Scientific Research and Essay Vol.4 (9), pp. 853-860, September 2009
Available online at http://www.academicjournals.org/sre
ISSN 1992-2248 © 2009 Academic Journals

Review

History of ordinal variables before 1980

Justin R Chimka[1]* and Harvey Wolfe[2]

[1]University of Arkansas, 4207 Bell Engineering Center, Fayetteville, AR 72701, USA.
[2]University of Pittsburgh, Pennsylvania, United States.

Accepted 9 June, 2009

It is common to see statistics that apply only to interval or ratio scales applied to ordinal data. The origin of this is very old and unique in that methods based on less than adequate research conclusions continue to be used. This paper explores the origin of scales purported to measure attitudes and the various arguments that have ensued over the years for and against the use of statistical measures. The advent of ordinal scale techniques actually began a quarter of a century after the first advocacy of using interval statistics for evaluating ordinal data. However ordinal statistical methods' complexity has resulted in most researchers and teachers reverting to the simplicity of using interval statistics and ignoring their underlying assumptions. We present a historical contribution to the literature on statistical models for ordinal data. Books on the subject maintain a relatively strong emphasis on methodological development (Agresti, 1984; Clogg and Shihadeh, 1994; Johnson and Albert, 1999), and the commentary by Velleman and Wilkinson (1993) deemphasizes the terminology that finally bore theoretically appropriate regression models for ordinal data (McCullagh, 1980). These proportional odds models are still practically state-of-the-art which is the reason we emphasize the history of ordinal variables before 1980.

Key words: Measurement scales, ordinal variables, proportional odds.

INTRODUCTION

A method for assigning values to ordered categories was attributed to Edward L. Thorndike as early as 1913. For the next two decades authors debated the distribution of attitudes, an issue important to the conversion of ordinal data to numerical values. Then in 1932, Rensis Likert emerged with his now famous scale and the simple method of assigning the values 1, 2, 3, 4 and 5 to its alternatives. The next decade, however, would sober this movement culminating with SS Stevens' (1946) theory of measurement scales.

Between the years 1932 and 1939, nineteen representatives of the British Association for the Advancement of Science Sections a (mathematical and physical sciences) and j (psychology) considered the possibility of estimating sensory events quantitatively. The committee's goal was to answer the question, "Is it possible to measure human sensation?" Disagreement about the definition of about the definition of measurement ensued, and after an interim report in 1938, the committee asked for a year continuance. Still, in their final report in 1940, members' views were wide ranging. In response, Stevens sought to clarify the meaning of measurement, writing, "Perhaps agreement can better be achieved if we recognize that measurement exists in a variety of forms and that scales of measurement fall into certain definite classes."

Stevens defined four types of scales (nominal, ordinal, interval and ratio) according to three parameters: Basic empirical operations, mathematical group Structure and permissible statistics (Table 1). The column basic empirical operations, listing the operations necessary to create each scale, is cumulative. For example, one can construct an ordinal scale provided there are operations for determining greater or less and operations for determining equality. One must add to these operations an additional operation for determining equality of intervals or differences in order to achieve an interval scale. Finally, to construct a ratio scale, one must add the ability to determine equality of ratios. Stevens wrote, "The criterion for appropriateness of a statistic is invariance

*Corresponding author. E-mail: jchimka@uark.edu. Tel.: (479) 575-7392. Fax: (479) 575-8431.

Table 1. Types of scales (Stevens, 1946).

