Psychological Bulletin
1989, Vol. 105, No. 2, 290–301

0033-2909/89/$00.75

Tutorial on Modeling Ordered Categorical Response Data

Alan Agresti
Department of Statistics, University of Florida

In the past decade there has been great progress in the development of methodology for analyzing ordered categorical data. Logit and log linear model-building techniques for nominal data have been generalized for use with ordinal data. There are many advantages to using these procedures instead of the Pearson chi-square test of independence to analyze ordered categorical data. These advantages include (a) more complete description of the nature of associations and (b) greater power for detecting population associations. This article introduces logit models for categorical data and shows two ways of adapting them to model ordered categorical data. The models are used to analyze a cross-classification table relating mental impairment and parents' socioeconomic status.

In his presidential address delivered at the annual meeting of the American Statistical Association in 1967, Frederick Mosteller stated, "I fear that the first act of most social scientists upon seeing a contingency table is to compute chi-square for it. Sometimes this process is enlightening, sometimes wasteful, but sometimes it does not go quite far enough" (Mosteller, 1968). Twenty years later, methodology for analyzing categorical data has advanced considerably, and most social scientists realize that there is more that can be done besides computing chi-square. Nevertheless, it is still true that analysis "sometimes does not go quite far enough." This applies particularly to distinguishing between methods for nominal and ordinal variables. Social scientists often analyze variables that have ordered categories using methods (such as the Pearson chi-square test) designed for variables that have unordered categories. This practice results in loss of power for both description and inference.

Social scientists are not to blame for this situation. Nearly all elementary statistics books introduce Pearson's chi-square statistic for testing independence of categorical variables; few of those books point out that the Pearson test is generally inappropriate when at least one of the classifications is ordered. Even the book by Bishop, Fienberg, and Holland (1975), regarded as the bible for multivariate analysis of categorical data, makes little distinction between methods for nominal and ordinal data.

One reason for the common lack of differentiation between nominal and ordinal analyses is that specialized methods for ordered categorical data have become well developed only within the past 10 years. Articles by Goodman (1979) and McCullagh (1980) have had a major impact on this development. These articles have extended log linear and logistic regression models so they can be applied to ordinal data.

The purpose of this article is to show how research psychologists can use some of these recent developments to improve the way they analyze ordinal data. The article should be technically accessible to researchers having an understanding of basic applied statistics, including regression modeling and the chi-square test. Detailed knowledge of logistic regression or loglinear models is not needed, though exposure to these topics at the level presented in the introductory statistics textbook by Agresti and Finlay (1986) or in the review article by Swafford (1980) is certainly helpful.

I first discuss a measure of association, the odds ratio, used to interpret models discussed in this article. After introducing the logit model for binary responses, I use it to model cumulative probabilities for ordinal response variables. I then introduce an alternative logit model, for which the odds ratio interpretion extends simultaneously to all pairs of adjacent response categories. I also show how to construct statistics for testing independence that utilize the ordering of categories and are more powerful than the standard Pearson statistic.

To illustrate the models, I analyze Table 1, taken from Srole et al. (1978, p. 289), relating mental impairment to parents' socioeconomic status. Both classifications in this table have ordered categories, and the 24 cell counts can be well represented by a single odds ratio measure that describes the nature and structure of the relationship.

Odds Ratios

First, I discuss some basic principles. For simplicity, these are illustrated for two-way contingency tables. The most commonly applied analysis for two-way tables is the test of independence, usually conducted with Pearson's statistic

$$X^2 = \Sigma\,(\text{Observed} - \text{Expected})^2/\text{Expected}.$$

Here, the observed counts in cells of the table are compared to expected values satisfying independence. Large values of this chi-square statistic contradict the null hypothesis of independence.

Suppose the contingency table has r rows and c columns. Then the degrees of freedom for the Pearson statistic are $df = (r-1)(c-1)$. This means there are $(r-1)(c-1)$ bits of information in the table regarding the association. Pearson's statistic

Correspondence concerning this article should be addressed to Alan Agresti, Department of Statistics, University of Florida, Gainesville, Florida 32611.

is designed to detect whether *any* of these bits indicate that the variables are dependent.

There are many ways to represent the $(r-1)(c-1)$ bits of information in a table. I now present a way that is useful when one of the variables (say, the column variable) is a response (dependent) variable, measured on an ordinal scale. Denote this variable by Y. Let $P(Y \leq j)$ denote the probability of a response in category j or below (i.e., in Category 1, 2, $\cdots$, or j of the column classification). This is called a *cumulative* probability. The ratio

$$P(Y \leq j)/P(Y > j)$$

is called the *odds* of response less than or equal to j. For instance, an odds equal to 3 means that subjects are three times as likely to make a response in category j or below than they are to make a response above category j.

Suppose the row variable is an explanatory (independent) variable, and denote it by X. Let Y_i denote the value of Y for a subject randomly selected from level i of X. The odds of response less than or equal to j can be compared for two rows, say, Rows 1 and 2, by the ratio

$$\lambda = \frac{P(Y_1 \leq j)/P(Y_1 > j)}{P(Y_2 \leq j)/P(Y_2 > j)}.$$

This measure is called a cumulative *odds ratio*. If $\lambda = 4$, for instance, the odds of response less than or equal to j are four times higher in the first row than in the second row. In a table having r rows and c columns, there are $(r-1)$ pairs of adjacent rows that can be compared like this. Also, there are $(c-1)$ ways of splitting the response into two parts (i.e., the cut point j could be any of 1, 2, $\cdots$, $c-1$). Therefore, there are $(r-1)(c-1)$ cumulative odds ratios that can be formed.

X and Y are statistically independent if all $(r-1)(c-1)$ of the cumulative odds ratios equal 1.0 or equivalently, if the logarithms of all of these odds ratios equal zero. Values of cumulative odds ratios can be used to describe the nature of the association between an ordinal response variable and a nominal or ordinal explanatory variable. For instance, let

Table 1
Cross-Classification of Mental Impairment and Parents' Socioeconomic Status

Parents' socioeconomic status	Mental impairment			
	Well	Mild symptom formation	Moderate symptom formation	Impaired
A (high)	64	94	58	46
B	57	94	54	40
C	57	105	65	60
D	72	141	77	94
E	36	97	54	78
F (low)	21	71	54	71

Note. From *Mental Health in the Metropolis: The Midtown Manhattan Study* (p. 289) by L. Srole, T. S. Langner, S. T. Michael, P. Kirkpatrick, M. K. Opler, and T. A. C. Rennie, 1978, New York: NYU Press. Copyright 1978 by New York University Press. Reprinted by permission.

Table 2
Cumulative Odds Ratios for Table 1, Treating Mental Impairment as Response Variable

Pair of socioeconomic status levels	Cut point between categories		
	Well, mild	Mild, moderate	Moderate, impaired
A, B	1.07	.95	.92
B, C	1.22	1.24	1.35
C, D	1.07	1.04	1.23
D, E	1.47	1.24	1.29
E, F	1.47	1.37	1.17

$$\lambda_{ij} = \frac{P(Y_i \leq j)/P(Y_i > j)}{P(Y_{i+1} \leq j)/P(Y_{i+1} > j)}$$

for $i = 1, \cdots, r-1$ and $j = 1, \cdots, c-1$. When $\lambda_{ij} > 1$, subjects at level i of X are more likely to make lower responses on Y (i.e., category j or below) and less likely to make higher responses on Y, in comparison with subjects at level $i+1$ of X. The logarithm of the cumulative odds ratio is then positive. Thus when these log odds ratios exceed 0 for all pairs of adjacent rows and all cut points for the response, larger values of Y are more likely to occur at higher levels of X, and the association is characterized as positive.

