the same ordering as the categories, $a_1 \leq a_2 \leq \cdots \leq a_c$. Their weighted average with respect to the sample distribution satisfies

$$\sum_{j=1}^{c} p_j a_j = \sum_{j=1}^{c} p_j \left(\sum_{k=1}^{j-1} p_k + \frac{1}{2} p_j \right)$$
$$= \frac{2 \sum \sum_{k<j} p_j p_k + \sum_j p_j^2}{2} = \frac{\left(\sum_j p_j \right)^2}{2} = 0.50.$$

The ridits are linearly related to the *midranks*, which are the averages of the ranks that would be assigned if the observations in a category could be ranked without ties. The midrank r_1 for category 1 is the average of the ranks $1, \ldots, n_1$ that pertain to the n_1 observations in category 1, so $r_1 = (1 + n_1)/2$. The midrank for category 2 is $r_2 = [(n_1 + 1) + (n_1 + n_2)]/2$. Generally, the midrank for category j is

$$r_j = \frac{\left[\left(\sum_{i=1}^{j-1} n_i\right) + 1\right] + \sum_{i=1}^{j} n_i}{2}.$$

Whereas midrank scores fall between 1 and n, ridit scores fall between 0 and 1. The linear relationship between them is

$$r_j = na_j + 0.5, \qquad a_j = \frac{r_j - 0.5}{n}.$$

Ridit and midrank scores take directly into account the way the response is categorized. For example, if two adjacent categories are combined, the ridit (or midrank) score for the new category falls between the original two scores, with the other scores being unaffected. If the category ordering is reversed, the ridit score for category j transforms from a_j to $1 - a_j$.

Another way to form data-dependent scores assumes a particular distribution for an unobserved continuous *latent variable* assumed to underlie Y. This approach regards the ordinal scale as representing a partition of intervals of values of the latent variable. For example, suppose that we assume an underlying standard normal distribution, with cumulative distribution function Φ. Then we could use some variation of *normal scores* as applied in some nonparametric statistical methods. For example, we could let v_1 be the mean of the truncated normal distribution falling between $-\infty$ and $\Phi^{-1}(p_1)$ [where $\Phi^{-1}(p_1)$ denotes the standard-normal score for which the cumulative probability below it equals p_1], let v_2 be the mean of the truncated normal distribution falling between $\Phi^{-1}(p_1)$ and $\Phi^{-1}(p_1 + p_2)$, and so on, up to v_c, which is the mean of the truncated normal distribution falling between $\Phi^{-1}(p_1 + \cdots + p_{c-1})$ and ∞. More simply, we could let $v_j = \Phi^{-1}(a_j)$ where a_j is the ridit score in category j. A very similar score based on the midranks $\{r_j\}$ is $v_j = \Phi^{-1}[r_j/(n + 1)]$.

We used scores in this section to summarize ordinal data, but it is not necessary to do so. In this chapter we learn about other methods that do not require assigning scores, and this is also true of most models for ordinal response variables presented in later chapters.

2.1.2 Example: Belief in Heaven

Every other year, the National Opinion Research Center at the University of Chicago conducts the General Social Survey (GSS). This survey of adult Americans provides data about the opinions and behaviors of the American public. It is simple to download results from the surveys.[1] In this book we use several data sets from the GSS to illustrate methods.

[1]This can currently be done at sda.berkeley.edu/GSS.

TABLE 2.1. Responses About Belief in Heaven

	Does Heaven Exist?				
	Definitely	Probably	Probably Not	Definitely Not	Total
Count	1546	498	205	138	2387
Proportion	0.648	0.209	0.086	0.058	1.0
Ridit score	0.324	0.752	0.899	0.971	

Source: General Social Survey.

