Applying R^2-Type Measures to Ordered Categorical Data

Author(s): Alan Agresti

Source: *Technometrics*, Vol. 28, No. 2 (May, 1986), pp. 133-138

Published by: Taylor & Francis, Ltd. on behalf of American Statistical Association and American Society for Quality

Stable URL: http://www.jstor.org/stable/1270449

Accessed: 20-12-2016 16:23 UTC

Applying R^2-Type Measures to Ordered Categorical Data

Alan Agresti

Department of Statistics
University of Florida
Gainesville, FL 32611

The concentration and entropy measures for categorical data tend to be highly dependent on the choice of response categories. If the categorization of the response variable is arbitrary but it is reasonable to assume an underlying continuous distribution, then an adaptation of the regression R^2 measure can be useful for describing multiple association.

KEY WORDS: Analysis of dispersion; Concentration; Contingency table; Correlation ratio; Entropy; Log-linear models; Ordinal data; Proportional reduction in variance; R squared.

1. INTRODUCTION

Let Y denote a categorical response variable, and let $\mathbf{X}$ denote a set of explanatory variables. This article considers measures of multiple association between Y and $\mathbf{X}$ that are analogous to the R^2 measure for regression models. The measures describe how well Y can be predicted for the model chosen to fit the data, and they can all be expressed in proportional reduction in dispersion form.

Let $(\hat{\pi}_{1(a)}, \ldots, \hat{\pi}_{r(a)})$ denote the distribution for the response variable that is estimated by the model at the setting $\mathbf{x}_a$ of the explanatory variables corresponding to the ath observation, $a = 1, \ldots, n$. Let $D(Y_a)$ denote a measure of dispersion for the ath observation relative to the estimated marginal distribution $(\hat{\pi}_1, \ldots, \hat{\pi}_r)$ of Y, and let $D(Y_a \mid \mathbf{X}_a)$ represent this measure computed for $(\hat{\pi}_{1(a)}, \ldots, \hat{\pi}_{r(a)})$. Then measures of association based on proportional reduction in dispersion have the form

$$\frac{\sum_{a=1}^{n} D(Y_a) - \sum_{a=1}^{n} D(Y_a \mid \mathbf{X}_a)}{\sum_{a=1}^{n} D(Y_a)}. \qquad (1.1)$$

Haberman (1982) and Magidson (1981) presented two measures of multiple association for categorical data. The Gini *concentration* measure C has $D(Y_a) = 1 - \sum_j \hat{\pi}_j^2$ for all a, and $D(Y_a \mid \mathbf{X}_a) = 1 - \sum_j \hat{\pi}_{j(a)}^2$, so

$$C = \frac{\sum_a \sum_j \hat{\pi}_{j(a)}^2 - n \sum_j \hat{\pi}_j^2}{n\left(1 - \sum_j \hat{\pi}_j^2\right)}. \qquad (1.2)$$

The *entropy* measure H has $D(Y_a) = -\sum_j \hat{\pi}_j \log \hat{\pi}_j$ for all a, and $D(Y_a \mid \mathbf{X}_a) = -\sum_j \hat{\pi}_{j(a)} \log \hat{\pi}_{j(a)}$, so

$$H = \frac{n \sum_j \hat{\pi}_j \log \hat{\pi}_j - \sum_a \sum_j \hat{\pi}_{j(a)} \log \hat{\pi}_{j(a)}}{n \sum_j \hat{\pi}_j \log \hat{\pi}_j}. \qquad (1.3)$$

The concentration and entropy measures are the two R^2-type measures that are provided, at present, when multinomial response models (Haberman 1982) are fitted using the LOGLINEAR routine in SPSS^X.

These measures C and H share the properties $0 \le (C, H) \le 1$, with $(C, H) = 0$ equivalent to $\{\hat{\pi}_{j(a)} = \hat{\pi}_j,\ a = 1, \ldots, n, j = 1, \ldots, r\}$ and $(C, H) = 1$ equivalent to {for each a, $\hat{\pi}_{j(a)} = 1$ for some j}. The lower bound occurs for models in which Y is independent of $\mathbf{X}$, and the upper bound occurs when the model suggests that Y can be perfectly predicted using $\mathbf{X}$. When the $\{\hat{\pi}_{j(a)}\}$ are obtained by maximum likelihood for a multinomial response model, Haberman (1982) noted that H is necessarily nondecreasing as the model is generalized (e.g., as additional predictors are used in $\mathbf{X}$). Although C is not necessarily nondecreasing, in practice it seems to behave much like H.