Scale	Basic empirical operations	Mathematical group structure	Permissible statistics
Nominal	Determination of equality	Permutation group x' = f(x) f(x) means any one-to-one substitution	Number of cases Mode Contingency correlation
Ordinal	Determination of greater or less	Isotonic group x' = f(x) f(x) means any monotonic increasing function	Median Percentiles
Interval	Determination of equality of intervals or differences	General linear group x' = ax + b	Mean Standard deviation Rank-order correlation Product-moment correlation
Ratio	Determination of equality of ratios	Similarity group x' = ax	Coefficient of variation

under the transformations in column 3." For example, the median of a distribution maintains under all transformations of the isotonic group, but the mean maintains only under transformations of the linear group. He also wrote, "The last column presents examples of the type of statistical operations appropriate to each scale. (The rank order correlation coefficient is usually deemed appropriate to an ordinal scale, but actually this statistic assumes equal intervals between successive ranks and therefore calls for an interval scale.)" This last column is also cumulative; all listed statistics are legitimate for ratio scale data. Clyde H. Coombs (Festinger and Katz, 1953; Coombs, 1964) and Warren (1958) published additional, interesting discussions about scales and measurement of underlying variables in the decades after Stevens' article in Science.

Measurement of attitudes

A common context for discussion about scales of measurement was borne out of "A technique for the measurement of attitudes" by Likert (1932) that is known as the Likert scale. Likert's study presented and attempted to evaluate a new method of measuring attitudes. The scale first appeared in a questionnaire to aid a project conceived by Likerts and Gardner Murphy in 1929. The project was to study the attitude areas of economic conflict, international relations and race relations and, to a lesser degree, religion and political conflict among students at typical American Universities. The questionnaires, or attitudes tests, were given to undergraduates in nine colleges and universities in states including Illinois, Connecticut, Ohio, Pennsylvania and Virginia. Only Columbia College (Chicago) was identified. Of the over 2000 participants, 650 cases were randomly sampled from seven of the nine groups for analysis. The questionnaire, called a “survey of opinions” was administered to seven groups in fall 1929 and to the two remaining in 1931. Participants were given both an initial survey and a retest 30 days later. The second test included some items from the first as well as other new items. The first test required approximately 40 min and the retest slightly more.

"In order to compare one type of statement with another such as the 'multiple choice' with the 'strongly approve,' it was necessary to devise some technique whereby they might be made comparable. In attempting to work out such a technique, it was noticed that a great number of the five-point statements yielded a distribution resembling a normal distribution (Likert, 1932)". Eight typical distributions were presented from a sample of 100 male cases from one University Likert wrote of them.

On the basis of this experimental evidence and upon the results of Rice (1928) and Folsom (1931), it seems justifiable for experimental purposes to assume that attitudes are distributed fairly normally and to use this assumption as the basis for combining the different statements. The possible dangers inherent in this assumption are fully realized. This assumption is made simply as part of an experimental approach to attitude measurement (Likert, 1932).

Likert's evidence of normality appears questionable. Of the empirical distributions in Tables 2 and 3 (Likert, 1932), not one distribution has as its mode the central, or third, ordered category. Furthermore, the distributions associated with statement numbers 6 and 9 appear to be bimodal (have two local, non contiguous maxima). In fact, only the data for statements 3 and 6 passed Shapiro-Wilk W tests for normal data (Shapiro and Wilk, 1965) with 95% confidence (administered by the authors of this review and displayed in Table 4).

Prior to Likert’s analysis of his work, Stuart A. Rice and Joseph K. Folsom both considered the distribution of attitudes and reviewed findings of Floyd H. Allport and DA Hartman (1925) published in the American Political

Table 2. Percentage of individuals checking alternatives of multiple choice statements.

Statement	A	B	C	D	E
3	11	43	27	15	4
7	1	1	3	8	87
8	29	42	26	3	0

Table 3. Percentage of individuals checking alternatives of strongly approve statements.

Statement	Strongly approve	Approve	Undecided	Disapprove	Strongly disapprove
5	32	52	10	5	1
6	10	27	17	35	11
9	3	17	14	44	22
10	24	49	17	7	3
16	13	43	21	13	10

Table 4. Shapiro-Wilk W tests for normal data of Likert statements.