For cross-classifications of two ordinal variables, there is often a positive or a negative trend in the association, in the sense that all or nearly all of these log odds ratios have the same sign. Then it may be possible to summarize the $(r-1)(c-1)$ bits of information about the association by a single number, for instance, by an average of the $\{\lambda_{ij}\}$. The models I present in this article provide a structure for summarizing the $(r-1)(c-1)$ odds ratios by a single bit (or at least fewer bits) of information.

The benefits of summarizing the association by fewer bits of information include (a) model parsimony and (b) improved power for detecting important alternatives to independence. These can be simply explained by analogy to regression analysis for continuous variables. Suppose one wants to test statistical independence. If the variables are statistically dependent, one often expects that as X goes up, Y tends to go up or as X goes up, Y tends to go down. Thus one normally fits a linear regression model and tests independence by testing that the slope coefficient equals zero. Though it may not perfectly describe the relationship, the linear model is simple to interpret: It is parsimonious. An alternative approach for testing independence is to fit a high-order polynomial and test that all coefficients (other than the Y intercept) equal zero. Though this more complex model fits better, one rarely uses this approach, because one expects the linear term to explain the major portion of the trend. If the relationship is close to linear, the test for the slope in the linear model is more powerful for detecting statistical dependence than is a more general test regarding several parameters, most of which are near zero. The same principles apply in analyzing categorical data. I prefer to use simple models, partly because interpretations are simpler and partly because inferences based on them are more powerful.

Table 2 gives sample values of the $\{\lambda_{ij}\}$ for the data in Table

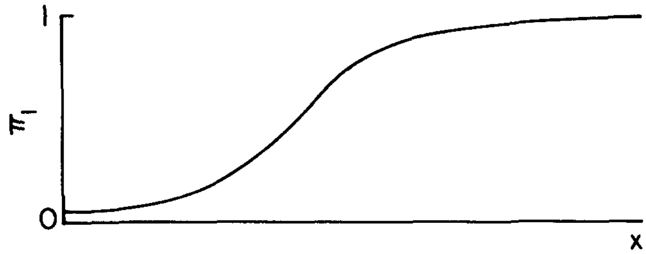

Figure 1. Logistic regression model for binary response.

1, treating Y = Mental impairment as the response variable. The first entry in Table 2 is 1.07. The odds that mental health is classified in the *well* category are estimated to be 1.07 times higher for socioeconomic status (SES) Level A than for SES Level B. Thus, in this sample, mental health is slightly more likely to be well at the higher SES level (Level A). In fact, in Table 2 all but two of the cumulative odds ratios exceed 1.0, and those two are close to 1.0. Thus, there seems to be a tendency for mental health to be better at the higher SES levels. Later in this article, I show that the sample $\{\lambda_{ij}\}$ values are not significantly different from an estimated common value of 1.18. The models that are presented next enable one to determine when information can be condensed in this way.

Logit Model for Ordinal Response

This section describes a model for ordered categorical response variables that is interpreted using cumulative odds ratios. I begin by reviewing the simpler case in which Y has only two categories. Let π_1 denote the probability of classification in the first of these categories. When Y is predicted by a quantitative variable X, the model

$$\pi_1 = \alpha + \beta x$$

corresponds to a linear regression model with a dummy variable for Y. This is an unsuitable model for a binary response, because probabilities are restricted to fall between 0 and 1, whereas linear predictors take values over the entire real line.

An S-shaped curve of the form shown in Figure 1 is generally a more appropriate response shape for a model. The formula

$$\pi_1 = \frac{\exp(\alpha + \beta x)}{[1 + \exp(\alpha + \beta x)]}$$

gives a curve of this shape, where $\exp(z)$ denotes the exponential function e^z, the antilog function for natural logarithms. This model for π_1 is called a *logistic regression model.* The probability π_1 increases as x increases when $\beta > 0$ (as in Figure 1), and it decreases as x increases when $\beta < 0$. For particular parameter values α and β, $\pi_1 = 0.5$ when $x = -\alpha/\beta$. Because the relationship is curvilinear, the slope of a line tangent to the curve changes as x changes. At a particular x value, that slope is $\beta\pi_1(1 - \pi_1)$, which depends on the probability of making Response 1 at that x value; this slope is greatest in absolute value where $\pi_1 = 0.5$. Thus, for this model, the effect on π_1 of an incremental change in x is less when π_1 is near 0 or 1 than it is when π_1 is near the middle of its range.

The logistic regression formula for π_1 is not especially simple to interpret. It is easier to interpret the corresponding formula for the odds, which is

$$\pi_1/(1 - \pi_1) = \exp(\alpha + \beta x) = e^{\alpha}(e^{\beta})^{x}.$$

In other words, there is an exponential relationship between the odds and the explanatory variable. The term e^{β} is a multiplicative effect on the odds: For every unit change in x, the odds $\pi_1/(1 - \pi_1)$ change by a multiplicative factor of e^{β}. If $\beta = 0$, then $e^{\beta} = 1$, and the odds do not change as x changes: X and Y are independent. If $\beta > 0$, then $e^{\beta} > 1$, and the odds increase as x increases; that is, π_1 is larger at higher values of x.

The *logit* for a binary response is defined as

$$\text{logit}(\pi_1) = \log[\pi_1/(1 - \pi_1)],$$

which is the log of the odds of making Response 1. For the logistic regression model, the logit transformation (with natural logarithm) linearizes the relationship, giving

$$\text{logit}(\pi_1) = \alpha + \beta x.$$

In this form, the model is called a *linear logit model.* Like other regression models, the linear logit model can be generalized to include multiple explanatory variables (dummy variables being used to represent nominal factors), and effects of explanatory variables can be tested using the ratio of parameter estimates to their standard errors.

