THE WORKSHOP

Estimating Interval Scale Values for Survey Item Response Categories

Carl Hensler, *Social Research Data Systems*
Brian Stipak, *Pennsylvania State University*

Analysts of social science data often need the flexibility and power of interval-level statistics, even though strictly interval measurement has not been achieved. In such cases, analysts should attempt to assign category values that minimize distortion of the underlying variables. This paper describes a number of methods for category value estimation. The methods include estimation from the observed frequencies based on an assumption about the underlying distribution, estimation from single or multiple criterion variables, and estimation from item text. These techniques can be easily used, and their sensible application can avoid needless measurement error and resulting statistical bias.

Analysts of social science data often use statistical techniques that assume interval measurement. A common practical problem the analyst faces is how to assign values to ordered categories. If the observed item measures a continuous variable, category values should be chosen to minimize distortions of the underlying metric. This paper will first discuss why this methodological issue is important, and formalize a criterion of proper numerical assignment. After a discussion of current assignment practices, the bulk of the paper is devoted to presenting different techniques for calculating category values, and illustrating the computational procedures for a number of these techniques with a specific example.

The Practical Necessity and Potential Importance of Numerical Assignment

Because they frequently use interval-level statistics, social science researchers often face the problem of choosing an appropriate scoring system for ordered item categories. Although statisticians have been developing improved nominal and ordinal statistical methods, the large array of interval-level statistics based on the general linear model still offers the most flexible and best understood statistical tools for sophisticated social science research. For this reason, Abelson and Tukey (1970), Tufte

American Journal of Political Science, Vol. 23, No. 3, August 1979
 0092-5853/79/030627-23$01.85

(1970), Labovitz (1970), and others have argued against over-reliance on nonparametric methods. A debate has ensued as to whether certain interval-level statistics should *routinely* be used with data that are merely ranked-ordered (Labovitz, 1970, 1971; Mayer, 1970, 1971; Mayer and Robinson, 1977; Vargo, 1971; Schweitzer and Schweitzer, 1971). Unfortunately, this debate has tended to be polarized and polemic rather than addressed to the practical concerns of research.

Because of the clear advantages of interval-level statistics, data analysts will continue to use them even when measurement is not strictly interval. Resorting to rank-order statistics not only limits the analyst to a class of statistical techniques less adaptable to sophisticated social science research, but also discards information about inter-category distances. Widely-used statistical packages make interval-level statistics based on the general linear model conveniently available. An important practical question therefore is how to best assign category values when (1) measurement is somewhere between interval and ordinal, and (2) the analytic task requires the flexibility and power of interval-level statistics.

Researchers using available datasets cannot avoid this question simply by using pre-existing code values. Category values are typically assigned for coding convenience or to facilitate categorical and cross-tabular analysis[1] —*not* to represent properly the underlying metric. Therefore, social science data often have code values that clearly distort the underlying variable. For example, the 1968 and earlier ICPSR national election studies use the following code for income:[2]

10	<$1,000	20	$6,000–$6,999	30	$10,000–$11,999
11	$1,000–$1,999	21	$7,000–$7,999	31	$12,000–$14,999
12	$2,000–$2,999	22	$8,000–$8,999	32	$15,000–$19,000
13	$3,000–$3,999	23	$9,000–$9,999	33	$20,000–$24,999
14	$4,000–$4,999			35	$25,000 and over
15	$5,000–$5,999				

As Figure 1 illustrates, this coding produces a markedly nonlinear representation of income.[3] Thus, even in the case of a variable like income with

[1] As an historical note, coding schemes like that shown below were originally employed for ease in using card sorters or counter sorters to create crosstabulation tables.

[2] See variable 261 in the 1968 ICPSR American National Election study.

[3] The discontinuity of the relationship shown in Figure 1 is a necessary consequence of categorization, and is in fact inherent in all measures of truly continuous variables. Our concern lies in wisely choosing category values that maximize the linear relationship between the true variable and its measure.

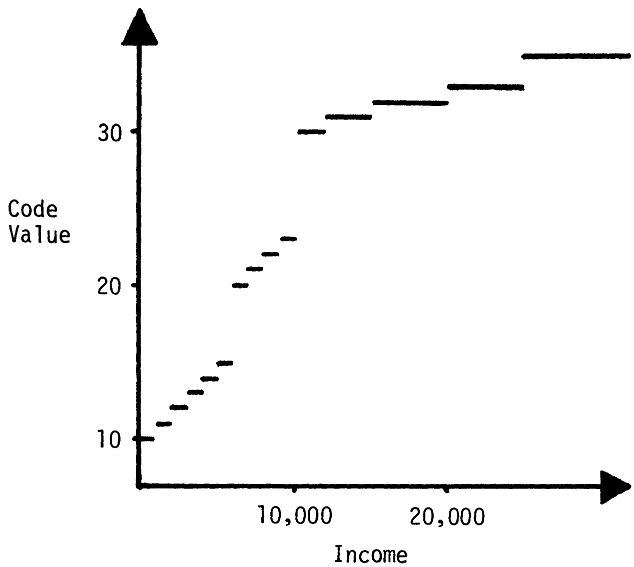


FIGURE 1

Example of Markedly Non-Linear Coding of Response Categories

a simple metric, researchers must be careful to scrutinize the appropriateness of category values before employing interval-level statistics.