Comparison method	Obs	Pr > *z*
3	100	0.10765
7	100	0.00000
8	100	0.00838
5	100	0.00006
6	100	0.33959
9	100	0.00489
10	100	0.00041
16	100	0.02892

Science Review. Allport and Hartman wanted to identify the psychological characteristics of individuals that lead them to adopt extreme social and political views. Their procedure for answering the question was to: first, measure distributions of opinions within a group and second, psychoanalyze individuals representing various regions of the distributions. A set of graded standardized scales was constructed for each of seven timely issues (e.g., the League of Nations, the qualifications of President Coolidge, prohibition) and administered to a freshman class at Syracuse University's College of Liberal Arts.

Of Allport and Hartman's work, Rice (1928) wrote; "The experimental procedure tended to substantiate the belief that attitudes can be thought of as distributed along scales, but it tended to discredit the hypothesis that they are distributed normally." Still, Rice did not abandon the hypothesis entirely, believing attitudes to be distributed normally apart from some distorting situation, under ideal conditions that cannot be exactly, actually attained. He suggested research to determine under what circumstances conditions are approximately ideal and distributions practically normal. Finally, he submitted; "If the distribution is not normal, it is because the factors determining individual attitudes are not numerous equipotent (sic) and independent."

In chapter XI of Folsom's (1931) book Social Psychology, there is a section with the title "No General Law of Distribution of Attitude-Variables." Folsom begins, "In the writer's view it is useless to seek any general law of distribution of social attitudes." Folsom claimed that attitudes are determined by social interaction and not by laws of probability. It was suggested that a factor describing an individual could be a normal variable only if the determination of that factor was independent of its determination in other individuals. The author claimed that individual attitudes are dependent on the attitudes of others. Folsom concluded, "This, of course, is a theoretical analysis. But it helps to show how absurd it is to expect even theoretically a normal distribution of any particular attitude."

Likert's analysis

Returning to Likert's scale analysis (1932), he converted the percentage of students that indicated a given position on a statement into a sigma value. His method, following that of Thorndike (1913), converts measurements of relative position into units of σ, deviation from the average, when the underlying distribution is approximately known. Assuming the probability surface of range $+3\sigma$ to -3σ is divisible into 100 equal areas that represent 100 successive percents, the average distance from the mean in units σ can be calculated for each percent. So, to obtain the sigma value for the highest 10 percent, for example, add the figures for the first 10%, and divide by 10. Percents in each ordered category for a group of responses are converted in this way, effectively assigning values to the categories.

Likert (1932) compared this sigma technique to the simpler method of assigning values 1, 2, 3, 4 and 5 to each of five alternatives in order from negative to positive.

"The scores obtained by this method and the sigma method correlated almost perfectly. The same results were obtained when the values 1, 3, 4, 5 and 7 were assigned to the different positions corresponding, respectively, to the values 1, 2, 3, 4 and 5. These results seem to justify the use of the simpler methods of scoring." This might be true, and the simpler methods valid, if the sigma method were able to expose the underlying variables of ordinal scales.

These early methods, results, and conclusions of Likert certainly foreshadowed those controversial ones of Sanford Labovitz (1970) that would come years later. Unfortunately, Likert's (1932) technique for the measurement of attitudes also predated Stevens' (1946) theory of measurement scales. Had Likert (1932) the fortune of reading Stevens' (1946) Science publication before developing his measurement techniques, they might have been different. It is interesting to consider these two papers of Likert (1932) and Stevens (1946) published fourteen years apart and largely contradictory of each other, but both very influential.

Likert (1932) compared his new method of measuring attitudes to other widely used methods of the day: especially those developed by Thurstone (1928) in the American Journal of Psychology article "Attitudes Can Be Measured." Likert (1932) wrote, "This method not only retains most of the advantages present in methods now used, such as yielding scores, the units of which are equal throughout the entire range, but it has additional advantages." Likert's method 1) eliminated judges or raters and, therefore, errors arising from them; 2) was a simpler one for constructing attitude scales; and 3) "yields the same reliability with fewer items." After assigning numerical values to the different alternatives, a score for each individual may be determined by finding the sum or the average of the numerical values corresponding to appropriate responses.