Consider the special case in which the explanatory variables are all categorical. Let $\{\pi_{ij}\}$, satisfying $\sum_i \sum_j \pi_{ij} = 1$, denote cell probabilities in an $s \times r$ contingency table in which the r columns are the levels of Y and the s rows are the combinations of levels of $\mathbf{X}$; that is, if the ath observation is in row i, then $\pi_{j(a)} = \pi_{ij}/\pi_{i+}$ $(j = 1, \ldots, r)$. Let $\{p_{ij}\}$ denote corresponding sample proportions, and let $\{\hat{\pi}_{ij}\}$ denote estimates of $\{\pi_{ij}\}$ based on fitting some model to the sample. In this case,

$$C = \frac{\sum_i (p_{i+}/\hat{\pi}_{i+}) \sum_j \hat{\pi}_{ij}^2/\hat{\pi}_{i+} - \sum_j \hat{\pi}_{+j}^2}{1 - \sum_j \hat{\pi}_{+j}^2} \qquad (1.4)$$

and

$$H=\frac{\sum_{j}\hat{\pi}_{+j}\log\hat{\pi}_{+j}-\sum_{i}(p_{i+}/\hat{\pi}_{i+})\sum_{j}\hat{\pi}_{ij}\log(\hat{\pi}_{ij}/\hat{\pi}_{i+})}{\sum_{j}\hat{\pi}_{+j}\log\hat{\pi}_{+j}}. \tag{1.5}$$

Most models that are specifically designed to treat Y as a response variable (such as logit models) satisfy the constraints $\{\hat{\pi}_{i+}=p_{i+}\}$. In this case these measures simplify to

$$C=\frac{\sum_{i}\sum_{j}\hat{\pi}_{ij}^{2}/\hat{\pi}_{i+}-\sum_{j}\hat{\pi}_{+j}^{2}}{1-\sum_{j}\hat{\pi}_{+j}^{2}} \tag{1.6}$$

and

$$H=-\frac{\sum_{i}\sum_{j}\hat{\pi}_{ij}\log(\hat{\pi}_{ij}/\hat{\pi}_{i+}\hat{\pi}_{+j})}{\sum_{j}\hat{\pi}_{+j}\log\hat{\pi}_{+j}}. \tag{1.7}$$

This concentration measure generalizes Goodman and Kruskal's (1954) tau measure, which is C applied to the estimates $\{\hat{\pi}_{ij}=p_{ij}\}$ obtained with the saturated model [also see related papers by Efron (1978), Gray and Williams (1975), and Margolin and Light (1974)]. This entropy measure was suggested by Theil (1970) for the $\{p_{ij}\}$.

The measures C and H are most appropriate for a fixed set of nominal response categories. When the response variable has several possible categorizations, these measures tend to take smaller values as the number of categories increases. For instance, the dispersion measure for C gives the probability that two independent observations occur in different categories. It is not surprising to have this probability tend to 1.0 for both the conditional and marginal distributions as finer measurement is used, in which case $C\rightarrow 0$. In addition, the dispersion functions for C and H are invariant to the ordering of response categories, and alternative dispersion measures may be more appropriate when the categories are ordered. In the next section I give an adaptation of the R^2 measure for regression models that is often better suited than C or H for ordinal response variables, especially when the response categorization is rather arbitrary.

2. A REDUCTION IN VARIANCE MEASURE

Suppose now that the response variable Y is ordinal. If one can assume an underlying continuous distribution, then it may be reasonable to use a variance expression in the proportional reduction in dispersion measure. This is especially appealing for the many log-linear models for ordinal variables that require the assignment of scores to the levels of Y (e.g., see Agresti 1984, chap. 5).