The researcher's choice among alternative sets of category values, all consistent with the same rank order, can have potentially critical statistical consequences, especially for multivariate analyses. If the number of categories is large and if the inaccuracies introduced by the choice of category values is random, then inappropriate category coding has the deleterious impact of random measurement error: the effects of that variable will be underestimated, and some of its effects will be erroneously attributed to correlated variables. If the number of categories is small or if the assignment inaccuracies occur in some systematic fashion, all sorts of statistical biases become possible. Even the sign of the correlation coefficient is not necessarily invariant with respect to order-preserving transformations of category values.[4] Careless numerical assignment can result, therefore, in misleading statistical findings.

[4] See Grether (1974, 1976) for a discussion of the conditions that determine if the sign of a correlation is invariant with respect to order-preserving transformations of one or both variables.

A Criterion for Proper Numerical Assignment

Proper assignment of category values requires assigning values that minimize nonlinear distortions of the underlying variable being measured. The task can be thought of as preserving the relative inter-category distances—that is, constructing an observed variable that is a linear transformation of the underlying variable (plus an error term due to loss of intra-category information). The original scale usually cannot be established, since the units of the underlying metric are typically undefined, as in the case of a political attitude.[5]

Consistent with the analytical methods most likely to be used, the criterion we have chosen is to minimize the mean squared error in linearly predicting the underlying variable from the assigned category values. This maximizes the correlation between the underlying variable and its observed measure, and thus minimizes the average squared error of measurement. The minimum variance estimator of the underlying variable given an observed category is its conditional expected value. Thus, the values assigned to categories of the observed variable must be a linear function of the conditional expectations of the underlying variable given the observed variable:

$$x_j = a + b \cdot E(X^*|j)$$

where: x_j = value assigned j^{th} category
X^* = underlying variable

Estimating these conditional expectations requires some knowledge about how the underlying variable is mapped onto the observed variable. Such knowledge can be of an a priori nature or can be derived from empirical analysis.

Current Practices and General Assignment Strategies

One current practice of researchers, when using existing datasets such as the ICPSR election studies, is simply to analyze the variables as they are coded. As shown above, the unthinking use of pre-existing code values can result in markedly nonlinear relationships between the observed and true variables. Even if the coding scheme is not obviously distorted, pre-existing category values seldom were assigned with the underlying metric explicitly

[5] Inability to establish the underlying metric is of no consequence when using statistics (e.g. correlations) invariant under changes of scale. Also, for regression applications the analyst can sometimes give the observed variable a more meaningful metric by setting its standard deviation to a convenient quantity.

in mind. Rather, category values probably resulted from the application of a simple general assignment strategy.

Current practice typically uses the equal-interval strategy of assigning rank-order numbers as category values. Arguments can be made that simple equal-interval scoring is a good general assignment strategy. Labovitz (1970, p. 520) has recommended imposing an equidistant scoring system on ordinal categories on the grounds that such a system has the advantage of lying midway between other possible scoring systems. Also, survey respondents may tend to impose an equal-interval interpretation on item alternatives independent of their wording.[6] Unfortunately, researchers often use rank-order numbers naively—perhaps because the data were originally coded that way, and without recognizing that rank-order assignment implies equal inter-category intervals.

Abelson and Tukey (1970) have proposed another general assignment strategy that maximizes the minimum possible correlation of the chosen assignment with the proper assignment. This "maximin" strategy leads to markedly nonequidistant assignment solutions.[7] We feel that a pure maximin strategy is overly conservative for most social science research. An equal-interval scheme is probably preferable as a general strategy.[8]

However, rigid application of any general assignment strategy is not advisable. These strategies are inefficient because they fail to use all information available to the analyst. Such information can be used either formally or informally to improve the choice of category values and better satisfy our criterion of proper numerical assignment.

Relevant Theoretical Literature

The assignment of values to response categories is referred to as item scoring in the test theory literature (e.g. Lord and Novick, 1968, pp. 302–326). This literature typically uses our minimum mean squared error criterion, but unfortunately it concentrates on the effects of guessing and other problems in the measurement of knowledge and ability. Thus test

[6] This is not to say that category labels have no effect, merely that their effect on inter-category distances is somewhat circumscribed.

[7] See Abelson and Tukey (1970, p. 410). The correct maximin category values can be quickly approximated by (1) choosing an equal-interval set of values having a mean of zero, (2) quadrupling the extreme values, and (3) doubling the next-to-end values (Abelson and Tukey, 1970, p. 411).

[8] Abelson and Tukey (1970, p. 411) point out that the analyst should replace the general maximin strategy with an almost equal-interval assignment strategy if he is willing to assume that categories are approximately equidistant.

theory is an interesting theoretical source and might be useful for scoring political knowledge items, but otherwise offers little help for the survey data analyst. The item scoring problem is often simply ignored, even in comprehensive treatments of practical test construction (e.g. Nunnally, 1967).

The psychometric scaling literature (e.g. Guilford, 1954, ch. 10) also considers the problem of assigning scale values to ordered response categories. Due both to its theoretical perspective and the type of data it requires, psychometric scaling is not very useful for estimating category values directly from one's survey data. However, various scaling methods can be used to estimate category values from the item text, as we will discuss later.