Finally, Likert's method of constructing an attitude scale was presented in two parts:

1) The selection of statements.
2) Constructing the scale.

Regarding the selection of statements, Likert wrote, "Each statement should be of such a nature that persons, with different points of view so far as the particular attitude is concerned, will respond to it differently." Regarding scale construction, he continued, "For purposes of tabulation and scoring, a numerical value must be assigned to each of the possible alternatives. If five alternatives have been used, it is necessary to assign values from one to five with the three assigned to the undecided position on each statement. The ONE end is assigned to one extreme of the attitude continuum and the FIVE to the other."

Thus, a lasting precedent was set for the measurement of attitudes and the assignment of numerical values to them. Likert's landmark paper substituting relevant terms for his specific comments about Thurstone reads, "It is feared that some will mistakenly interpret this [text] as an 'attack' on [Likert's] methods. I therefore wish to emphasize in the strongest terms that I am simply endeavoring to call attention to certain problems of the method, and that I am very far from convinced that the present data close the question."

ORDINAL VARIABLES AS INTERVAL SCALES

A formal debate over "The Assignment of Numbers to Rank Order Categories" began with Labovitz (1970) in the American Sociological Review. Labovitz claimed to show that, by assigning numbers to ranks, ordinal variables conform well to interval scales. One can often use "more powerful, more sensitive, better developed, and more clearly interpretable statistics" such as correlation, regression and analysis of variance with interval and ratio variables. Labovitz believed these advantages offset any error resulting from treating ordinal data as interval and ratio.

Labovitz determined the error resulting from treating ordinal data as interval by relating a measure of occupational prestige to suicide rate. Thirty-six occupations (e.g., dentists, chemists, architects and engineers) were ordered on a prestige rating scale based on a questionnaire employing a method of paired comparisons. The prestige of each occupation was assigned numerical value by Labovitz using twenty different scoring systems, and correlated with suicide rate. The first system resulted from the original paired comparisons method; the 36 occupations were assigned values ranging from seven to 97 with some ties. The second system assigned monotone, equidistant numbers to occupational levels, retaining the order determined by the paired comparisons. The remaining eighteen scoring systems assigned to occupational categories random increasing numbers between one and 10,000. Labovitz found that these systems of scoring yielded highly (Pearson) correlated variables. In addition, correlation coefficients of occupational prestige and suicide rate were similar given the different systems of scoring. Labovitz concluded that "different systems yield interchangeable variables," but that an equidistant system of scoring ordinal categories was preferred. He also reported that confidence in assigning interval scores to ordinal data increases with the number of categories and that one should not collapse them. Although at the time of Labovitz's study multivariate analysis was underdeveloped for ordinal data, his publication ended on a cautionary note.

The researcher should know and report the actual scales of his data, and any interval statistic should be interpreted with care. Further exploration and tests are necessary for added confidence in treating ordinal data as if they are interval. The more conservative procedure, of course, is to treat ordinal data as strictly ordinal, and thereby avoid the possibility of attributing a property to a given scale [that] it does not possess.

The American Sociological Review published Labovitz's study along with a comment by Lawrence (1970) who found Labovitz's correlation coefficients "interesting although not surprising." He believed the paper to show not that one should apply numbers to ranks for the applications of multivariate techniques but simply "that the Pearson *r* is fairly stable with respect to non-linear monotone transformations." Mayer further criticized Labovitz for not mentioning that most interval level procedures require assumptions about the distributions of data. Even after assigning numbers to rank ordered categories and accepting an interval scale for ordinal data, departures from distribution assumptions will further weaken interval level procedures for ordinal data.