Let $\{v_j\}$ be scores that satisfy $v_1<v_2<\cdots<v_r$, let $\hat{\mu}=\sum_j v_j\hat{\pi}_j$, and let $\hat{\mu}_{(a)}=\sum_j v_j\hat{\pi}_{j(a)}$. Let Y_a denote the score on the ordinal response for the ath observation in the sample; that is, $Y_a=v_j$ if the ath observation falls in the jth response category. Then letting

$$D(Y_a)=(Y_a-\hat{\mu})^2,\quad D(Y_a\,|\,\mathbf{X}_a)=(Y_a-\hat{\mu}_{(a)})^2,$$

we obtain the *proportional reduction in variance* measure

$$\hat{\eta}=\frac{\sum_{a}(Y_a-\hat{\mu})^2-\sum_{a}(Y_a-\hat{\mu}_{(a)})^2}{\sum_{a}(Y_a-\hat{\mu})^2}. \tag{2.1}$$

The value $\hat{\eta}=0$ occurs if $\hat{\mu}_a=\hat{\mu}$ for $a=1,\ldots,n$. This happens if $\hat{\pi}_{j(a)}=\hat{\pi}_j$ for all a and j, but it can occur for other $\{\hat{\pi}_{j(a)}\}$ as well. Many models for ordinal variables imply, however, that levels of $\mathbf{X}$ are stochastically ordered with respect to Y and that the $\{\hat{\pi}_j\}$ equal the sample response proportions. For such models, $\hat{\eta}=0$ is equivalent to $\{\hat{\pi}_{j(a)}=\hat{\pi}_j\}$. The value $\hat{\eta}=1$ is equivalent to $Y_a=\hat{\mu}_{(a)}$ for each observation.

We now consider properties of $\hat{\eta}$ in detail for the case in which the explanatory variables are categorical. The measure can then be expressed as

$$\hat{\eta}=\frac{\sum_{j}(v_j-\hat{\mu})^2p_{+j}-\sum_{i}\sum_{j}(v_j-\hat{\mu}_i)^2p_{ij}}{\sum_{j}(v_j-\hat{\mu})^2p_{+j}}, \tag{2.2}$$

where $\hat{\mu}_i=\sum_j v_j\hat{\pi}_{ij}/\hat{\pi}_{i+}$. Let $M_i=\sum_j v_jp_{ij}/p_{i+}$ and let $M=\sum_j v_jp_{+j}$. For any model that satisfies $\{\hat{\mu}_i=M_i\}$,

$$\hat{\eta}=\sum_{i}(\hat{\mu}_i-\hat{\mu})^2p_{i+}\Big/\sum_{j}(v_j-\hat{\mu})^2p_{+j}. \tag{2.3}$$

In this case $\hat{\eta}$ is simply the ratio of the variation "between" levels of $\mathbf{X}$ to the "total" variation. Hence it is analogous to the correlation ratio that is used for continuous response variables, where the mean of the response variable is directly modeled. Like C and H, $\hat{\eta}$ then must fall in the range [0, 1]. For case (2.3), $\hat{\eta}$ cannot decrease as the set of explanatory variables is expanded, since the sum of squares "within" levels of $\mathbf{X}$ cannot increase.

For a given contingency table and a given set of scores, a certain class of models gives the maximum value for (2.2). For models (such as hierarchical log-linear models) that satisfy $\hat{\pi}_{+j}=p_{+j}$ for all j, $\hat{\mu}$ and $\sum_j(v_j-\hat{\mu})^2p_{+j}$ are constant, so $\hat{\eta}$ achieves its maximum value when $\sum_i\sum_j(v_j-\hat{\mu}_i)^2p_{ij}$ is minimized. For each i, $\sum_j(v_j-\hat{\mu}_i)^2p_{ij}/p_{i+}$ is minimized for $\hat{\mu}_i=\sum_j v_jp_{ij}/p_{i+}$. Hence the maximum $\hat{\eta}$ is achieved for any model that satisfies $\{\hat{\mu}_i=M_i\}$. Moreover, $\hat{\mu}=M$ when $\{\hat{\pi}_{+j}=p_{+j}\}$, so the maximum value for (2.2) is

$$\sum_{i}(M_i-M)^2p_{i+}\Big/\sum_{j}(v_j-M)^2p_{+j}. \tag{2.4}$$

Table 1. Cross-Classification of Spin Speed and Mask Dimension With Size of Contact Window in Fabrication of 3.5-μm CMOS Circuits

Mask Dimension	Spin Speed	Window Size				
		I	*II*	*III*	*IV*	*V*
1	1	30	0	0	0	0
1	2	10	3	3	7	2
1	3	6	4	6	12	0
2	1	17	5	6	2	0
2	2	7	4	7	9	3
2	3	6	0	1	3	9

Source: Based on table XII in Phadke et al. (1983), ignoring missing data.

The measure (2.4) was proposed for the observed data by Anderson and Landis (1982).