An Empirical Example

For purposes of illustrating the computational procedures of several of our estimation techniques, we will estimate category values for an item from the 1972 CPS American National Election Study. This item is a standard measure of political involvement used in the Michigan election studies:[9]

> Some people seem to follow what's going on in government and public affairs most of the time, whether there's an election going on or not. Others aren't that interested. Would you say you follow what's going on in government and public affairs most of the time, some of the time, only now and then, or hardly at all?

The marginal percentages obtained in 1972 were the following, based on a sample size of 1076:[10]

36% most of the time
36% some of the time
17% only now and then
10% hardly at all

One approach to scoring this item is simply to use rank-order numbers as category values, which is the way the item is coded in the ICPSR dataset. In the following sections we will show how alternative scoring systems for this item can be estimated. Since some of these techniques require using similar items as criterion variables, this item is a good choice

[9] This item appears as variable 476 in the 1972 ICPSR codebook.

[10] This sample of 1076 consists of cases that (1) are Form I interviews, (2) have both pre- and post-election interviews, and (3) have no missing data on any of the six criterion variables used in later sections. This subset constitutes the sample used for analyses presented in this paper.

for illustration because a number of other involvement items are available in the 1972 study.

Estimation from Observed Frequencies and a Distributional Assumption

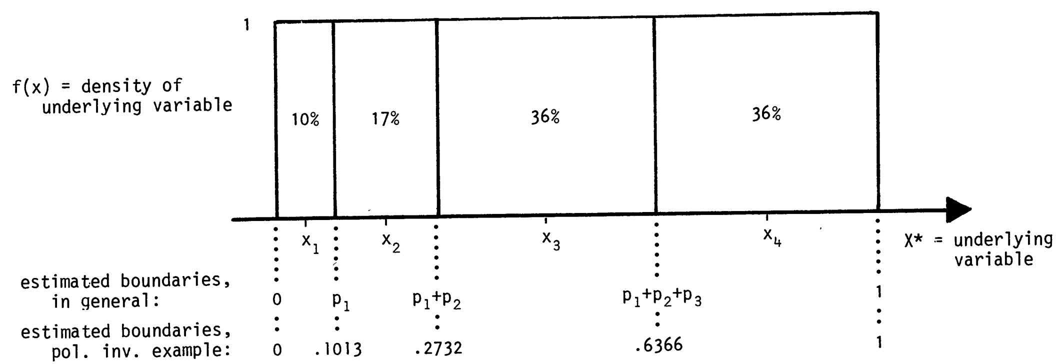
If the researcher believes the underlying variable is distributed in a particular way, he can often estimate category values from the observed frequencies. This is possible by assuming that the observed categories correspond to separate segments under the density function of the underlying variable. That is, we think of the observed category proportions as estimates of areas under the probability density function of the underlying variable, which are mapped into the observed categories.[11] The expected value of the underlying variable conditioned on category j having been observed, $E(X^*|j)$, is then simply the mean value for the segment of the density function corresponding to category j. Thus, the researcher need only calculate those means and assign them as category values. This section will show how to use this method of estimating category values assuming (1) a rectangular distribution, and (2) a normal distribution.[12]

If the researcher assumes that the underlying variable is rectangularly (evenly) distributed, then he simply assigns to each category the average of the two endpoints of the corresponding segment of the density function. That is, the mean of segment j is $(a_j + b_j) \div 2$, where a_j is the lower bound and b_j is the upper bound. Letting p_i $(i = 1, \ldots, n)$ be the proportion of cases in each of n categories of the observed variable, and arbitrarily defining the underlying distribution over the interval 0 to 1, the estimated lower and upper boundaries of segment j of the density function are $\sum_{i=1}^{j-1} p_i$ and $\sum_{i=1}^{j} p_i$. The lower boundary is understood to be 0 if j is 1. Thus, the boundaries are 0 and p_1 for the first segment, p_1 and $p_1 + p_2$ for the second segment, and $1-p_n$ and 1 for the last segment.

Figure 2 illustrates this procedure for our four-category political involvement example. The calculated category values $(x_1, \ldots, x_4)$ are .051,

[11] This model assumes a perfect measurement process in which an alternative is chosen if and only if the underlying variable falls in a particular range. More realistic response models could be used, but would seriously complicate estimation. We feel that whatever minimal improvement in estimation might result would not warrant the complication.

[12] An appendix showing the formal derivation of the procedures can be obtained from Brian Stipak, Institute of Public Administration, 211 Burrowes Bld., Pennsylvania State University, University Park, PA 16802.



<u>Estimated Category Values</u>

in general, $E(X^*\|j) \approx x_j =$	$\frac{p_1}{2}$	$p_1+\frac{p_2}{2}$	$p_1+p_2+\frac{p_3}{2}$	$p_1+p_2+p_3+\frac{p_4}{2}$
for the pol. inv. example:	$\frac{.1013}{2}$	$.1013+\frac{.1719}{2}$	$.2732+\frac{.3634}{2}$	$.6366+\frac{.3634}{2}$
	= .051	.187	.455	.818

FIGURE 2

Category Value Estimates for a Four-Category Item,
Assuming a Rectangularly Distributed Underlying Variable

.187, .455, and .818. The researcher can linearly transform these values as he desires to yield a computationally convenient range.