Next, one can apply these ideas to an ordinal response variable by forming logits of cumulative probabilities. Denote the possible values of Y by $1, \cdots, c$, and denote their probabilities by $\pi_1, \cdots, \pi_c$. Then

$$\text{logit}[P(Y \le j)] = \log\left(\frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_c}\right), \qquad j = 1, \cdots, c - 1,$$

treats the response as binary by combining the first j categories and by combining the remaining $(c - j)$ categories. There are $(c - 1)$ of these logits, one for each possible cut point for collapsing the response. I shall refer to them as *cumulative logits.*

A model that simultaneously describes all $c - 1$ cumulative logits is

$$\text{logit}[P(Y \le j)] = \alpha_j + \beta x, \qquad j = 1, \cdots, c - 1.$$

For fixed j, this has a similar form as the linear logit model just discussed for the collapsing of the response into two categories. As j increases, the α_j parameters increase, reflecting the increase in the logits as additional probabilities are added into the numerator. The parameter of primary interest is β. It describes the effect of X, equaling 0 when X has no effect on Y. If $\beta > 0$, cumulative probabilities tend to be higher at higher values of X. That is, the likelihood that Y is below any fixed level is relatively greater at higher values of X; so, small Y values are relatively more likely at large X values. Alternatively, the right-hand side of the model formula could be expressed as $\alpha_j - \beta x$, or the logit could be defined with $P(Y > j)$ in the numerator and $P(Y \le j)$ in the denominator; with either change, a positive β corresponds

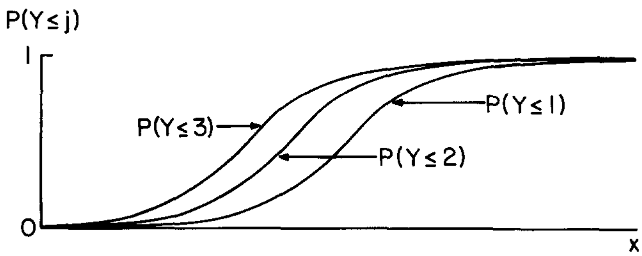

Figure 2. Depiction of cumulative logit model with effect independent of cut point.

to a positive association, in the sense that large Y values are relatively more likely to occur at large X values.

For this model, β has the same value for each of the $c - 1$ cumulative logits. This means that the effect of X is assumed to be the same for each cumulative probability; it does not depend on the cut point for forming the logit. The parameter e^{β} is interpreted as a multiplicative effect on the cumulative odds: The odds that the response is less than or equal to j (for any fixed j) is multiplied by e^{β} for every unit change in X. Figure 2 depicts this model for the case of $c = 4$ responses, for which there are three cumulative logits. The constant value for β means that the response curves for the cumulative probabilities are assumed to have the same shape. The curve for $P(Y \leq 1)$ is the curve for $P(Y \leq 2)$ moved $(\alpha_2 - \alpha_1)/\beta$ units to the right; the curve for $P(Y \leq 3)$ equals the curve for $P(Y \leq 1)$ moved $(\alpha_1 - \alpha_3)/\beta$ units to the right. This single model simultaneously describes three relationships: for the effect of X on the odds that $Y \leq 1$ instead of $Y > 1$; for the effect of X on the odds that $Y \leq 2$ instead of $Y > 2$; and for the effect of X on the odds that $Y \leq 3$ instead of $Y > 3$. When it is fitted, one gets a single estimate of β (rather than three separate estimates, as one would get if one fitted the binary logit model separately for each collapsing of the response).

I have described the cumulative logit model for an ordinal response Y and a continuous explanatory variable X. When X is also ordinal, the data are counts in a two-way contingency table in which both classifications are ordered. Suppose one represents the ith level of X by the score x_i, where one sets $x_1 < x_2 < \cdots < x_r$ to reflect the ordering of the categories. The cumulative logit model is then

$$\text{logit}\,[P(Y_i \leq j)] = \alpha_j + \beta x_i.$$

For integer-spaced scores (such as $x_1 = 1, x_2 = 2, \cdots, x_r = r$), e^{β} then reflects the multiplicative effect on the cumulative odds of each row change in X.

The effect parameters in cumulative logit models are related to cumulative odds ratios. For instance, for Rows 1 and 2,

$$\text{logit}\,[P(Y_1 \leq j)] - \text{logit}\,[P(Y_2 \leq j)] = \log \frac{P(Y_1 \leq j)/P(Y_1 > j)}{P(Y_2 \leq j)/P(Y_2 > j)} = \log \lambda_{1j}.$$

For the cumulative logit model, this difference in logits equals $\beta(x_1 - x_2)$, and similar results apply to other pairs to rows. When one uses integer-spaced scores, so that $x_{i+1} - x_i = 1$ for all i, then this model implies that all $\log \lambda_{ij} = -\beta$. (For the model formulation in which the right-hand side of the equation is denoted $\alpha_j - \beta x_i$, all $\log \lambda_{ij} = \beta$.) For this cumulative logit model, then, all cumulative odds ratios for pairs of adjacent rows are equal, and their common value is $e^{-\beta}$. That is, this model implies uniform association for cumulative odds ratios.

If $\beta = 0$ in the cumulative logit model, then X and Y are independent. When this model fits adequately, one can test independence by testing that $\beta = 0$. The estimate $\hat{\beta}$ for β has approximately a normal sampling distribution, and one can use

$$z = \hat{\beta}/SE(\hat{\beta})$$

as the test statistic. Equivalently, z^2 and another statistic to be discussed later have approximate chi-square distributions with $df = 1$. Like the Pearson statistic, these statistics approximate their sampling distributions better as the sample size increases. Unlike the Pearson test, this test utilizes the ordinal nature of the variables. The statistic z (or its square) uses a single degree of freedom to search for statistical dependence in the form of an approximately constant cumulative odds ratio.

Fitting the cumulative logit model is not computationally simple, but it can be done using statistical computer packages. These details are discussed in a later section of this article, as are chi-square statistics for testing the fit of the model.

For the data in Table 1 on mental health and parents' SES, one sees that the cumulative logit model fits very well. When mental impairment is the response variable and the scores (1, 2, 3, 4, 5, 6) are assigned to the levels of SES, the uniform association model has an estimated parameter value of $\hat{\beta} = -0.167$. This corresponds to an estimate of $e^{0.167} = 1.18$ for the uniform cumulative odds ratio. One has the parsimonious result that the 15 sample odds ratios reported in Table 2 can be adequately represented by a single number, 1.18. For instance, the odds that mental health is classified *well* (instead of *mild* or *moderate symptom formation* or *impaired*) are estimated to be 1.18 times higher for each increase of a level in parents' SES. The odds that mental health is classified *well* are estimated to be 1.18 times higher for parents' SES Level A than for parents' SES Level B and are estimated to be $1.18^5 = 2.3$ times higher for parents' SES Level A than for parents' SES Level F. The standard error of the $\hat{\beta}$ value is 0.0277. The statistic $z = \hat{\beta}/(SE) = -0.167/0.0277 = -6.0$ gives very strong evidence that mental impairment tends to be less at the higher levels of parents' SES.

Incorporating Nominal or Multiple Explanatory Variables

This discussion of the cumulative logit model has focused on the case of a single explanatory variable having ordered categories. More general models can be formulated in which the effect β need not be the same for each cut point, though the common effect model is usually adequate for describing the most important component of the association. Cumulative logit models can also be formulated for nominal explanatory variables.

When the row variable X in a cross-classification table is nominal, the model

$$\text{logit}\,[P(Y_i \leq j)] = \alpha_j + \beta_i$$

is useful. The $\{\beta_i\}$ parameters for the levels of the nominal variable are usually scaled so that $\Sigma\,\beta_i = 0$ (as in analysis of variance [ANOVA] coding) or so that $\beta_r = 0$ (as in coding for parameters that are coefficients of dummy variables). These parameters are called *row effects.* Two rows that have the same β parameter value have the same distribution on Y; that is, the probability of making any given response is then the same for subjects at each of those two levels of X. Independence of X and Y corresponds to the distribution of Y being the same in each row and is the special case $\beta_1 = \cdots = \beta_r$. Cumulative probabilities in a particular row tend to be higher when the β parameter for that row is higher. The farther apart β values are for two rows, the greater the difference between their distributions on Y. The difference between β values has an interpretation using cumulative odds ratios.