Of course, the maximum value (2.4) for $\hat{\eta}$ for a given table is achieved when the saturated model is fitted. The maximum likelihood fit for the log-linear model that has an interaction term of the form $\tau_i v_j$ between **X** and Y (plus all of the corresponding lower-order relatives) has df = $(s - 1)(r - 2)$, and it also satisfies $\{\hat{\mu}_i = M_i\}$. For instance, when there is a single explanatory variable X, Goodman's (1979) "row effects" model

$$\log \pi_{ij} = \alpha + \lambda_i^X + \lambda_j^Y + \tau_i v_j \qquad (2.5)$$

satisfies the likelihood equations $\{\hat{\pi}_{i+} = p_{i+}\}$, $\{\hat{\pi}_{+j} = p_{+j}\}$, and $\{\sum_j v_j \hat{\pi}_{ij} = \sum_j v_j p_{ij}\}$.

Another class of models for which $\hat{\eta}$ is particularly well suited is the one consisting of models for the mean of an ordinal variable having assigned response scores. The weighted least squares (WLS) solution for this class is quite simple and was presented by Bhapkar (1968); Grizzle, Starmer, and Koch (1969); and Williams and Grizzle (1972). When $r > 2$, the WLS solution for these models does not produce cell probability estimates $\{\hat{\pi}_{ij}\}$, but it does yield predicted means $\{\hat{\mu}_i\}$. Hence $\hat{\eta}$ can be calculated for these models, whereas C and H cannot. Moreover, the cumulative logit and probit models discussed by McCullagh (1980) can be regarded as mean response models for underlying logistic and normal response distributions (see Agresti 1984, pp. 153–154).

I illustrate $\hat{\eta}$ using Table 1, which is based on an experiment described in Phadke, Kackar, Speeney, and Greico (1983) for analyzing the effects of several variables on the process for forming contact windows in 3.5-μm complementary metal-oxide semiconductor (CMOS) circuits. Table 1 gives the 3 × 2 × 5 cross-classification of two of the factors, spin speed and mask dimension, with the response variable, window size. The categories for window size are ordered, with the following description (in micrometers): I—window not open or not printed, II—(0, 2.25), III—[2.25, 2.75), IV—[2.75, 3.25], V—(3.25, ∞). To describe location effects of these factors on window size, I used the row effects model (2.5) with scores $v_1 = .000$, $v_2 = 1.125$, $v_3 = 2.500$, $v_4 = 3.000$, $v_5 = 5.000$ for the response categories. In fitting this model to the entire 3 × 2 × 5 table, I adjusted the table by adding $1/r = .2$ to each cell so that maximum likelihood (ML) estimates $\{\hat{\pi}_{ij}\}$ exist. For the 3 × 5 marginal cross-classification of spin speed with window size, the likelihood-ratio goodness-of-fit chi-squared statistic equals 3.72 based on residual df = 6, and it has $\hat{\mu}_1 = .444$, $\hat{\mu}_2 = 1.925$, $\hat{\mu}_3 = 2.383$, and $\hat{\eta} = .257$. Thus there is about a one-fourth reduction in variation for this factor. For the 2 × 5 marginal cross-classification of mask dimension with window size, the model has a chi-squared statistic of 6.18 based on df = 3, and it gives $\hat{\mu}_1 = 1.173$, $\hat{\mu}_2 = 1.862$, and $\hat{\eta} = .043$. The reduction in variance is much less than with spin speed. When model (2.5) is applied to the 6 × 5 cross-classification for the interaction of both factors simultaneously with window size, $\hat{\eta} = .291$. These $\hat{\eta}$ values are identical to the ones obtained for the actual data (i.e., for the saturated model applied to the 3 × 5 and 2 × 5 marginal tables and then to the adjusted 3 × 2 × 5 table). A simpler row effects model for the 3 × 2 × 5 table that has main effects but no interaction in the effects of spin speed and mask dimension on window size gives $\hat{\eta} = .275$.