An assumption of a normal rather than a rectangular underlying variable may be more often applicable, since by the Central Limit Theorem variables that can be thought of as the sum of a number of roughly independent factors will tend to be approximately normal.[13] If the researcher assumes that the underlying variable is normally distributed, then he can estimate the upper and lower boundaries of each segment of the density function using a table of areas under the standard normal curve. The lower boundary of the first category and the upper boundary of the last category extend to $-\infty$ and ∞, respectively. For example, if the first category contains 5 percent of the cases and the second category contains 45 percent of the cases, the estimates for their respective boundaries under the standard normal curve are $-\infty$ and -1.96, and -1.96 and 0.

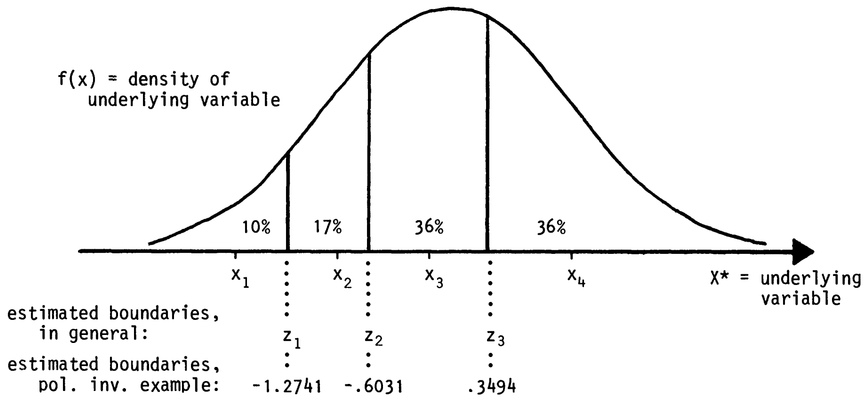
Once upper and lower boundaries have been obtained using a table of areas under the standard normal curve, category values are easily estimated using the procedure presented in Guilford (1954, p. 237). Consulting a table of the ordinates of the standard normal curve, the researcher simply subtracts the ordinate for the upper boundary from the ordinate for the lower boundary, and divides the result by the proportion in that category. Figure 3 illustrates this method for the political involvement item.

Even when the researcher does not use formal methods for estimating category values, he should be aware of the logical relationship among (1) the observed frequencies, (2) the inter-category distances, and (3) the shape of the underlying distribution. For example, if the category frequencies are about equal, then the analyst can justify assigning rank-order numbers only if he can reasonably assume the underlying variable is rectangularly distributed. If the analyst feels sure that the underlying distribution is not rectangular, he must employ a nonequidistant assignment scheme. If he assumes the underlying variable is normal, then equal category frequencies imply unequal inter-category intervals. Researchers aware of these relationships can at least assign category values that are logically consistent with their assumptions and the data.

Estimation from Criterion Variables

When the researcher has reason to assume that the probability distribution of the underlying variable in the observed population is rectangular,

[13] Invoking the Central Limit Theorem formally requires that the component variables be independent and identically distributed. For practical purposes a tendency towards normality can usually be expected when the component variables are roughly independent and not greatly different in variance.



<u>Estimated Category Values</u>

in general, $E(X^*|j) \approx x_j =$

$$\frac{-f(z_1)}{p_1} \quad \frac{f(z_1)-f(z_2)}{p_2} \quad \frac{f(z_2)-f(z_3)}{p_3} \quad \frac{f(z_3)}{p_4}$$

for the pol. inv. example:

$$\frac{-.1772}{.1013} \quad \frac{.1772-.3326}{.1719} \quad \frac{.3326-.3753}{.3634} \quad \frac{.3753}{.3634}$$

$$= -1.749 \quad -.904 \quad -.118 \quad 1.033$$

FIGURE 3

Category Value Estimates for a Four-Category Item, Assuming a Normally Distributed Underlying Variable normal, or otherwise mathematically tractable, estimation from the observed frequencies is probably simplest. However, in many cases there is no basis to assume a particular distribution. On the contrary, there is often reason to think that the distribution is skewed, multi-modal, or otherwise intractable in some uncertain fashion. The analyst must then look beyond the item in question for some information about its functional relationship to the underlying variable.

The analyst can obtain this information from one or more observed variables, given an assumption about the relationship of those variables to the underlying variable. We will call such variables "criterion variables." A criterion variable can be an alternative measure of the underlying variable, or a variable which is causally linked to it. The critical requirements are (1) the analyst must have data from the same cases for both the target variable—the variable for which category values are to be estimated—and the criterion variable(s), and (2) the functional relationship between the underlying variable and the criterion variable(s) must be known or assumed. Note that the dataset used to estimate category values for the target variable need not be the same dataset used for substantive analysis. Indeed, keeping the estimation of category values entirely separate from the substantive analysis has the advantage of making sampling and measurement errors in the category estimates statistically independent of these types of errors in the data used for substantive analysis.

Single Alternative Measure Criterion

The simplest case involves a single criterion variable that is an alternative, linear measure of the underlying variable. For example, suppose a researcher has responses for two items requiring respondents to evaluate the President: first, a graphic rating scale such as the Michigan SRC "feeling thermometer" (Weisberg and Rusk, 1970, pp. 1175, 1185), and second, a purely verbal evaluation item such as "Do you think ——— has been doing a very good job, a good job, a fair job, or a poor job as President?" If the researcher assumes the rating scale is a linear measure of the underlying variable, the respondent's evaluation of the President, the researcher can use that scale to estimate category values for the verbal item.