Louis (1971) echoed Mayer's sentiments regarding Labovitz's (1970) ascription of interval scales to ordinal variables. Specifically, Vargo (1971) addressed the issue of correlation measurement, claiming that Labovitz (1970) had demonstrated only its response to monotone transformations of data, the significance being "numerical, not methodological (Vargo, 1971). "Sybil and Donald (1971) offered further quantitative evidence of the insensitivity of the Pearson *r* to scale transformations. They wrote, "Labovitz's use of this statistic is a poor justification for the general claim that ordinal scales can be treated as if they conform to interval scales." Mayer (1971) then returned to present a theoretical argument against treating ordinal data as interval. He added some monotone transformations not considered by Labovitz (1970) that may lead one to underestimate the true relationship between an interval variable and an ordinal variable.

Labovitz (1971) wrote in defense of assigning numbers to ranks that it "is an aid to data analysis, and the procedure is not risky if care is taken to avoid extreme exponential distributions." Unfortunately, no clues were given for identifying the distributions making Labovitz's procedure risky. He defended the legitimacy of assigning numbers to ordinal rankings and answered Mayer's (1970) skepticism regarding the distributions that arise from assigning numbers to ranks. He did this with two strategies:

1) Use robust statistics so that violating distribution assumptions barely affects interpretation.
2) Assign numbers to meet distribution assumptions.

In reference to Schweitzer and Schweitzer (1971), Labovitz suggested thinking of the Pearson *r* as robust rather than insensitive. Schweitzer and Schweitzer argued that the Pearson *r* is not an appropriate test of whether ordinal variables conform to interval scales because of its insensitivity to scale transformations and number of ordered categories. Labovitz failed to see these as flaws in his study. He wrote, "It is precisely these characteristics, however, that warrants [the use of Pearson's *r*] for the problem, and permits accurate interpretations of *r* for transformed scales."

ORDINAL SCALE TECHNIQUES

Statistics appropriate in theory for ordinal data began to take shape soon after Stevens' (1964) theory of measurement scales was published. Biometrics published an article in 1957 that began to develop what would come to be known as an “ordered logit model” (Aitchison, 1957). After discussion inspired by Labovitz (1970) and a short hiatus, publication resumed in the area of ordinal scale techniques in 1975. Interest resumed through the early 80's, giving way to relatively well developed models and discussions about the practical consequences of ordered logit modeling compared to those of interval and ratio level techniques.

Classification of silverfish

Aitchison and Silvey (1957) of the University of Glasgow may have been the first to describe an ordinal procedure for modeling an ordinal response: "The Generalization of Probit Analysis to the Case of Multiple Responses." The generalization arose from the following entomological problem: The *Petrobius* leech, a primitive relative of the silverfish, passes through different stages in its lifetime. Since it was difficult to keep the leech alive in a laboratory, entomologists observed leeches on different dates in the field and classified the leeches according to stage. The problem was to estimate the mean time spent in each stage, given the observational data. Aitchison and Silvey reduced the problem to one of estimating (by maximum likelihood) the probability of observing the insect on each date, in each stage; date is an independent variable, or stimulus, and the probabilities of occurring in each ordered categorical stage comprise the dependent variable or response.

The authors effectively solved the leech problem by a generalization of Finney's (1971) probit analysis. Aitchison and Silvey cited Finney's original1947 edition of Probit Analysis. Methods of probit analysis are applicable in the following case: Random samples of subjects are given various doses of some stimulus to which a subject may or may not respond. Therefore, any dose results in two groups of subjects to which the dose is applied: those responding and those not responding. A dose is said to be effective if it produces a response, and the minimum effective dose for a subject is called that subject's tolerance. One of the main objectives of probit analysis is to estimate mean tolerance. Situations arise, of course, where subjects divide into three or more classes, rather than into a simple dichotomy, as in an experiment where subjects receive a stimulus and are placed in one of $s + 1$ classes based on the effect. For example, Aitchison and Silvey (1957) cited a paper by Tattersfield et al. (1925) that classified insects subjected to a poison as unaffected, slightly affected, moribund or dead. Aitchison and Silvey (1957) created generalized probit analysis for situations such as these.