Table 2.1 shows results of 2387 responses from the GSS to a question about whether heaven exists. The ridit scores for the counts in this ordinal categorical scale are

$$a_1 = \left(\frac{1}{2}\right)\frac{1546}{2387} = 0.32, \qquad a_2 = \frac{1546}{2387} + \left(\frac{1}{2}\right)\frac{498}{2387} = 0.75,$$

$$a_3 = \frac{1546 + 498}{2387} + \left(\frac{1}{2}\right)\frac{205}{2387} = 0.90,$$

$$a_4 = \frac{1546 + 498 + 205}{2387} + \left(\frac{1}{2}\right)\frac{138}{2387} = 0.97.$$

The ridit scores of 0.90 for "probably not" and 0.97 for "definitely not" are relatively close. Whenever two adjacent categories both have relatively small proportions, this necessarily happens.

The normal scores based on ridits, $v_j = \Phi^{-1}(a_j)$, are (−0.457, 0.681, 1.277, 1.897), where, for example, $\Phi(-0.457) = a_1 = 0.32$ is the probability that a standard normal variable falls below −0.457. The very similar normal scores based on midranks, $v_j = \Phi^{-1}[r_j/(n+1)]$, are (−0.457, 0.680, 1.276, 1.894), where, for example, $\Phi(-0.457) = [(1 + 1546)/2]/2388 = 0.324$.

This example illustrates that ridit scores or scores based on them, such as normal scores, need not represent an underlying scale realistically. For the ridit scores (0.32, 0.75, 0.90, 0.97) for (definitely, probably, probably not, definitely not), the score of 0.75 for "probably" is closer to the score of 0.97 for "definitely not" than it is to the score of 0.32 for "definitely." Yet we would not be likely to regard "probably" and "definitely not" as closer together than "probably" and "definitely." Similarly, note that the normal scores treat "definitely" and "probably" as being nearly twice as far apart as "probably" and "probably not" or "probably not" and "definitely not."

For descriptive summaries of this ordinal scale, such as comparing mean responses for different groups, it is often more sensible to use fixed scores instead of ridit scores or normal scores. The scores (1, 2, 3, 4) would treat (definitely, probably, probably not, definitely not) as equidistant for pairs of adjacent categories. Scores such as (0, 1, 4, 5) would treat the distance between "probably" and "probably not" as greater than the distance between "definitely" and "probably" and the distance between "probably not" and "definitely not."

2.1.3 Two-Way Contingency Tables with an Ordinal Response

In practice, observations on ordinal response variables are usually accompanied by observations on explanatory variables and are sometimes accompanied by observations on other response variables. When the other variables are categorical, a contingency table can display the frequencies of observations for the various combinations of levels of the variables. Each cell in the contingency table shows the number of observations that have that combination. In this chapter we consider primarily the case of two categorical variables. We denote the second variable by X if it is another response variable and by x if it is an explanatory variable. We let r denote the number of rows and let c denote the number of columns in the contingency table. Let n_{ij} denote the number of observations in the cell of the table in row i and column j.

For a two-way cross-classification of an ordinal response variable Y with another categorical response variable X, let $\{p_{ij}\}$ denote the cell proportions for the possible values of (X, Y). That is, $p_{ij} = n_{ij}/n$, where n is the total sample size. Then $\sum_i \sum_j p_{ij} = 1$, and $\{p_{ij}\}$ is the sample *joint distribution*. The sample marginal distributions are the row totals and column totals obtained by summing the joint proportions. We denote marginal proportions by p_{i+} for row i and p_{+j} for column j. Note that $p_{+j} = \sum_i p_{ij} = \sum_i n_{ij}/n$ and $\sum_j p_{+j} = 1$.

Although the second variable could also be a response variable, more commonly it is an explanatory variable. Then *conditional distributions* for the response variable are usually more relevant than joint distributions. We let the columns refer to the ordinal response variable Y and the rows refer to the explanatory variable x. For the observations in row i, we denote the proportion in category j of Y by $p_{j|i}$. Hence, $p_{j|i} = n_{ij}/n_{i+}$, where n_{i+} is the total count in row i and $\sum_j p_{j|i} = 1$ for each i. The values $(p_{1|i}, p_{2|i}, \ldots, p_{c|i})$ form a sample conditional distribution. Different levels of x can be compared with respect to the proportions of observations in the various categories of Y. The sample conditional cumulative proportions,

$$\hat{F}_{j|i} = p_{1|i} + \cdots + p_{j|i}, \qquad j = 1, 2, \ldots, c,$$

specify the proportion of observations classified in one of the first j columns, given classification in row i.