As was discussed previously, our goal is to assign values to categories of the target variable which are a linear function of the conditional expected values of the underlying variable. When the criterion variable is linearly related to the underlying variable, the sample means of the criterion variable for cases falling in each of the target variable's categories will be unbiased, minimum variance estimates of a linear function of the conditional expected values, providing no other variables distort the relationship between the target and criterion. Additional independent variables that are correlated with the target variable and affect the criterion variable can be accounted for statistically, as a later section describes. If it is not necessary to account for other variables introducing correlated error, one simply computes the mean value of the criterion variable for each of the subgroups of cases defined by the categories of the target variable. The researcher can linearly rescale these conditional means to produce computationally more convenient category values for the target variable.

The analyst should be concerned about the statistical confidence he can place in these category value estimates, particularly if some of the categories contain small numbers of cases. Assuming simple random sampling, the standard error of the estimated category value, x_j, for the j^{th} category is:

$$\sigma_{x_j} = \sigma(\bar{C}|j) = \frac{\sigma\,(C|j)}{\sqrt{n_j}}$$

where: $\sigma(C|j)$ = standard deviation of the criterion variable within the j^{th} category

n_j = number of cases in the j^{th} category

The standard error for the estimated interval between the i^{th} and j^{th} categories is:

$$\sigma_{x_i - x_j} = \sqrt{\sigma_{x_i}^2 + \sigma_{x_j}^2}$$

Plus or minus two times the standard error is approximately a 95 percent confidence interval for an estimate. Of course, if the category value estimates have been rescaled by multiplying by some constant, the standard errors must be rescaled by the same constant.

An inevitable question is how strongly must the criterion variable be related to the target variable. Consider what happens to the category value estimates as the strength of the relationship between the target and criterion variable declines. First, their variance about the overall mean of the criterion variable decreases as they collapse toward this common value. Second, their sampling variances increase as the within category variances increase. The weaker the relationship between the target and criterion variables, the larger the sampling variances of the category values compared to their variance about their overall mean, and the less confidence one can place in the estimates. Practically speaking, the target-criterion relationship must be sufficiently strong that the inter-category distances are large compared to their confidence intervals. Obviously, larger sample sizes permit the analyst to use criterion variables having weaker relationships.

Single Casually Related Criterion

It is not necessary that a criterion variable be an alternative measure of the underlying variable, only that it be statistically related to the underlying variable in a known functional way. If the relationship is linear, the conditional means of the criterion variable serve as category value estimates, as in the case of a single alternative measure discussed above.

However, if the relationship is a known nonlinear function, then the criterion variable must be transformed first. For example, suppose one wished to estimate occupational prestige scores for the traditional census occupational classes (professional and technical; managers, officials and proprietors; clerical and sales; etc.) from the annual income of heads-of-household. On several grounds (see Simon, 1957) one might assume that the prestige of an occupation is proportional to the logarithm of its pay rate, so that equal percentage increments in income correspond to equal prestige intervals. If the income data have been categorized into ranges, one would transform the income variable by assigning to each income category the average of the logs of the two category endpoints. Taking the income range of $10,000 to $12,000 a year as an example:

$$\{\log_{10}(10{,}000) + \log_{10}(12{,}000)\}/2 = \{4.00 + 4.08\}/2 = 4.04$$

Then one would compute the mean of this log-income variable for each occupational category to estimate the prestige scale values.

Single Criterion Determined by a Target Variable and Other Independent Variables

The type of bivariate estimating procedures described above are not always appropriate to estimate category values for an independent (target) variable from its effect on a dependent (criterion) variable. In particular, if other independent variables that are correlated with the target variable affect the criterion variable, failing to take them into account will produce biased category estimates.[14] However, this estimation bias will cause non-

[14] This follows from the basic principle that omission of a relevant variable (one that has an effect upon the dependent variable) from an estimating equation biases estimates of the effects of all other independent variables that are correlated with the omitted variable. Failure to include a relevant variable is referred to as "specification error" in econometrics. See Kmenta (1971, pp. 391–395) or any other econometrics text.

linear distortion of the category value estimates only if the effects of the omitted variables are nonlinearly related to the target variable. Purely linear omitted effects may affect the efficiency of estimation, but will not cause nonlinear biases.

Whenever other independent variables must be taken into account to estimate properly the relative effects of the categories of the target variable, the analyst can no longer simply compute category means. Instead, he must use multiple regression or some other analytic technique[15] to estimate simultaneously the effects of several independent variables upon the criterion variable. The standard procedure for estimating the effects of a qualitative variable in a multivariate model is to use dummy (binary, 1/0) variables to represent *n*-1 of its *n* categories.[16] The regression coefficients obtained for the dummy variables can then be treated as category value estimates relative to the zero point defined by the omitted category. The standard errors for the dummies can be multiplied by two to obtain approximately 95 percent confidence intervals for the category value estimates. In this way the analyst can estimate category values which reflect the target variable's effects upon a criterion variable, using a model that includes other independent variables.