3. DEPENDENCE OF MEASURES ON RESPONSE CATEGORIES

When there are only $r = 2$ response categories, it is easily seen that $\hat{\eta}$ and C are identical for the saturated model. The measures tend to be quite different, however, for large values of r. For instance, suppose that at each level of **X** there is an underlying continuous distribution for Y, and consider a sequence of categorizations of the response with $r \to \infty$ in such a way that $\max_{i,j} \pi_{j(i)} \to 0$. Then the concentration measure converges to zero, regardless of how the conditional distributions compare to the marginal distribution of Y. The entropy measure also tends to be small for large values of r. For instance, suppose $\{\pi_i = 1/r\}$ and suppose that each conditional distribution has t probabilities equal to $1/t$ and the remaining ones equal to 0. Then $H = 1 - (\log t)/(\log r)$; and if r and $t \to \infty$ with fixed $f = t/r > 0$, we obtain $H \to 0$. For the joint (**X**, Y) distribution, denote the conditional variance by $\sigma^2_{Y|\mathbf{X}}$ and the marginal variance of Y by σ^2_Y. If the cutpoints for each categorization are evenly spaced and if equal-interval scores are assigned to the responses, then the population value η of (2.1) converges to the correlation ratio $(\sigma^2_Y - E\sigma^2_{Y|\mathbf{X}})/\sigma^2_Y$ for the underlying distribution.

The dependence of C and H on the response categorization can be illustrated using Table 1. The values of the measures for the adjusted sample counts in that 6 × 5 table are $C = .178$, $H = .199$, and $\hat{\eta} = .291$. If the table is collapsed to a 6 × 2 table by

Table 2. Measures of Association Computed for 2 × 2 × r Tables, With Underlying Trivariate Normal Distribution

			r				
Measure	ρ_{X_1Y}	ρ_{X_2Y}	*2*	*3*	*4*	*5*	*10*
C	.4	.0	.069	.043	.027	.019	.009
	.4	.4	.137	.090	.059	.041	.019
	.8	.0	.348	.208	.126	.092	.043
	.8	.4	.417	.278	.181	.135	.063
H	.4	.0	.050	.039	.035	.032	.024
	.4	.4	.104	.082	.073	.068	.050
	.8	.0	.268	.207	.177	.161	.117
	.8	.4	.349	.276	.238	.218	.161
η	.4	.0	.069	.084	.092	.095	.100
	.4	.4	.137	.168	.184	.191	.200
	.8	.0	.348	.402	.411	.404	.417
	.8	.4	.417	.486	.503	.499	.518

combining responses II–V, we obtain $C = .297$, $H = .252$, and $\hat{\eta} = .297$.

Table 2 also illustrates the behavior of the C, H, and η measures in terms of the categorization of the response. The three measures were calculated for a set (X_1, X_2, Y) of continuous variables categorized in tables of sizes $2 \times 2 \times r$ with $r = 2, 3, 4, 5, 10$. (For the formulas in Sections 1 and 2, this is treated as a $4 \times r$ table). The cell proportions $\{\pi_{ijk}\}$ corresponded to an underlying trivariate normal distribution with correlations $\rho_{X_1X_2} = 0$ and (a) $\rho_{X_1Y} = .4$, $\rho_{X_2Y} = 0$; (b) $\rho_{X_1Y} = .4$, $\rho_{X_2Y} = .4$; (c) $\rho_{X_1Y} = .8$, $\rho_{X_2Y} = .0$; and (d) $\rho_{X_1Y} = .8$, $\rho_{X_2Y} = .4$. The cut points for forming the tables were chosen at the means for the marginal distributions of X_1 and X_2. For the marginal $N(\mu_Y, \sigma_Y^2)$ distribution of Y, they were chosen at μ_Y for $r = 2$, at $\mu_Y \pm .4\sigma_Y$ for $r = 3$, at μ_Y and $\mu_Y \pm .8\sigma_Y$ for $r = 4$, at $\mu_Y \pm .4\sigma_Y$ and $\mu_Y \pm 1.2\sigma_Y$ for $r = 5$, and at μ_Y, $\mu_Y \pm .4\sigma_Y$, $\mu_Y \pm .8\sigma_Y$, $\mu_Y \pm 1.2\sigma_Y$, $\mu_Y \pm 1.6\sigma_Y$ for $r = 10$. The measure values reported in the table were computed for the saturated model. Hence for these cases the η measure is equivalent to the R^2 measure of Anderson and Landis (1982).

As r increases, the concentration measure decreases dramatically from its initial value toward its limiting value of zero. The entropy measure is somewhat more stable. Its values are also very small, however, when r is large, the values at $r = 10$ being less than half the size as when $r = 2$. For the proportional reduction in variance measure, the values at $r = 10$ are about 25%–45% higher than at $r = 2$. Like other correlation measures, $\hat{\eta}$ tends to be attenuated by grouping. A Sheppard correction can be made to the denominator variance to reduce the dependence on r. In terms of relative size, however, the variation in values is not nearly so great with η as with C and H. Most important, meaningful limiting values can be obtained with η but are generally not obtained with C or H, as $r \to \infty$. If the **X** categorization were also suitably refined, η would converge in the limit to $\rho_{X_1Y}^2 + \rho_{X_2Y}^2$.