2.1.4 Probabilistic Comparisons of Two Ordinal Distributions

Now consider the special case of a $2 \times c$ table, for comparing two groups on an ordinal response variable Y. Let Y_1 and Y_2 denote the column numbers of the response variable for subjects selected at random from rows 1 and 2, independent of each other. A measure that summarizes their relative size is

$$\alpha = P(Y_1 > Y_2) + \tfrac{1}{2}P(Y_1 = Y_2) \tag{2.1}$$

(Kruskal 1957; Klotz 1966). If Y_1 and Y_2 are identically distributed or if they have symmetric distributions over all c categories, then $\alpha = 0.50$. When $\alpha > 0.50$ (< 0.50), outcomes of Y_1 tend to be larger (smaller) than outcomes of Y_2.

A related measure that has null value equal to 0 rather than 0.50 is

$$\Delta = P(Y_1 > Y_2) - P(Y_2 > Y_1). \tag{2.2}$$

The measures α and Δ are functionally related,

$$\alpha = \frac{\Delta + 1}{2}, \qquad \Delta = 2\alpha - 1,$$

with α having range [0, 1] and Δ having range [−1, 1]. We refer to them as measures of *stochastic superiority*, a term introduced by Vargha and Delaney (1998). In Chapter 7 we present related measures for $r \times c$ tables.

With sample data we can estimate α from the conditional distributions by

$$\hat{\alpha} = \sum_{j>k}\sum p_{j|1} p_{k|2} + \tfrac{1}{2} \textstyle\sum_j p_{j|1} p_{j|2}.$$

The sample version of Δ is

$$\hat{\Delta} = \sum_{j>k}\sum p_{j|1} p_{k|2} - \sum_{j<k}\sum p_{j|1} p_{k|2}.$$

Another useful comparison of $P(Y_1 > Y_2)$ and $P(Y_2 > Y_1)$ is

$$\theta = \frac{P(Y_1 > Y_2)}{P(Y_2 > Y_1)}.$$

Its sample value is

$$\hat{\theta} = \frac{\sum\sum_{j>k} p_{j|1} p_{k|2}}{\sum\sum_{j<k} p_{j|1} p_{k|2}} = \frac{\sum\sum_{j>k} n_{1j} n_{2k}}{\sum\sum_{j<k} n_{1j} n_{2k}}.$$

When $c = 2$, $\hat{\theta}$ is an odds ratio. For $c > 2$, $\hat{\theta}$ is a generalized odds ratio for ordinal responses (Agresti 1980), which we refer to as an *ordinal odds ratio* for comparing two groups. In Section 2.2 we introduce other ways of forming odds ratios for ordinal responses.

The ordinal odds ratio θ differs slightly from

$$\frac{\alpha}{1-\alpha} = \frac{P(Y_1 > Y_2) + \frac{1}{2} P(Y_1 = Y_2)}{P(Y_2 > Y_1) + \frac{1}{2} P(Y_1 = Y_2)},$$

which approximates $P(Y_1 > Y_2)/P(Y_2 > Y_1)$ for an underlying continuous scale. The measure $\alpha/(1-\alpha)$ is closer to 1.0 than is θ. Similarly, usually $P(Y_1 > Y_2)/P(Y_2 > Y_1)$ for the underlying continuous scale is closer to 1.0 than θ is for the observed ordinal scale. This is because observations that are tied on the observed ordinal scale usually have similar relative frequencies of the two orders for the underlying scale. By contrast, we can interpret α or Δ either for the observed scale or an underlying continuous scale. For example, suppose that $\Delta = 0.40$. Then, in comparisons of the groups with independent observations for the underlying continuum, we expect a higher response for group 1 about 70% of the time and a higher response for group 2 about 30% of the time, since $0.70 - 0.30 = 0.40$ and $0.70 + 0.30 = 1.0$.