Multiple Criterion Variables

When the analyst has several criterion variables, he can sum them with arbitrary (usually equal) weights to form a criterion index. Alternatively, he can use canonical correlation to estimate weights that maximize the strength of the relationship of the criterion index with the target variable, as described in the next section. In either case, using multiple criteria rather than a single criterion has at least two advantages. First, since the sum of several criterion variables will generally have a higher reliability, the category value estimates will be more precise. Second, while it is often impossible to find a single criterion that displays the requisite metric characteristics, the sum of several monotone items can usually be accepted as a relatively linear interval measure. Ideally, the items should represent a wide range of item difficulties to maintain linearity near the scale endpoints (Nunnally, 1967, p. 71).

To illustrate the simplest use of multiple criterion variables, we will estimate category values for the political involvement item using an equally-weighted criterion index. This index is the sum of six dummy variables

[15] Logistic function analysis or discriminant function analysis are appropriate if the dependent variable is dichotomous or polytomous.

[16] See Nie et al. (1975, pp. 373–383) or Kmenta (1971, pp. 409–425) for lucid discussions of the use of dummy variables.

created from other political involvement items in the 1972 CPS National Election Study dataset. Each of these six dummy variables takes the value one under the following six respective conditions:

(1) The respondent says he remembers something about the controversy over the selection of the Democratic Party's vice presidential candidate (VAR47=1).
(2) The respondent says he has been very much interested in following the political campaigns (VAR163=1).
(3) The respondent is able to rate himself, Nixon, and McGovern on the governmental guarantee of job and standard of living scale (VAR173, VAR174=1−7. Note if VAR172≠1−7, VAR173, VAR174=0).
(4) The respondent says he reads newspaper articles about the election regularly or often (VAR457=1, 2).
(5) The respondent says people generally consider Republicans more conservative than Democrats (VAR500=2, 4).
(6) The respondent correctly states the term of a U.S. Senator is six years (VAR944=6).

The unweighted sum of these dummy variables constitutes an index of the respondent's political involvement and knowledge.[17] To use this index to estimate category values for our four-category involvement item, we simply calculate the mean of the index for respondents falling in each of the four item categories.[18] The resulting values, which can of course be linearly rescaled as desired, are the following:

3.893 most of the time
2.913 some of the time
2.060 only now and then
1.312 hardly at all

The appropriate criterion in some cases may be a special type of algebraic function of two or more variables. For example, suppose one wants to estimate category values for the standard Michigan SRC party identification question (Campbell et al., 1960, pp. 122–124). One criterion might be the algebraic difference between the respondent's evaluation of "Republicans" and "Democrats" on a "feeling thermometer" rating scale (Weisberg and Rusk, 1970, p. 1168).

[17] See Hensler (1971) for a structural analysis of the orientations toward government tapped by these types of items.

[18] All cases with missing value codes of 9 were excluded from these analyses. This yields a total number of cases of 1076 (see note 10).

Differential Weights for Criterion Variables

When the criteria form a conceptually homogenous and highly intercorrelated set of variables, there is little need for differential weighting. But if the criterion variables are fairly heterogeneous, weighting becomes necessary. One approach is to factor analyze the criterion variables and use the factor score, if the researcher can decide which factor is appropriate, as the criterion. A neater approach is to use canonical correlation to simultaneously estimate weights for the criterion variables and category values for the target variable.

Canonical correlation blends factor analysis and multiple regression. The input data are two sets of variables. The procedure estimates a canonical variate—a linear weighted sum of the variables—for each of the two sets. These weights are chosen to maximize the correlation between the canonical variates.

In this application, the criterion variables form one set of input variables, and *n*-1 dummies representing the *n*-category target variable form the other.[19] Thus, canonical correlation seeks a set of category values—the weights for the category dummies—that maximizes the correlation of the target canonical variate with the criterion canonical variate. This is quite similar to the use of dummies and multiple regression to estimate category values as described in a previous section, with two exceptions. First, the canonical coefficients generated by most programs are for standardized rather than raw-score variables. Thus, it is usually necessary to divide the coefficients for the target category dummies by their standard deviations to obtain category values estimates. Second, while multiple regression produces only one solution, canonical correlation produces a series of orthogonal solutions analogous to factors in factor analysis. In this application, the first solution should produce results that both appear reasonable and have a much larger canonical correlation than the second solution. If not, either the set of criterion variables is too heterogeneous, or the analyst is wrong about the nature of the relationship between the target and the criteria.

We used this technique to estimate category values for our political involvement example, employing the same six criterion variables used in the prior section. The first set of input variables consists of three dummy variables, each of which takes the value one if the respondent chose the

[19] Klatzky and Hodge (1971) originally described this use for canonical correlation.

category it represents. Since we let the second category be the omitted category, its coefficient is assumed to be zero. The second set of input variables is composed simply of the six criterion variables used earlier. The following unstandardized coefficients for the target dummy variables were calculated:

1.148 most of the time
0 some of the time
—1.009 only now and then
—1.855 hardly at all

These coefficients, or any convenient linear transformation of them, serve as the category value estimates.

Estimating Category Values from Item Text

Judgmental Estimation

In judgmental estimation one or more judges rate the relative strengths of an item's categories. At the most elementary level the analyst, acting as sole judge, may feel that some inter-category intervals are larger than others, and assign values to reflect that belief. A more rigorous approach is to have a panel of judges rate the categories using established psychometric techniques. Spector (1976) has used the normalized rank method (Guilford, 1954, pp. 181–183) to estimate scale values for commonly used item response categories. Unfortunately, this technique makes the tenuous assumption that the scale values are normally distributed, and is unnecessarily complicated. We feel that the graphic rating scale method (Guilford, 1954, p. 265) is preferable.