The model previously discussed for an ordinal explanatory variable is the special case of this model in which the $\{\beta_i\}$ have the linear trend $\beta_i = \beta x_i$, for some set of ordered scores $\{x_i\}$. The row effects model treats X as nominal, but it might also be used when X is ordinal (a) if Y is not a linear function of X (on the logit scale) or (b) if we do not wish to impose a pattern for the association by assigning scores to the levels of X.

For illustrative purposes, apply the row effects model to Table 1. When scaled so that $\hat{\beta}_6 = 0$, the estimates of the row effect parameters are $\hat{\beta}_1 = 0.82$ ($SE = 0.17$), $\hat{\beta}_2 = 0.84$ ($SE = 0.17$), $\hat{\beta}_3 = 0.62$ ($SE = 0.16$), $\hat{\beta}_4 = 0.52$ ($SE = 0.15$), $\hat{\beta}_5 = 0.26$ ($SE = 0.16$), and $\hat{\beta}_6 = 0.0$. To compare rows 1 and 3 of Table 1, take the difference $\hat{\beta}_1 - \hat{\beta}_3 = 0.20$. For any j, the odds that mental impairment is less than or equal to j are estimated to be $e^{20} = 1.2$ times higher for SES Level A than for SES Level C. For these data, the row effect estimates show roughly a decreasing trend, indicating that the cumulative probabilities starting at the *well* end of the scale tend to decrease as parents' SES level moves toward the F end of the scale. The $\hat{\beta}_1$ and $\hat{\beta}_2$ values are nearly identical, indicating that subjects at Levels A and B of parents' SES in this sample had very similar distributions on mental impairment.

More generally, cumulative logit models can be formulated for multiple explanatory variables, some of which may be nominal or ordinal categorical and some of which may be continuous. To illustrate the interpretation of models with multiple explanatory variables, I next analyze the artificial data in Table 3, relating mental impairment to a binary measurement of socioeconomic status ($Z = 1$, high; $Z = 0$, low) and to a life events index X, a composite measure of both the number and severity of important life events (such as birth of a child, new job, or death in the family) that occurred to the subject within the past 3 years.

Consider first the main effects model,

$$\text{logit}\,[(P(Y \leq j)] = \alpha_j + \beta_1 x + \beta_2 z.$$

The parameter estimates are $\hat{\beta}_1 = -0.319$ ($SE = 0.121$) and $\hat{\beta}_2 = 1.111$ ($SE = 0.611$). The cumulative probability increases (so relatively more subjects are at low levels of mental impairment) as the life events score decreases and at the higher level of SES. At a fixed life events score, the odds that mental impairment is below any given level is estimated to be $e^{1.111} = 3.04$ times as great at the high SES level as at the low level.

Table 3
Mental Impairment by Socioeconomic Status and Life Events

Subject	Mental impairment	SES	Life events
1	Well	1	1
2	Well	1	9
3	Well	1	4
4	Well	1	3
5	Well	0	2
6	Well	1	0
7	Well	0	1
8	Well	1	3
9	Well	1	3
10	Well	1	7
11	Well	0	1
12	Well	0	2
13	Mild	1	5
14	Mild	0	6
15	Mild	1	3
16	Mild	0	1
17	Mild	1	8
18	Mild	1	2
19	Mild	0	5
20	Mild	1	5
21	Mild	1	9
22	Mild	0	3
23	Mild	1	3
24	Mild	1	1
25	Moderate	0	0
26	Moderate	1	4
27	Moderate	0	3
28	Moderate	0	9
29	Moderate	1	6
30	Moderate	0	4
31	Moderate	0	3
32	Impaired	1	8
33	Impaired	1	2
34	Impaired	1	7
35	Impaired	0	5
36	Impaired	0	4
37	Impaired	0	4
38	Impaired	1	8
39	Impaired	0	8
40	Impaired	0	9

Note. SES = socioeconomic status.

A model allowing interaction is obtained by adding the term $\beta_3 xz$ to the model. The estimates are then $\hat{\beta}_1 = -0.420$ ($SE = 0.186$), $\hat{\beta}_2 = 0.371$ ($SE = 1.136$), and $\hat{\beta}_3 = 0.181$ ($SE = 0.238$). The estimated effect of life events on the cumulative logit is $-.420$ for the low SES group and $-.239$ for the high SES group. The effect is weaker for the high SES group, though the difference in effects is not significant for this small sample.

Alternative Logit Model

There are alternative ways of constructing models for ordinal response variables. I next discuss a model based on a different way of forming logits of response proportions. I utilize the ordinality of Y by forming logits for adjacent pairs of response categories. For a cross-classification of an ordinal response Y and an ordinal explanatory variable X having assigned scores $x_1 < \cdots < x_r$, the model

$$\log\left[\frac{P(Y_i = j)}{P(Y_i = j+1)}\right] = \alpha_j + \beta x_i, \qquad j = 1, \cdots, c-1,$$

assumes a linear effect β that is the same for each adjacent pair of response categories. The variables are independent if $\beta = 0$. The odds of making response j instead of response $j + 1$ are multiplied by e^{β} for each unit change in X.

Let

$$P^*(Y_i = j) = P(Y_i = j)/[P(Y_i = j) + P(Y_i = j+1)].$$

This is the conditional probability of making response j, given that the response is either j or $j + 1$. The model has the logit form

$$\log\left[\frac{P^*(Y_i = j)}{1 - P^*(Y_i = j)}\right] = \alpha_j + \beta x_i, \qquad j = 1, \cdots, c-1,$$

for these conditional probabilities. This model can also be represented by Figure 2, if one replaces the cumulative probability label for the vertical axis with the label $P^*(Y = j)$. If $\beta > 0$, then outcome j is relatively more likely than outcome $j + 1$ at larger values of x. The case $\beta = 0$ corresponds to the statistical independence of X and Y.

Logit models for adjacent response categories utilize a different type of odds ratio to summarize the association. The odds ratio

$$\theta_{ij} = \frac{P(Y_i = j)/P(Y_i = j+1)}{P(Y_{i+1} = j)/P(Y_{i+1} = j+1)}$$

compares the odds of making the lower of two adjacent responses (j and $j + 1$) at two adjacent levels (i and $i + 1$) of the explanatory variable. There are also $(r - 1)(c - 1)$ of these odds ratios, for the different possible combinations of pairs of adjacent rows and pairs of adjacent columns. The $\{\theta_{ij}\}$ are called *local odds ratios,* because they describe associations in localized regions of the table. As was true for cumulative odds ratios, independence is equivalent to all $\theta_{ij} = 1.0$, and logit models can provide structure for their values. For instance, when both variables are ordinal, all sample $\{\theta_{ij}\}$ may have about the same magnitude; then the table can be described by an estimate for a common value of the $\{\theta_{ij}\}$.