According to Aitchison and Silvey, two crucial conditions must be satisfied in order to apply the generalization of probit analysis. They are:

1) That classes must be ordered, mutually exclusive, and exhaustive.
2) If dose *x* places a subject in the *i*th class, then a dose greater than *x* is required to place this subject in the *j*th class whenever *j* is greater than *i*.

In addition, one must assume that mean time spent in each stage *s* is a non-negative, random variable, and that its distribution is approximately normal. This is possible only when the standard deviation of the variable "is small relative to its mean, say, $3\sigma_i <= \lambda_i$." Finally, the authors remarked that, "if $s = 1$ then the present analysis becomes an ordinary probit analysis and it is in this sense that we have generalized probit analysis."

Application to voting behavior

The Aitchison and Silvey model stood alone for seventeen years until McKelvey and Zavoina (1975) further extended probit analysis for more than one independent variable. Between the years of the two publications, other models were developed for treating the Aitchison and Silvey (1957) problem (Theil, 1969, 1970; Aitchison and Bennet, 1970; Grizzle, 1971), but they generally assumed nominal, rather than ordinal, dependent variables. Around the time of McKelvey and Zavoina (1975), log-linear models were also developed, requiring the assignment of scores to levels of an ordinal scale response (Haberman, 1974; Simon, 1974; Goodman, 1979; Fienberg, 1980; Clogg, 1982; Agresti, 1984).

McKelvey and Zavoina (1975) illustrated theoretical problems related to using linear regression with ordinal data. The authors showed that the assumptions of the linear model are not met when *Y* is ordinal and values are assigned to each of the ordered categories. For example, McKelvey and Zavoina presented data of a trichotomous dependent variable *Y*, a function of one continuous independent variable X_1. A plot of the data indicated a definite relationship between the two variables; however, it seemed that no linear model could have possibly produced the data. The least squares line had positive errors for small X_1 and negative errors for large X_1. Given these difficulties, the authors sought an alternative model.

McKelvey and Zavoina distinguished between a dependent variable of theoretical interest, *Y*, and an observed dependent variable, *Z*. The authors assumed that the theoretical dependent variable was interval level and would, given one's ability to measure it, satisfy a linear model, with error distributed normally and mean equal to zero. However, due to inadequate measurement, only an ordinal version *Z* of *Y* is observable, for which the linear model is not appropriate, because the error term of *Z* will have non-constant variance. McKelvey and Zavoina assumed that the variable *Z* is ordinal with a number of categories. They developed a model for ordinal level dependent variables that estimated the probability of occurrence in each category and so predicted the ordinal outcome. Estimators for the model parameters are obtained using maximum likelihood methods.

McKelvey and Zavoina compared regression analysis and their ordinal probit analysis for modeling voting behavior in the 89th Congress on the Medicare bill of 1965. Prior to the passage of the bill by the House of Representatives, Representative John Byrnes (R-Wisconsin) moved to recommit the bill to the Ways and Means Committee with instructions to weaken the bill by making Medicare a voluntary program financed by participants. McKelvey and Zavoina built an ordinal scale (with ordered categories 0, 1, 2) out of voting behavior both for recommittal (CQ34) and passage (CQ35) of the Medicare bill: 0) Voting for recommittal and against passage was defined as a vote against 1) Voting for recommittal and for passage was defined as a vote weakly for 2) Voting against recommittal and for passage was defined as a vote strongly for. Independent variables included Party (0 if Democrat, 1 otherwise), Region (0 if Non-South, 1 otherwise), Employment (Percent civilian unemployed in district) and Population (Population density in thousands per square mile).

Four different models were considered. The authors hypothesized that the probability of voting a particular way was a linear function of various combinations of the independent variables. The four different models were estimated with both regression analysis and probit analysis. The authors observed what they considered a general tendency of the regression analysis to return a lower value of R^2 than the probit analysis. This tendency was thought sensible. It was obvious to them that an excellent fit to the probit model would translate, due to the categorization of the dependent variables, into a poor fit to the regression model. It is not surprising that a model designed especially for an ordinal dependent variable would fit ordinal data better than would a model designed for interval or ratio data.