As an example of the latter approach, suppose we wished to estimate category values for our political involvement example. Each judge would be given the text of the item, and asked to place each of the responses on the political interest rating scale shown in Figure 4. Note that, in order to avoid error due to number preference and rounding, the scale is not numbered. The judge marks the position he estimates for each category, and the analyst measures the distance in millimeters from the left end of the scale. The judges' ratings would be aggregated by transforming the four category values obtained from each judge to a common mean and standard deviation, and then computing a mean for each category over all the judges. These average, standardized category ratings would then be used as category values in subsequent data analysis.

Where on this line would you place a person giving this answer to the question?

No interest in politics Never pays any attention to politics Knows nothing about what is going on in politics	Extremely interested in politics Spends much time following political news Knows all about what is going on in politics

FIGURE 4

Political Interest Rating Scale

Psycholinguistic Estimation

Psycholinguistic research has shown that the strength of adverbs and adjectives can be expressed quantitatively, and that the effect of an adverb upon an adjective is multiplicative (Cliff, 1959; Howe, 1963; Lilly, 1968, 1969). Since the alternatives presented in survey items are often adverb-adjective combinations, e.g. "very favorable, somewhat favorable, etc.," it is sometimes possible to use the strength measure tables produced by this research to estimate relative values for item categories. The tabled strength values can also be used to guide the choice of wording for new items. Thus, psycholinguistic research can contribute an objective base to the traditionally subjective art of writing survey items.

Comparison of Alternative Assignment Strategies for the Political Involvement Example

Table 1 compares the relative inter-category distances for the sets of category values estimated earlier for the political involvement example. Sets of category values based on equal-interval assignment and on Abelson and

TABLE 1

Comparison of Relative Inter-Category Distances for the Political Involvement Item Using Different Estimation Techniques

	Category 1-2	Category 2-3	Category 3-4
equal-interval	1.00	1.00	1.00
maximin	2.73	1.00	2.73
even assumption	.51	1.00	1.36
normal assumption	1.07	1.00	1.46
equally-weighted criterion index	.88	1.00	1.15
differentially-weighted criterion index	.84	1.00	1.14

Note: Each figure is the inter-category distance, under that particular choice of category values, expressed as a proportion of the distance between the two middle categories.

Tukey's (1970) maximin strategy are included also.[20] For each of these alternative sets, Table 1 shows the relative distance between each extreme category and the nearest middle category, compared to the distance between the two middle categories.

The techniques differ considerably in the relative inter-category distances that resulted for this example. The inter-category distances for the values derived from criterion variable estimation are almost equal. The maximin inter-category distances, on the other hand, are very unequal, since the outer categories are pushed toward the extremes. In the case of the category values based on an even and normal assumption, the inter-category distances are unequal, but less markedly unequal than for the maximin values.

These differences in relative inter-category distances would have little consequence for simple statistical procedures. The product-moment correlations among the alternative versions of the political involvement item range from .973 to .999. This is not surprising, since Labovitz (1970) has shown that the product-moment correlation is highly stable over a variety of nonlinear monotone transformations of the numbers assigned to categories.

However, in some multivariate analyses even minor differences in assignment schemes that allow only a small percentage of a variable's variance to be unaccounted for may create important distortions. That portion of the variable's effects can then be attributed erroneously to correlated variables. If errors are introduced in a variable with high correlations with other variables, even small distortions can substantially bias parameter estimates for other variables. Moreover, when differences among the estimation techniques are greater than in the case of our illustration, the statistical implications of the choice become greater. For example, assume that the true relative inter-category distances for our political involvement item were 1, 1, and 10. If equal-interval category values were then used instead of the true values (1, 2, 3, 13), the variable would correlate only .853 with the variable as correctly scored. Thus, over 27 percent of the correctly-scored variable's variance would be unaccounted for and free to contaminate the results for correlated variables in the analysis.

Although the criterion variable estimation techniques resulted in close to equal-interval category values for our example, this is not a necessary result of such techniques. Indeed, for some poorly-worded items or items tapping underlying variables with peculiar distributions, we would expect

[20] The maximin sequence for four categories is —.866, —.134, .134, .866 (Abelson and Tukey, 1970, p. 410).

criterion variable estimates to be very unequal-interval. In the case of the political involvement item, the nearly equal-interval criterion variable estimates support the reasonableness of an equal-interval assignment. In other cases criterion variable estimates might indicate that an equal-interval strategy would be a poor choice.

Recommendations

Analysts applying interval-level statistics to survey data should carefully scrutinize pre-existing category values to guard against using any markedly nonlinear measures of the true variables. However, simple statistical analyses will usually not be highly sensitive to minor variations in the choice of category values. Since the product-moment correlation is quite stable over a wide variety of nonlinear monotone transformations of the numbers assigned to categories, researchers using only bivariate correlations or other very simple statistical techniques need not expend great effort in choosing among slightly different assignment schemes. In such cases equal-interval assignment will often be a reasonable choice.