4. ASYMPTOTIC VARIANCES

Next I present asymptotic variance formulas for $\hat{\eta}$, C, and H, for the important case in which the explanatory variables are categorical and the model satisfies $\{\hat{\pi}_{i+} = p_{i+}\}$. Let $\boldsymbol{\pi}$, $\hat{\boldsymbol{\pi}}$, and $\mathbf{p}$ denote the $\{\pi_{ij}\}$, $\{\hat{\pi}_{ij}\}$, and $\{p_{ij}\}$ expressed in column vector form. Let C_π, H_π, and η denote population values of C, H, and $\hat{\eta}$. Assuming the model holds, these have the same form as the sample expressions with $\hat{\boldsymbol{\pi}}$ (and $\mathbf{p}$ in $\hat{\eta}$) replaced by $\boldsymbol{\pi}$. Now $\sqrt{n}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \xrightarrow{d} N(\mathbf{0}, \sum)$, where $\sum$ is given in Bishop, Fienberg, and Holland (1975, p. 512, eq. 14.8-19) for arbitrary model form and is given in Fienberg (1980, p. 170) for the special case of log-linear models. For the saturated model, $\hat{\boldsymbol{\pi}} = \mathbf{p}$, so $\sum = \text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'$.

In (1.6) and (1.7), C and H are simple functions of $\hat{\boldsymbol{\pi}}$, so they are asymptotically normally distributed by the delta method. For C let

$$v = \sum_i \sum_j \pi_{ij}^2/\pi_{i+} - \sum_j \pi_{+j}^2$$

$$\delta = 1 - \sum_j \pi_{+j}^2$$

$$v'_{kl} = \frac{\partial v}{\partial \pi_{kl}} = \frac{2\pi_{kl}\pi_{k+} - \sum_j \pi_{kj}^2}{\pi_{k+}^2} - 2\pi_{+l}$$

$$\delta'_{kl} = \partial\delta/\partial\pi_{kl} = -2\pi_{+l},$$

and let $\mathbf{d}'_C = (d_{11}, \ldots, d_{sr})$ with $d_{kl} = (\delta v'_{kl} - v\delta'_{kl})/\delta^2$. Then if $0 < C_\pi < 1$, $\sqrt{n}(C - C_\pi) \xrightarrow{d} N(0, \sigma_C^2)$ with $\sigma_C^2 = \mathbf{d}'_C \sum \mathbf{d}_C$. For H let

$$v = \sum_i \sum_j \pi_{ij} \log(\pi_{i+}\pi_{+j}/\pi_{ij}), \qquad \delta = \sum_j \pi_{+j} \log \pi_{+j},$$

$$v'_{kl} = 1 + \log(\pi_{k+}\pi_{+l}/\pi_{kl}), \qquad \delta'_{kl} = 1 + \log \pi_{+l},$$

and let $\mathbf{d}'_H = (d_{11}, \ldots, d_{sr})$ with $d_{kl} = (\delta v'_{kl} - v\delta'_{kl})/\delta^2$. Then if $0 < H_\pi < 1$, $\sqrt{n}(H - H_\pi) \xrightarrow{d} N(0, \sigma_H^2)$ with $\sigma_H^2 = \mathbf{d}'_H \sum \mathbf{d}_H$.

The formulas for σ_C^2 and σ_H^2 are special cases of formulas given by Haberman (1982). He gave formulas that also apply if **X** is partly or wholly continuous. Our reason for treating the fully categorical case separately here is that the formulas are somewhat easier to use.

The measure $\hat{\eta}$ in (2.2) is a function of both $\hat{\boldsymbol{\pi}}$ and **p**. If the model truly holds, these jointly satisfy

$$\sqrt{n}\left[\begin{pmatrix}\mathbf{p}\\ \hat{\boldsymbol{\pi}}\end{pmatrix}-\begin{pmatrix}\boldsymbol{\pi}\\ \boldsymbol{\pi}\end{pmatrix}\right] \xrightarrow{d} N(\mathbf{0}, V),$$

where V is given in Bishop et al. (1975, p. 517, eq. 14.9-26). Now let

$$\hat{v} = \sum_j (v_j - \hat{\mu})^2 p_{+j} - \sum_i \sum_j (v_j - \hat{\mu}_i)^2 p_{ij}$$