McKelvey and Zavoina argued, "Regression analysis is an inappropriate technique to apply to ordinal level dependent variables [due to] the failure of the regression model to describe the observed data." The authors assumed "that the true relation is described by the linear model, [but that there is] inherent loss of information when the continuous dependent variable is measured by techniques [that put together in a single group different parts] of the scale. Consequently, there is a correlation between error and regressor when the regression is applied to the observed data [introducing a bias] into the estimate of βp̃ıı9." McKelvey and Zavoina believed that this bias might, in some cases, cause regression analysis to "severely underestimate the relative impact of certain variables." More simply, the fact that regression is an interval or ratio level technique indicates that it is not well suited for ordinal dependent variables. The authors may have found evidence of this in a high probability of Type II error.

Contribution of Peter McCullagh

Five years after McKelvey and Zavoina's paper, the Royal Statistical Society published Peter McCullagh's (1980) (University of Chicago), "Regression Models for Ordinal Data." McCullagh's paper cited Aitchison and Silvey (1957) [but not McKelvey and Zavoina (1975)], further described their model and changed it slightly. McCullagh's (1980) model is also known as an ordered logit model. The difference between logit and probit specifications is in their assumptions regarding the distribution of error. Whereas previous authors assumed normally distributed errors, McCullagh (1980) popularized ordinal techniques by suggesting the simpler logistic specification. He called his model the "proportional odds model," a generalization of logistic regression for cumulative probabilities, which requires the assumption that odds ratios are independent of ordered category.

Clogg and Edward (1994) chapter "Logit-Type Regression Models for Ordinal Dependent Variables" serves well as an introduction to the topic. The authors considered that there are many possible ways that response functions for such variables can be formed, and that each way implies a different regression model.

Clogg and Shihadeh (1994) also found in an interesting comparison between the cumulative logit model and linear regression that "the two give funda-mentally different inferences." The example involves the 1977 and 1989 General Social Survey question to measure attitudes about mothers who work outside the home. Responses were according to a Likert scale of five choices. Predictors examined include gender, year (of survey), education (completed years of schooling) and age. To illustrate the difference between model conclusions, Clogg and Shihadeh compared the sizes of the gender and education effects: The cumulative logit model gave a ratio of 8.0; "the same ratio of the least squares estimates for the integer-scored version of *Y* is 2.2, which is radically different."

That the cumulative logit model and linear regression may give different inferences has inspired more questions as to the accuracy of inferences from ordinal data.

The reader is left to discover a more recent history and perhaps join efforts to educate about and research in statistics of ordinal variables. Articles important to a more recent history of ordinal variables address the practical use of theoretically appropriate models (Anderson, 1984; Brant, 1990; Greenland, 1985; Peterson and Harrell, 1990). Hopefully we have motivated the reader to approach these more technical treatments and analysis of ordinal data with greater sensitivity to its scale.

REFERENCES

Agresti A (1984). Analysis of Ordinal Categorical Data, New York: John Wiley & Sons.

Aitchison J, Bennet JA (1970), Polychotomous Quantal Response by Maximum Indicant, Biometrika, 57(2): 253-262.

Aitchison J, Silvey SD (1957). The Generalization of Probit Analysis to the Case of Multiple Responses, Biometrika, 44: 131-140.

Allport FH, Hartman DA (1925). The Measurement and Motivation of Atypical Opinion in a Certain Group, Am. Political Sci. Rev. 19(4): 735-760.

Anderson JA (1984). Regression and Ordered Categorical Variables, J. Royal Stat. Soc., 46, Ser. B, (1): 1-30.

Brant R (1990). Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression, Biometrics, 46: 1171-1178.

Clogg CC (1982). Some Models for the Analysis of Association in Multi-way Cross Classifications Having Ordered Categories, J. Am. Stat. Assoc., 77(380): 803-815.