2.1.5 Means of Conditional Distributions in Two-Way Tables

Next we consider $r \times c$ tables. With ordered scores $\{v_j\}$ for the categories of Y, in each row we can use the sample conditional distribution to find a sample mean response. In row i this is

$$\bar{y}_i = \sum_{j=1}^{c} v_j p_{j|i}.$$

When x is ordinal, we often expect a trend (upward or downward) in $\{\bar{y}_i\}$ across the rows.

Alternatively, we could find the means using data-generated scores. For example, we could use ridit scores for Y calculated from the proportions in its marginal distribution. For outcome category j,

$$a_j = \sum_{k=1}^{j-1} p_{+k} + \tfrac{1}{2} p_{+j}, \qquad j = 1, 2, \ldots, c.$$

The *mean ridit* for the sample conditional distribution in row i is

$$\bar{A}_i = \sum_{j=1}^{c} a_j p_{j|i}.$$

The weighted average of the mean ridits satisfies

$$\sum_{i=1}^{r} p_{i+} \bar{A}_i = 0.50.$$

When the data in the full sample are ranked, using midranks $\{r_j\}$, the *mean rank* for the sample conditional distribution in row i is

$$\bar{R}_i = \sum_{j=1}^{c} r_j p_{j|i}.$$

Their weighted average over the r rows is $(n+1)/2$. The mean ridits and mean ranks are related by

$$\bar{A}_i = \frac{\bar{R}_i - 0.50}{n}.$$

Bross (1958) argued that an advantage of ridit scoring is their lack of sensitivity to the way the ordinal response variable is categorized (e.g., with different numbers of categories). Two researchers who categorize an ordinal response in different ways for a particular sample would, nevertheless, obtain similar mean ridits for the rows.

2.1.6 Mean Ridits and Mean Ranks Relate to Stochastic Superiority Measures

For $2 \times c$ tables, the sample values of the stochastic superiority measures $\alpha = P(Y_1 > Y_2) + \frac{1}{2}P(Y_1 = Y_2)$ and $\Delta = P(Y_1 > Y_2) - P(Y_2 > Y_1)$ relate to the mean ridit scores in the two rows by[2]

$$\hat{\alpha} = (\bar{A}_1 - \bar{A}_2 + 0.50) \quad \text{and} \quad \hat{\Delta} = 2(\bar{A}_1 - \bar{A}_2).$$

Vigderhous (1979) presented other connections between mean ridit measures and ordinal measures of association. In terms of the mean ranks $\bar{R}_1$ and $\bar{R}_2$ in the two rows,

$$\hat{\alpha} = \frac{\bar{R}_1 - \bar{R}_2}{n} + 0.50 \quad \text{and} \quad \hat{\Delta} = \frac{2(\bar{R}_1 - \bar{R}_2)}{n}.$$

For $r \times c$ tables, let Y_i denote the response outcome for a randomly selected subject at level i of x, and let Y^* denote the response outcome for a randomly selected subject from the marginal distribution of Y. The sample mean ridit $\bar{A}_i$ using the marginal ridit scores estimates

$$P(Y_i > Y^*) + \tfrac{1}{2}P(Y_i = Y^*).$$

In analogy with the terms *logit* and *probit*, Bross (1958) chose the term *ridit* because $\bar{A}_i$ describes how the distribution of Y in row i compares *r*elative to an *i*dentified *d*istribution (in this case, the marginal distribution of Y). The $\{\bar{A}_i\}$ or the corresponding population values can be used to compare each row to an overall marginal distribution of the response (Kruskal 1952). In some of the literature on nonparametric statistical methods they are referred to as *relative effects*.

For underlying continuous distributions, $\bar{A}_i$ estimates the probability that an observation from row i ranks higher on the ordinal response variable than does an observation from the marginal distribution of Y. Such a probability inference is approximate, since besides sampling error, it is unknown how tied observations for the observed discrete scale would be ordered for an underlying continuum. Also, the sample marginal distribution of Y, which determines the ridit scores, reflects the study design. For some sampling schemes, this need not be close to the population marginal distribution.

[2]For fully ranked data, analogous connections exist between Wilcoxon statistics using mean ranks and Mann–Whitney statistics using pairwise orderings.