$$\delta = \sum_j (v_j - \hat{\mu})^2 p_{+j}$$

$$\begin{aligned}\hat{v}'_{kl1} &= \partial\hat{v}/\partial\hat{\pi}_{kl}\\ &= -2v_l(M - \hat{\mu}) + 2(v_l - \hat{\mu}_k)(p_{k+}/\hat{\pi}_{k+})(M_k - \hat{\mu}_k)\\ \hat{v}'_{kl2} &= \partial\hat{v}/\partial p_{kl} = (v_l - \hat{\mu})^2 - (v_l - \hat{\mu}_k)^2\\ \hat{\delta}'_{kl1} &= \partial\hat{\delta}/\partial\hat{\pi}_{kl} = -2v_l(M - \hat{\mu})\\ \hat{\delta}'_{kl2} &= \partial\hat{\delta}/\partial p_{kl} = (v_l - \hat{\mu})^2.\end{aligned}$$

Then if $d_{klm} = (\delta v'_{klm} - v\delta'_{klm})/\delta^2$ (all k, l, m) and if $\mathbf{d}' = (d_{111}, \ldots, d_{sr2})$, we have $\sqrt{n}(\hat{\eta} - \eta) \xrightarrow{d} N(0, \sigma^2_{\hat{\eta}})$ with $\sigma^2_{\hat{\eta}} = \mathbf{d}'V\mathbf{d}$.

If a model is fitted that satisfies $M = \hat{\mu}$ and $M_i = \hat{\mu}_i$ (all i), then $\hat{v}'_{kl1} = \hat{\delta}'_{kl1} = 0$ in the expression for the asymptotic variance of $\hat{\eta}$. In that case, the value of $\hat{\eta}$ matches that for the saturated model, and the asymptotic variance simplifies to $\sigma^2_{\hat{\eta}} = \sum\sum \pi_{kl} d^2_{kl}$ with

$$d_{kl} = \frac{(v_l - \mu_k)^2 - (v_l - \mu)^2(1 - \eta)}{\sum_j (v_j - \mu)^2 \pi_{+j}}.$$

In particular, this is the asymptotic variance of the Anderson and Landis (1982) R^2 measure.

The asymptotic standard errors can be estimated by substituting the model estimates $\{\hat{\pi}_{ij}\}$ for $\{\pi_{ij}\}$. For instance, the estimated standard error of $\hat{\eta} = .291$ for the row effects model fitted to Table 1 is $\hat{\sigma}_{\hat{\eta}}/\sqrt{n} = .683/(162)^{1/2} = .054$. An approximate 95% confidence interval for η is $.291 \mp 1.96(.054)$, or (.186, .396). Similarly, the estimated standard error of $C = .139$ is $\hat{\sigma}_C/\sqrt{n} = .026$, and the estimated standard error of $H = .139$ is $\hat{\sigma}_H/\sqrt{n} = .027$.

In describing the R^2-type measures in Sections 1 and 2 and giving their standard errors, I have assumed full multinomial sampling, since such measures are usually of less interest when one variable is fixed. If it were more reasonable to assume independent multinomial sampling at each level of **X** (perhaps the case for Table 1), one might also use a fixed population distribution for **X** (rather than the observed one) in defining the measures. These alternative measures and their standard errors can be formulated using the methodology discussed in Goodman and Kruskal (1972).

5. CHOICE OF SCORES

One disadvantage of the $\hat{\eta}$ measure is the necessity of assigning scores to the response categories. It is often not obvious how to assign distances between categories of an ordinal variable. If $\hat{\eta}$ is used for a model that requires assigning scores to the response categories, then normally one would use the same scores for $\hat{\eta}$ as are used in the model. For the models proposed by Goodman (1979), including the row effects model (2.5), the simplest descriptions of model parameters occur for equal-interval scores. Unless the particular classification suggests a more natural scoring (e.g., as in Table 1), I suggest these scores as a "default" choice. In any case, the researcher should try a few "reasonable" choices to determine the dependence of the value of $\hat{\eta}$ on that choice. For the adjusted Table 1, for instance, the equal-interval scoring also gives $\hat{\eta} = .291$, and another "reasonable" choice (0, 2.0, 2.5, 3.0, 3.5) gives $\hat{\eta} = .304$.