For the logit model for adjacent response categories just discussed, all

$$\log \theta_{ij} = -\beta$$

when the x scores are integer spaced, such as $x_1 = 1, \cdots, x_r = r$. (Again, $-\beta$ is replaced by β if one uses the parameterization $\alpha_j - \beta x_i$ for the right-hand side of the model formula, or if one defines the logit with $P(Y_i = j + 1)$ in the numerator.) Thus, $e^{-\beta}$ respresents the common value of the local odds ratios. This model is a *uniform association* model for local odds ratios. For this, the odds of making response j instead of response $j + 1$ are $e^{-\beta}$ times higher in row i than in row $i + 1$, for all i and j. The hypothesis of independence can be tested using $z = \hat{\beta}/SE(\hat{\beta})$. Like the uniform association model for cumulative odds ratios, this model has one more parameter than the independence model, which is the special case of either model with $\beta = 0$.

This model also gives a good fit to Table 1. The estimate $\hat{\beta} = -0.091$ has a standard error of 0.015, so there is again strong evidence that mental impairment tends to diminish as parents' SES increases. The estimated common value of the local odds ratio is $e^{0.091} = 1.09$. The odds that mental health is well instead of showing mild symptom formation is estimated to be 1.09 times higher for SES Level A than for Level B, 1.09 times higher for Level B than for Level C, and so forth. The same statement applies to each pair of adjacent response categories. Estimated odds ratios for other comparisons are obtained by taking 1.09 to the power given by the product of the distance between the rows and the distance between the columns. For instance, the estimated odds that mental health is well instead of impaired is estimated to be $1.09^{15} = 3.9$ times higher for SES Level A than for Level F. (There are three levels between well and impaired, five levels between Levels A and F, and $15 = 3 \times 5$.) The β parameter (and its estimate) for this model differs from the β parameter for the cumulative logit model, because they refer to different types of odds ratios. However, the substantive conclusions reached with the two models are similar: There is strong evidence of a negative association between mental impairment and parents' SES; the degree of association is relatively weak; and the association has similar value throughout the table, whether measured with local or cumulative odds ratios.

When X is nominal, or when it is ordinal but does not have a linear effect on the logit, one can use the more general row effects model

$$\log\left[\frac{P(Y_i = j)}{P(Y_i = j+1)}\right] = \alpha_j + \beta_i.$$

The model just discussed is the special case in which the $\{\beta_i\}$ parameters have a linear trend, and the independence model is the special case in which all $\{\beta_i\}$ are equal. If rows a and b have parameters $\beta_a = \beta_b$, then those rows have identical distributions on Y. Generally, $\beta_a - \beta_b$ is the log odds ratio for each pair of adjacent response categories, for the cells in rows a and b.

A useful characteristic of logit models for adjacent categories is that there are equivalent log linear models. Thus some statistical computer packages for log linear models can be used to fit them. Also, the models are equivalent to other models that form logits using pairs of categories (rather than groups of categories, as does the cumulative logit model). For instance, *baseline category* logit models contrast each response category with the final category. The adjacent-categories logit model having a linear effect is equivalent to the baseline-category logit model

$$\log\left[\frac{P(Y_i = j)}{P(Y_i = c)}\right] = \gamma_j + \beta x_i(c - j), \qquad j = 1, \cdots, c-1.$$

The effect parameter β in the adjacent-categories logit model

Table 4
Estimated Expected Frequencies for Independence Model and Uniform Association Models for Local and Cumulative Odds Ratios Fitted to Table 1

	Mental impairment			
Parents' socioeconomic status	Well	Mild symptom formation	Moderate symptom formation	Impaired
A (high)	64	94	58	46
Independence model	48.5	95.0	57.1	61.4
Local odds model	65.3	104.4	50.1	42.1
Cumulative odds model	65.9	103.7	48.9	43.5
B	57	94	54	40
Independence model	45.3	88.9	53.4	57.4
Local odds model	54.2	94.9	49.9	45.9
Cumulative odds model	54.3	94.8	49.3	46.6
C	57	105	65	60
Independence model	53.1	104.1	62.6	67.3
Local odds model	55.9	107.2	61.7	62.2
Cumulative odds model	55.7	107.3	61.6	62.4
D	72	141	77	94
Independence model	71.0	139.3	83.7	90.0
Local odds model	65.3	137.0	86.4	95.3
Cumulative odds model	65.0	137.3	86.8	94.9
E	36	97	54	78
Independence model	49.0	96.1	57.8	62.1
Local odds model	39.0	89.6	61.8	74.7
Cumulative odds model	39.0	89.6	62.4	74.0
F (low)	21	71	54	71
Independence model	40.1	78.7	47.3	50.9
Local odds model	27.3	68.8	52.0	68.8
Cumulative odds model	27.7	68.6	52.5	68.2

can be estimated by fitting the baseline-category logit model and replacing x_i with $x_i(c-j)$ in the design matrix.

Goodness of Fit

Let π_{ij} denote the probability of a subject being classified in the cell in row i and column j of a two-way table. Then for a sample of size n, $n\pi_{ij}$ is the expected value (mean) of the distribution for the number of subjects classified in that cell. The values $\{n\pi_{ij}\}$ are called *expected frequencies.* Estimation methods used for fitting logit models also can generate estimates of expected frequencies that satisfy the model. For a particular model, estimated expected frequencies are values that provide the closest fit to the observed cell counts, subject to the constraints that they satisfy the model and match the observed data in certain marginal totals.

Table 4 gives estimated expected frequencies for fitting to Table 1 the independence model, the uniform association model for cumulative odds ratios, and the uniform association model for local odds ratios. The independence model fits very poorly in the corners of the table. This happens for ordinal data whenever there is a monotone trend. The models permitting dependence fit the corner cells considerably better. The estimates of the cell probabilities for the various models are the entries in Table 4 divided by the overall sample size, 1,660.

The goodness of fit of a model can be quantified by comparing the observed counts to the estimated expected frequencies, using a statistic of the Pearson form. Another statistic used for this purpose is the *likelihood ratio statistic,* defined by

$$G^2 = 2\ \Sigma\ \text{Observed}\ \log\left(\frac{\text{Observed}}{\text{Expected}}\right),$$

where the sum is taken over all the cells in the table. Like the Pearson statistic, this statistic is nonnegative and tends to take larger values when the fit is poorer, for a given sample size.

Degrees of freedom for goodness-of-fit statistics equal the number of sample logits minus the number of parameters in the model. For cumulative logits or adjacent-categories logits applied to a $r \times c$ table, there are $c - 1$ logits in each of r rows, a total of $r(c-1)$ logits. The uniform association model has $c - 1$ $\{\alpha_j\}$ parameters and the β parameter, so

$$df = r(c-1) - (c-1) - 1 = rc - r - c.$$

This is one fewer than the degrees of freedom $(r-1)(c-1)$ for the Pearson chi-square test of independence, because the uniform association model has one more parameter (β) than the model corresponding to independence. The row effects model has $c - 1$ $\{\alpha_j\}$ parameters and $r - 1$ $\{\beta_i\}$ parameters, so

$$df = r(c-1) - (c-1) - (r-1) = (r-1)(c-2).$$

Table 5 contains G^2 and df values for several models fitted to Table 1. The independence model has $G^2 = 47.4$, based on $df =$

Table 5
Goodness-of-Fit Statistics for Models Fitted to Table 1

		G^2	
Model	*df*	Cumulative logits	Adjacent category logits
Independence	15	47.4	47.4
Uniform association	14	10.9	9.9
Row effects	9	7.8	6.3

15. This is a poor fit; the P value is less than 0.001 for testing the hypothesis that the independence model holds in the population. (The Pearson chi-square statistic for testing independence is 46.0, also showing a poor fit.) By contrast, the cumulative logit model implying uniform cumulative odds ratios has $G^2 = 10.9$, based on $df = 14$, a very good fit. Similarly, the cumulative logit model with row effects gives a good fit, as do the corresponding models for adjacent-categories logits.