For more sophisticated research the proper assignment of category values, as well as other measurement problems, can be crucial. Although the choice of category values may have only minor effects on simple correlations, it may make a critical difference when estimating the regression coefficients of two highly correlated independent variables, or when comparing alternative causal models. Researchers can with little additional trouble avoid needless distortions in multivariate statistical analyses by sensibly applying techniques of category value estimation described in this paper.

When the researcher strongly believes that the distribution of the underlying variable is approximately normal or rectangular, estimation from the observed frequencies is the most straightforward approach. This technique requires no data beyond the marginal frequencies, and can be done quickly using readily available statistical tables. Obviously, if the researcher has little basis to assume a particular mathematically tractable distribution, other estimation techniques must be used. We have described a number of easily applied methods of calculating category values using criterion variables. These methods can themselves yield important substantive results.[21] Even when unable to estimate category values using the data, analysts have methods available for estimating category values from item text. At the

[21] For example, see Klatzky and Hodge's (1971) analysis of occupational mobility.

most elementary level the analyst takes but a minute to assign category values which reflect his subjective belief about relative inter-category distances.

In short, social science researchers using sophisticated multivariate statistics have available to them a variety of easily applied methods of estimating category values. Researchers should, therefore, only resort to a simple general assignment strategy when they consciously decide it is the best alternative, rather than unthinkingly using rank-order numbers out of ignorance of available alternatives. Empirical estimates of category values should be reported in the research literature so that those values can be used and further tested by others.

Manuscript submitted 31 March 1978
Final manuscript received 29 August 1978

REFERENCES

Abelson, Robert P., and John Tukey. 1970. Efficient conversion of non-metric information into metric information. In Edward R. Tufte, ed., *The quantitative analysis of social problems.* Reading, Mass.: Addison-Wesley, pp. 407–417.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American voter.* New York: Wiley.

Cliff, Norman. 1959. Adverbs as multipliers. *Psychological Review,* 66 (January 1959): 27–44.

Grether, David M. 1974. Correlations with ordinal data. *Journal of Econometrics,* 2 (April 1974): 241–246.

———. 1976. On the use of ordinal data in correlation analysis. *American Sociological Review,* 41 (October 1976): 908–912.

Guilford, J. P. 1954. *Psychometric methods.* New York: McGraw-Hill.

Hensler, Carl. 1971. The structure of orientations toward government. Ph.D. Dissertation, Massachusetts Institute of Technology.

Howe, Edmund S. 1963. Probabilistic adverbial qualifications of adjectives. *Journal of Verbal Learning and Verbal Behavior,* 1 (January 1963): 225–241.

Klatzky, Sheila R. and Robert W. Hodge. 1971. A canonical correlation analysis of occupational mobility. *Journal of the American Statistical Association,* 66 (March 1971): 16–22.

Kmenta, Jan. 1971. *Elements of econometrics.* New York: Macmillan.

Labovitz, Sanford. 1970. The assignment of numbers to rank order categories. *American Sociological Review,* 35 (June 1970): 515–524.

———. 1971. In defense of assigning numbers to ranks. *American Sociological Review,* 36 (June 1971): 521–522.

Lilly, Roy S. 1968. The qualification of evaluative adjectives by frequency adverbs. *Journal of Verbal Learning and Verbal Behavior,* 7 (April 1968): 333–336.

———. 1969. Adverbial qualification of adjectives connoting activity. *Journal of Verbal Learning and Verbal Behavior,* 8 (April 1969): 313–315.

Lord, Frederic M., and Melvin R. Novick. 1968. *Statistical theories of mental test scores.* Reading, Mass.: Addison-Wesley.

Mayer, Lawrence S. 1970. Comment on the assignment of numbers to rank order categories. *American Sociological Review,* 35 (October 1970): 916–917.

———. 1971. A note on treating ordinal data as interval data. *American Sociological Review,* 36 (June 1971): 519–520.

Mayer, Lawrence S., and Jeffrey A. Robinson. 1977. Measures of association for multiple regression models with ordinal predictor variables. In Karl F. Schuessler, ed., *Sociological methodology* 1978. San Francisco: Jossey-Bass, pp. 141–163.

Nie, Norman H., C. Hadlai Hull, Jean G. Jenkins, Karin Steinbrenner, and Dale H. Bent. 1975. *Statistical package for the social sciences.* New York: McGraw-Hill.

Nunnally, Jum C. 1967. *Psychometric theory.* New York: McGraw-Hill.

Schweitzer, Sybil, and Donald G. Schweitzer. 1971. Comment on the Pearson R in random number and precise functional scale transformations. *American Sociological Review,* 36 (June 1971): 518–519.

Simon, Herbert A. 1957. The compensation of executives. *Sociometry,* 20 (March 1957): 32–35.

Spector, Paul E. 1976. Choosing response categories for summated rating scales. *Journal of Applied Psychology,* 61 (June 1976): 374–375.

Tufte, Edward R. 1970. Improving data analysis in political science. In Edward R. Tufte, ed., *The quantitative analysis of social problems.* Reading, Mass.: Addison-Wesley, pp. 437–449.

Vargo, Louis G., 1971. Comment on the assignment of numbers to rank order categories. *American Sociological Review,* 36 (June 1971): 517–518.

Weisberg, Herbert F., and Jerrold G. Rusk. 1970. Dimensions of candidate evaluation. *American Political Science Review,* 64 (December 1970): 1167–1185.