Alternatively, one could use the data to generate scores. For instance, one might choose the ridits for the marginal distribution of Y as scores; that is,

$$v_j = p_{+1} + \cdots + p_{+j-1} + p_{+j}/2, \qquad j = 1, \ldots, r.$$

In this case $\hat{\eta}$ is a natural summary measure for the models for mean ridits discussed by Williams and Grizzle (1972) and by Semenya, Koch, Stokes, and Forthofer (1983). An advantage of ridit scores is that they directly take into account the way the response is categorized. For instance, if two adjacent categories are combined, then the new ridit score is between the original two; the other ridit scores are unaffected.

As another possibility, one might choose a model that contains parameters that can be interpreted as scores. In particular, Goodman's (1979) multiplicative row and column effects model can be regarded as a generalization of the row effects model in which scores are estimated that produce the best fit of that model. Goodman (1981) showed that such scores are similar in many ways to scores that would be obtained in a canonical correlation analysis. The asymptotic variance given for $\hat{\eta}$ in Section 4 must be derived separately for cases in which the scores are generated by the data, since the $\{v_j\}$ are then random rather than fixed. Although it is difficult to make general remarks about the effect of the choice of scores, in my experience substantive interpretations have not depended on that choice.

ACKNOWLEDGMENT

I am very grateful to Ming Yang for a large amount of computing assistance and to Christy Chuang for some helpful comments. This research was partially supported by National Institutes of Health Grant R01 GM33210-01.

[*Received September 1984. Revised October 1985.*]

REFERENCES

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

Anderson, R. J., and Landis, J. R. (1982), "CATANOVA for Multidimensional Contingency Tables: Ordinal-Scale Response," *Communications in Statistics—Part A*, 11, 257–270.

Bhapkar, V. P. (1968), "On the Analysis of Contingency Tables With a Quantitative Response," *Biometrics*, 24, 329–338.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis*, Cambridge, MA: MIT Press.

Efron, B. (1978), "Regression and ANOVA With Zero-One Data: Measures of Residual Variation," *Journal of the American Statistical Association*, 73, 113–121.

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

Goodman, L. A. (1979), "Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories," *Journal of the American Statistical Association*, 74, 537–552.

——— (1981), "Association Models and Canonical Correlation in the Analysis of Cross-Classifications Having Ordered Categories," *Journal of the American Statistical Association*, 76, 320–324.

Goodman, L. A., and Kruskal, W. H. (1954), "Measures of Association for Cross-Classifications," *Journal of the American Statistical Association*, 49, 732–764.

——— (1972), "Measures of Association for Cross-Classifications IV: Simplification of Asymptotic Variances," *Journal of the American Statistical Association*, 67, 415–421.

Gray, L. N., and Williams, J. S. (1975), "Goodman and Kruskal's Tau *b*: Multiple and Partial Analogs," in *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 444–448.

Grizzle, J. E., Starmer, C. F., and Koch, G. G. (1969), "Analysis of Categorical Data by Linear Models," *Biometrics*, 25, 489–504.

Haberman, S. J. (1982), "Analysis of Dispersion of Multinomial Responses," *Journal of the American Statistical Association*, 77, 568–580.

Magidson, J. (1982), "Qualitative Variance, Entropy, and Correlation Ratios for Nominal Dependent Variables," *Social Science Research*, 10, 177–194.

Margolin, B. H., and Light, R. J. (1974), "An Analysis of Variance for Categorical Data, II: Small-Sample Comparisons With Chi Square and Other Competitors," *Journal of the American Statistical Association*, 69, 755–764.

McCullagh, P. (1980), "Regression Models for Ordinal Data," *Journal of the Royal Statistical Society*, Ser. B, 42, 109–142.

Phadke, M. S., Kackar, R. N., Speeney, D. V., and Grieco, M. J. (1983), "Off-Line Quality Control in Integrated Circuit Fabrication Using Experimental Design," *The Bell System Technical Journal*, 62, 1273–1309.

Semenya, K., Koch, G. G., Stokes, M. E., and Forthofer, R. N. (1983), "Linear Models Methods for Some Rank Function Analyses of Ordinal Categorical Data," *Communications in Statistics—Part A*, 12, 1277–1298.

Theil, H. (1970), "On the Estimation of Relationships Involving Qualitative Variables," *American Journal of Sociology*, 76, 103–154.

Williams, O. D., and Grizzle, J. E. (1972), "Analysis of Contingency Tables Having Ordered Response Categories," *Journal of the American Statistical Association*, 67, 55–63.