An advantage of the likelihood ratio statistic G^2 is that, unlike the Pearson form of statistic, it cannot increase as the model is made more complex. This feature makes it useful for comparing models. For instance, suppose one wishes to test whether a "complete" model M_c gives a significantly better fit than a "reduced" model M_r. Then $G^2(M_c) \leq G^2(M_r)$, and their difference is a chi-square statistic having degrees of freedom equal to the difference in the degrees of freedom for the two statistics. The larger the difference in G^2 values, the more evidence there is to reject the simpler model in favor of the more complex one. The test described here is analogous to the F test for comparing two regression models, in which G^2 is analogous to the residual sum of squares computed in regression analysis.

As an illustration, the difference in G^2 values between the independence model and the uniform association model for cumulative odds ratios is $47.4 - 10.9 = 36.5$, based on $df = 15 - 14 = 1$. This gives a P value less than 0.0001 for testing the hypothesis that the independence model is adequate against the alternative hypothesis that the uniform association model holds. Thus the uniform association model gives a better fit than the independence model. The uniform association model for local odds ratios also gives a better fit than the independence model. For either type of model, the test for comparing the model to the independence model usually gives very similar results to the statistic $z = \hat{\beta}/SE$ for testing $H_0: \beta = 0$ against $H_a: \beta \neq 0$.

The difference between G^2 values for the independence model and an ordinal model also gives a test for independence, one that takes into account the ordering of the categories. It is more efficient than the usual chi-square test of independence, because it is based on fewer degrees of freedom. For instance, by comparing G^2 values for the independence model and a uniform association model in Table 1, one obtains a test based on $df = 1$ rather than $df = 15$.

For either cumulative logits or adjacent-categories logits, one sees from Table 5 that the row effects model does not fit significantly better than the uniform association model. For instance, for cumulative logits, the difference $10.9 - 7.8 = 3.1$, based on $df = 14 - 9 = 5$, indicates that the more parsimonious model, uniform association, gives an adequate description of these data. This test is equivalently a way of checking whether an assigned set of scores is appropriate. That is, when one compares the uniform association model and the row effects model, one is testing whether the fit is as good when one replaces the parameters $\{\beta_i\}$ with βx_i for chosen scores $\{x_i\}$.

The G^2 statistic does not give a valid test of goodness of fit when any of the explanatory variables are continuous or when the cell counts tend to be small. However, in such cases, differences of G^2 values can still be useful for comparing complete and reduced models (see Agresti and Yang, 1986).

As with other statistical endeavors, there is danger in putting too much emphasis on statistical tests, whether of effects or of goodness of fit. Results are sensitive to sample size, and test statistics merely help indicate the level of parsimony that can be achieved. It is important to supplement these tests with estimation methods that describe the strength of associations and with residual analyses that detect parts of the data for which the general overall trend does not hold.

Model Selection

For a particular model type, such as cumulative logit, the G^2 statistic can be compared among several models to gauge the complexity of model needed, for instance, to determine which explanatory variables are needed, the forms of their effects, and whether interaction terms are needed. As in regression analysis, there is a trade-off between the desirability of a good fit (small G^2) and a parsimonious model. For Table 1, the uniform association model satisfies both criteria well, both for the cumulative logit and for the adjacent-categories logit type of model.

The cumulative logit and adjacent-categories logit both provided similar fits for Table 1. This similarity of results happens often in practice, so the choice between these model types must be based on other grounds. When one wants inferences to apply to an assumed underlying continuum for the response, the cumulative logit model is convenient. If that model fits well for an underlying continuous response variable, then it also fits well for any choice of categories for the response. The effect parameters will be about the same size regardless of the choice of response categories.

On the other hand, when one wants inferences to apply to a fixed set of response categories, the adjacent-categories logit model may be more convenient. With it, for each pair of response categories a and b, one can describe how the odds of response in category a, rather than category b, depend on explanatory variables. With the cumulative logit, description applies instead to the odds of response below y rather than above y for an arbitrary fixed point y. With the adjacent-categories logit, the values of effect parameters will depend strongly on the choice of categories for the response. The parameters tend to have smaller values when the number of response categories increases, because the local odds ratios then describe association over more restricted ranges.

Computational Issues

The ordinary least squares procedures used in fitting regression and ANOVA models cannot be used with models for categorical data. The main difficulty is that statistics such as sample logits do not have variance that is constant for all levels of the explanatory variables. There are two related estimation methods that can be used: maximum likelihood and weighted least squares. Weighted least squares is a generalization of ordinary least squares that gives relatively more weight to a sample logit as its variance decreases. Maximum likelihood estimates are the values for the parameters under which the observed data would have had the highest probability of occurrence; they are calculated through iterative use of a weighted least squares algorithm. The two methods give very similar results for large samples. For small samples, weighted least squares can be somewhat unreliable; for instance, estimates do not exist if there are zeros in some cells of the table. Also, weighted least squares cannot be used if any explanatory variables are continuous. Thus, maximum likelihood is the preferred method for most analyses, and this is the method used to obtain estimates reported in this article.

Though iterative methods are needed to solve the equations that determine maximum likelihood estimates, these methods are becoming increasingly available in statistical computer packages. Most packages use the Newton-Raphson iterative method, for which adequate convergence usually occurs within a few cycles for tables of only two or three dimensions. The estimated covariance matrix of the model parameter estimates is produced as a by-product of this method.

The statistical computer package SAS (SAS Institute, Inc., 1987) can be used to fit all models discussed in this article. The supplemental procedure LOGIST (Harrell, 1986) in SAS fits regular logistic regression models when the response variable has two categories, and it fits cumulative logit models when there are more than two response categories. This program uses maximum likelihood estimation, and it can be used with categorical or continuous explanatory variables. The newest version permits the testing of the assumption that the effect is the same for each cut point, and it permits the fitting of more general models in which this assumption is not made. The procedure CATMOD in SAS is a general procedure that enables the user to define a wide variety of models for a categorical response variable. Maximum likelihood is available as an option for the standard response function, which is the baseline-category logit. For other response functions, weighted least squares is used, in which case the explanatory variables must all be categorical.

Appendix A gives the code for using PROC LOGIST to fit cumulative logit models to Table 1, using maximum likelihood. For PROC LOGIST, $k + 1$ denotes the number of response categories, so $k = 3$ in this case. The model statement MODEL MENTAL = SES treats SES as a quantitative variable with the scores supplied with the input data, so in this case it gives the uniform association model for cumulative odds ratios. The variables A–E are dummy variables for the first five levels of SES. The second model statement fits these as explanatory variables and corresponds to the row effects models.