Clogg CC, Shihadeh ES (1994). Statistical Models for Ordinal Variables, Thousand Oaks, California: Sage Publications.

Coombs CH (1964). A Theory of Data, New York: John Wiley & Sons, Inc.

Festinger L, Katz D (1953). Research Methods in the Behavioral Sciences, "Theory and Methods of Social Measurement, by Clyde H. Coombs, New York: Holt, Rinehart and Winston, Chapter 11.

Fienberg SE (1980). The Analysis of Cross-Classified Categorical Data (2nd ed.), Cambridge: MIT Press.

Finney DJ (1971). Probit Analysis (3rd ed.), Cambridge: The University Press.

Folsom JK (1931). Social Psychology, New York: Harper & Brothers Publishers.

Goodman LA (1979). Simple Models for the Analysis of Association in Cross Classifications Having Ordered Categories, J. Am. Stat. Assoc., 74(367): 537-552.

Greenland S (1985). An Application of Logistic Models to the Analysis of Ordinal Response," Biom. J. 27 (2): 189-197.

Grizzle J (1971). Multivariate Logit Analysis, Biometrics, 27: 1057-1062.

Haberman SJ (1974). Log-Linear Models for Frequency Tables with Ordered Classifications, Biometrics, 30: 589-600.

Labovitz S (1970). The Assignment of Numbers to Rank Order Categories, Am. Sociol. Rev. 35(3): 515-524.

Labovitz S (1971). In Defense of Assigning Numbers to Ranks, Am. Sociol. Rev., 36(3): 521-522.

Likert R (1932). A Technique for the Measurement of Attitudes," Arch. Psychol., 140: 1-55.

Mayer LS (1970). Comment on 'The Assignment of Numbers to Rank Order Categories,'" Am. Sociol. Rev. 35 (5): 916-917.

Mayer LS (1971). A Note on Treating Ordinal Data as Interval Data," Am. Sociol. Rev. 36 (3): 510-520.

McCullagh P (1980). Regression Models for Ordinal Data," J. Royal

Stat. Society, 42 Ser. B (2): 109-142.
McKelvey RD, Zavoina W (1975). A Statistical Model for the Analysis of Ordinal Level Dependent Variables," J. Math. Sociol. 4: 103-120.
Peterson B, Harrell Jr., Frank E (1990). Partial Proportional Odds Models for ordinal Response Variables," Appl. Stat., 39 (2): 205-217.
Rice SA (1928). Quantitative Methods in Politics, New York: Alfred A. Knopf.
Schweitzer S, Schweitzer DG (1971). Comment on the Pearson *R* in Random Number and Precise Functional Scale Transformations, Am. Sociol. Rev., 36 (3): 518-519.
Shapiro SS, Wilk MB (1965). An Analysis of Variance Test for Normality (Complete Samples), Biometrika, 52 (3, 4): 591-611.
Simon G (1974). Alternative Analyses for the Singly-Ordered Contingency Table, J. Am. Stat. Assoc. 69 (348): 971-976.
Stevens SS (1946). On the Theory of Scales of Measurement," Sci. 103 (2684): 677-680.
Tattersfield F, Giminghan CT, Morris HM (1925). Ann. Appl. Biol. 12: 61.
Theil H (1969). A Multinomial Extension of the Linear Logit Model, Int. Econ. Rev. 10: 251-259.
Theil H (1970). On the Estimation of Relationships Involving Qualitative Variables," Am. J. Sociol. 76 (1): 103-154.
Thorndike EL (1913). An Introduction to the Theory of Mental and Social Measurements (2^{nd} ed.), New York: Teachers College, Columbia University.
Thurstone LL (1928). Attitudes Can Be Measured, Am. J. Sociol. 33 (4): 529-554.
Vargo LG (1971). Comment on 'The Assignment of Numbers to Rank Order Categories,' Am. Sociol. Rev. 36 (3): 517-518.