Appendix B gives the code for using PROC CATMOD in SAS to fit the logit model for uniform cumulative odds ratios using weighted least squares and the logit model for uniform local odds ratios using maximum likelihood. For either model, there are 18 sample ordinal logits, 3 for each row of Table 1. Each model has four parameters, the three cut point parameters $\{\alpha_1, \alpha_2, \alpha_3\}$ and the β effect parameter. Thus the design matrix has dimensions 18×4. The RESPONSE statement forms the 18 sample cumulative logits, 3 for each row. With Version 6 of SAS for the personal computer, this is done automatically with the statement RESPONSE CLOGITS;. When there is no RESPONSE statement before the MODEL statement, the model uses baseline-category logits. The maximum likelihood fit for the adjacent-categories logit model can be obtained by fitting the corresponding baseline-category logit model. The matrix entered in the second MODEL statement is the design matrix for the baseline-category logit model that makes it equivalent to the adjacent-categories logit model. The fourth element in each row of the design matrix has the value $i(4 - j)$, for $i = 1, \cdots, 6$ and $j = 1, 2, 3$; this is the coefficient of β for the model implying uniform local odds ratios.

The procedure LOGLINEAR in SPSS^X can be used to fit adjacent-category logit models but not cumulative logit models. The models are fitted by identifying them with corresponding log linear models (uniform association and row effects models). See Agresti (1984, Appendix D) and Norusis (1985) for details.

The computer package GLIM (Numerical Algorithms Group Inc., 1985) can also be used to fit these models. The adjacent-categories logit model is obtained by fitting the equivalent log linear model (see Agresti, 1984, Appendix D) and the cumulative logit model is fitted using a supplementary program (Hutchison, 1984).

Increased Power With Ordinal Models

When classifications are ordinal and one expects a positive or a negative trend, one may be able to summarize the association adequately by a single statistic, such as an estimated uniform cumulative odds ratio. The test of independence is then based on a chi-square statistic having only a single degree of freedom. If the model on which the statistic is based fits well, such a test is much more powerful than the usual Pearson test of independence; that is, it will be more likely to give a small P value when there truly is an association with a positive or a negative trend.

I now outline the reason for the advantage in power that a chi-square test with a single degree of freedom has. When the null hypothesis of independence is false, chi-square statistics have large-sample noncentral chi-square distributions. If two statistics capture the same basic information about the association, they share the same noncentrality. For fixed noncentrality, power increases as degrees of freedom decrease. Consequently, it is desirable to obtain the noncentrality using a statistic having as few degrees of freedom as possible. The chi-square statistic in the Pearson test of independence does not do this; it is designed to detect any type of deviation from the null hypothesis, so it uses all $(r - 1)(c - 1)$ degrees of freedom. The model-based statistics described in this article detect only certain types of deviations, but they are the ones of most importance for ordinal variables. Hence they are based on fewer degrees of freedom and usually have greater power than the Pearson test. Of course, they may be less powerful if the actual association is described poorly by the model on which the statistics are based (just as the t test for the regression slope will not do well if there is really a quadratic relationship rather than a linear one).

Literature Review

Nontechnical introductions to ordinary logit and log linear models were given by Upton (1978), Swafford (1980), Gilbert (1981), and Agresti and Finlay (1986, chap. 15). Slightly more advanced treatments were given by Fienberg (1980) and Agresti (1984, chaps. 1–4, 6). Haberman's (1978) text is considerably more technical but shows lots of applications. The book by Bishop et al. (1975) is a classic reference for log linear models, but it deals almost exclusively with models that treat all variables as nominal.

Discussions of the cumulative logit model include those by Williams and Grizzle (1972), Simon (1974), Bock (1975, pp. 544–546), McCullagh (1980), Agresti (1984, chap. 7), and Anderson (1984). Haberman (1974) and Goodman (1979) presented a class of log linear models, called association models, that utilize orderings of categories by assigning or estimating scores. Goodman (1979) characterized the models in terms of local odds ratios, and he showed how to present analogs of ANOVA tables for partitioning the association into component parts. Goodman (1983) showed how these models correspond to logit models for adjacent categories (see also Goodman, 1981).

An alternative analysis for an ordinal response involves assigning scores to response categories and fitting regression models for the mean response. Because a categorical response is not normally distributed with constant variance, this analysis uses weighted least squares rather than ordinary least squares. Grizzle, Starmer, and Koch (1969), Forthofer and Lehnen (1981, chap. 6), Semenya, Koch, Stokes, and Forthofer (1983), and Agresti (1984, section 8.3) described this approach. Mantel (1963) and Bhapkar (1968) described simple tests, similar to regression and ANOVA tests, that utilize assigned scores. Analogous tests for partial association were given by Landis, Heyman, and Koch (1978).

Yet another approach to the analysis of ordinal data is more nonparametric in flavor. An ordinal categorical version of the Kruskal-Wallis test can be used to test independence between an ordinal response and a nominal factor. An ordinal categorical version of Kendall's tau (such as Goodman and Kruskal's gamma or Kendall's tau-b) can be used as the basis of a single-degree-of-freedom test of independence between two ordinal variables. For instance, z = Measure/(SE of measure) is a suitable statistic for testing independence using these measures of association. See Goodman and Kruskal (1979), Lehmann (1975, p. 305), and Agresti (1984, chap. 10) for details of these nonparametric tests. A weakness of the nonparametric approach is that it is less suitable for describing multidimensional tables than is an approach involving model building. Many of these alternative analyses can be implemented in SAS with PROC CATMOD or PROC FREQ, or in BMDP-4F (Dixon 1981).

Up until now, models as discussed in this article have been applied in the social sciences primarily in methodological papers rather than in routine data analysis. This should change as the models become more widely known to statisticians and researchers. Most applications have involved the use of association models, either for scaling or for describing occupational mobility (see, for instance, Clogg, 1982, 1984); Smith & Garnier, 1987; and Sobel, Hout, & Duncan, 1985).

In summary, there is a growing body of new methods available for analyzing ordered categorical response data. I hope this presentation will encourage researchers to use such methods in their own analyses of ordinal data.

References

Agresti, A. (1984). *Analysis of ordinal categorical data.* New York: Wiley.

Agresti, A., & Finlay, B. (1986). *Statistical methods for the social sciences* (2nd ed.). San Francisco: Dellen (Macmillan).

Agresti, A., & Yang, M. (1986). An empirical investigation of some effects of sparseness in contingency tables. *Computational Statistics & Data Analysis, 5,* 9–21.

Anderson, J. A. (1984). Regression and ordered categorical variables (with discussion). *Journal of the Royal Statistical Society, Series B, 46,* 1–30.

Bhapkar, V. P. (1968). On the analysis of contingency tables with a quantitative response. *Biometrics, 24,* 329–338.

Bishop, Y. V. V., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis.* Cambridge, MA: MIT Press.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research.* New York: McGraw-Hill.

Clogg, C. C. (1982). Using association models in sociological research: Some examples. *American Journal of Sociology, 88,* 114–134.

Clogg, C. C. (1984). Some statistical models for analyzing why surveys disagree. In C. F. Turner & E. Martin (Eds.) *Surveying subjective phenomena* (Vol. 1) New York: Russell Sage Foundation.

Dixon, W. J. (Ed.). (1981). *BMDP statistical software 1981.* Berkeley: University of California Press.

Fienberg, S. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge, MA: MIT Press.

Forthofer, R. N., & Lehnen, R. G. (1981). *Public program analysis: A new categorical data approach.* Belmont, CA: Lifetime Learning.

Gilbert, G. N. (1981). *Modelling society.* London: Allen & Unwin.

Goodman, L. A. (1979). Simple models for the analysis of association in cross-classification tables having ordered categories. *Journal of the American Statistical Association, 74,* 537–552.

Goodman, L. A. (1981). Three elementary views of loglinear models for the analysis of cross-classifications having ordered categories. In *Sociological methodology 1981* (pp. 193–239). San Francisco: Jossey-Bass.

Goodman, L. A. (1983). The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds. *Biometrics, 39,* 149–160.

Goodman, L. A., & Kruskal, W. H. (1979). *Measures of association for cross classifications.* New York: Springer-Verlag. (Originally published as a series of articles in the *Journal of the American Statistical Association,* 1954, 1959, 1963, 1972)

Grizzle, J. E., Starmer, C. F., & Koch, G. G. (1969). Analysis of categorical data by linear models. *Biometrics, 25,* 489–504.

Haberman, S. J. (1974). Loglinear models for frequency tables with ordered classifications. *Biometrics, 30,* 589–600.

Haberman, S. J. (1978). *Analysis of qualitative data: Vol. 1. Introductory topics.* New York: Academic Press.

Harrell, F. E. (1986). The LOGIST procedure. In *SUGI supplemental library user's guide* (Version 5 ed., pp. 269–293). Cary, NC: SAS Institute.

Hutchison, D. (1984). Ordinal variable regression using the McCullagh (proportional odds) model. *GLIM Newsletter, 9,* 9–17.

Landis, J. R., Heyman, E. R., & Koch, G. G. (1978). Average partial association in three-way contingency tables: A review and discussion of alternative tests. *International Statistical Review, 46,* 237–254.

Lehmann, E. L. (1975). *Nonparametrics: Statistical methods based on ranks.* San Francisco: Holden-Day.

Mantel, N. (1963). Chi-squared tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association, 58,* 690–700.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B, 42,* 109–142.

Mosteller, F. (1968). Association and estimation in contingency tables. *Journal of the American Statistical Association, 63,* 1–28.

Norusis, M. J. (1985). *SPSS^X advanced statistics guide.* New York: McGraw-Hill.

Numerical Algorithms Group Inc. (1985). *The GLIM release 3.77 manual.* Downer's Grove, IL: Author.

SAS Institute, Inc. (1987). *SAS/STAT guide for personal computers* (Version 6 ed.). Cary, NC: Author.

Semenya, K., Koch, G. G., Stokes, M. E., & Forthofer, R. N. (1983). Linear models methods for some rank function analyses of ordinal categorical data. *Communications in Statistics Series A, 12,* 1277–1298.

Simon, G. A. (1974). Alternative analyses for the singly-ordered contingency table. *Journal of the American Statistical Association, 69,* 971–976.

Smith, H., & Garnier, M. (1987). Scaling via models for the analysis of association: Social background and educational careers in France. In *Sociological methodology 1987* (pp. 205–246). San Francisco: Jossey-Bass.

Sobel, M. E., Hout, M., & Duncan, O. D. (1985). Exchange, structure, and symmetry in occupational mobility. *American Journal of Sociology, 91,* 359–372.

Srole, L, Langner, T. S., Michael, S. T., Kirkpatrick, P., Opler, M. K., & Rennie, T. A. C. (1978). *Mental health in the metropolis: The midtown Manhattan study* (rev. ed.). New York: NYU Press.

Swafford, M. (1980). Three parametric techniques for contingency table analysis: A nontechnical commentary. *American Sociological Review, 45,* 664–690.

Upton, G. H. G. (1978). *The analysis of cross-tabulated data.* New York: Wiley.

Williams, O. D., & Grizzle, J. E. (1972). Analysis of contingency tables having ordered response categories. *Journal of the American Statistical Association, 67,* 55–63.

Appendix A

SAS (PROC LOGIST) Used to Fit Cumulative Logit Models to Table 1

```
DATA MENTAL;
INPUT SES ROW $ N0 N1 N2 N3 @@;
A=ROW='A';
B=ROW='B';
C=ROW='C';
D=ROW='D';
E=ROW='E';
LABEL MENTAL = 'O=WELL 1=MILD 2=MODERATE
  3=IMPAIRED';
MENTAL=0; DO I = 1 TO N0; OUTPUT; END;
MENTAL=1; DO I = 1 TO N1; OUTPUT; END;
MENTAL=2; DO I = 1 TO N2; OUTPUT; END;
MENTAL=3' DO I = 1 TO N3; OUTPUT; END;
CARDS;
1  A  64   94  58  46
2  B  57   94  54  40
3  C  57  105  65  60
4  D  72  141  77  94
5  E  36   97  54  78
6  F  21   71  54  71
PROC LOGIST  K=3;
MODEL MENTAL = SES;
PROC LOGIST  K=3;
MODEL MENTAL = A B C D E;
```

Appendix B

SAS (PROC CATMOD) Used to Fit Cumulative Logit and Adjacent-Categories Logit Model to Table 1

```
DATA MENTAL;
INPUT SES MENTAL COUNT @@;
CARDS;
1  1  64    1  2   94    1  3  58    1  4  46
2  1  57    2  2   94    2  3  54    2  4  40
3  1  57    3  2  105    3  3  65    3  4  60
4  1  72    4  2  141    4  3  77    4  4  94
5  1  36    5  2   97    5  3  54    5  4  78
6  1  21    6  2   71    6  3  54    6  4  71
;
PROC CATMOD   ORDER=DATA; WEIGHT COUNT;
DIRECT SES;
RESPONSE  1  -1  0  0  0  0,  0  0  1  -1  0  0,  0  0  0  0  1  -1  LOG
          0  0  0  1,  1  1  1  0,  0  0  1  1,  1  1  0  0,  0  1  1  1,  1
          0  0  0;
MODEL MENTAL = _RESPONSE_SES;

PROC CATMOD   ORDER=DATA; WEIGHT COUNT;
POPULATION SES;
MODEL MENTAL = (1  0  0   -3,  0  1  0   -2,  0  0  1  -1,
                1  0  0   -6,  0  1  0   -4,  0  0  1  -2,
                1  0  0   -9,  0  1  0   -6,  0  0  1  -3,
                1  0  0  -12,  0  1  0   -8,  0  0  1  -4,
                1  0  0  -15,  0  1  0  -10,  0  0  1  -5,
                1  0  0  -18,  0  1  0  -12,  0  0  1  -6)
       (1 2 3 = "CUTPOINTS", 4 = "SES")/ML NOGLS PRED=FREQ;
```

Received January 13, 1988
Revision received May 20, 1988
Accepted May 20, 1988