Statistics Canada Statistique Canada

Methodology Branch

Household Survey
Methods Division

Direction de la méthodologie

Division des méthodes
d'enquêtes des ménages

Canada

WORKING PAPER
METHODOLOGY BRANCH

DIVERSITY MEASURES OF INTERVIEWER ERROR IN CATEGORICAL SURVEY DATA

HSMD - 2001 - 007E

Takis Merkouris

Household Survey Methods Division
Statistics Canada

August 2001

Diversity measures of interviewer error in categorical survey data

Takis Merkouris[1]

ABSTRACT

This article considers the measurement of interviewer error in polytomous categorical responses in surveys. Treating this error as extra-multinomial variation in the categorical responses, a statistical analysis of measurement error is formulated in a factor-response framework that is analogous to the random effects analysis of variance for quantitative data. Unlike current methodologies based on standard ANOVA assumptions, the proposed methodology is based on a measure of diversity for categorical variables that allows valid inference procedures for interviewer errors in polytomous responses. An extension of the methodology to mixed effects can incorporate effects of factors other than interviewers, such as interviewing modes and question types, on hierarchical or cross-classified polytomous categorical survey data.

Key Words: Diversity; random effects; intraclass correlation; interviewer variance; Gini-Simpson index.

[1]Takis Merkouris, Household Surveys Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6

Des mesures de diversité de l'erreur due à l'intervieweur dans les données d'enquête catégoriques

Takis Merkouris[2]

RÉSUMÉ

Dans cet article on considère la détermination de l'erreur due à l'intervieweur dans les réponses à catégories polytomiques dans les enquêtes. En traitant cette erreur comme une variation extra multinomiale dans les réponses à catégories, on peut formuler une analyse statistique de l'erreur de mesurage dans une optique de facteur-réponse analogue à l'analyse des effets aléatoires sur la variance de données quantitatives. Contrairement aux méthodologies actuelles basées sur les suppositions ANOVA classiques, la méthodologie proposée est basée sur une mesure de la diversité pour les variables catégoriques qui permet des procédures d'inférence valides pour les erreurs dues à l'intervieweur dans les réponses polytomiques. Une généralisation de la méthodologie aux effets mixtes peut incorporer les effets de facteurs autres que les intervieweurs, tels le mode de collecte et le type de question, sur les données d'enquête hiérarchiques ou à catégories polytomiques croisées.

Mots clés: Diversité; effets aléatoires; corrélation intraclasse; variance due à l'intervieweur; indexe de Gini-Simpson.

---

[2]Takis Merkouris, Division des méthodes des enquêtes auprès des ménages, Statistique Canada, Ottawa (Ontario) K1A 0T6

1 Introduction

In this article, the measurement error introduced by interviewers into categorical survey data is considered. For the analysis of the interviewer error the existing methodology typically treats the categories of a polytomous categorical response as dichotomous responses. The interviewer error is then considered as the random component in a random effects (components of variance) model, which is directly analogous to the standard random effects ANOVA for quantitative variables; for a review of this modeling approach, see Biemer and Stokes [2]. The two basic quantities describing the random interviewer effect, namely interviewer variance and intraclass (intra-interviewer) correlation coefficient, are defined through this simple random effects model. Mixed models, with fixed effects due to sources of measurement error other than interviewers, have also been considered in various studies of interviewer error. For such models, however, standard ANOVA inference procedures for the interviewer variance and the intraclass correlation coefficient, based on normality assumptions and constant variance components, are not valid for dichotomous variables.

The problems with using standard ANOVA techniques for dichotomous variables has led investigators to consider alternative methods of inference for the interviewer effect. The existing methodology includes likelihood-based inference assuming a beta-binomial distribution for the random interviewer effect (Pannekoek [6]), quasi-likelihood inference procedures (Pannekoek [6]), and likelihood methods based on modeling the distribution of a continuous latent variable underlying the categorical response (Stokes [7]). These methods are based on distributional assumptions that may be unsubstantiated, and are computationally complicated. Moreover, whereas the existing methodology analyzes polytomous categorical responses component-wise, i.e., a separate intraclass correlation coefficient is computed for each category treated as dichotomous, an overall measure of intraclass correlation that would effectively combine these separate measures is of considerable interest.

In this article, the statistical analysis of interviewer effect on categorical survey responses is formulated in the general factor-response framework. In this framework, a general analysis of variation for categorical data known as analysis of diversity (ANODIV), Rao [10], analogous to the ANOVA for quantitative data, is used to determine the effect of the levels of the factor interviewer on a polytomous categorical response variable.

The formulation involves an extension of the ANODIV to random and mixed effects factors, which is suitable for hierarchical and cross classifications of categorical survey data. In this mixed effects set-up the explained diversity is partitioned by source, namely the random interviewer factor and a fixed factor which could be the interviewing mode, the interviewer shift or the question type. Unlike the standard ANOVA applied to dichotomous variables, the proposed formulation defines overall intraclass correlation coefficients, over the categories at given level of the fixed factor, as well as over the levels of the fixed factor. The intraclass correlation coefficient can then be used effectively as a single summary of interviewer effects on hierarchical or cross classified polytomous categorical survey data, or as a comparative measure of interviewer effects across fixed factor levels.

Much has appeared in the literature about the measurement of interviewer error, as well as about its effect on descriptive statistics or on methods of statistical inference; for a review of the various aspects of the study of interviewer error see Biemer and Trewin [3]. The attention in this article is confined to the measurement of the interviewer error in polytomous categorical responses. It is in this restricted sense that the term interviewer effect will be used here, i.e., to denote the component of variation due to interviewers in the survey responses.

In Section 2, we provide a background review of the problem and motivation for the proposed methodology. In Section 3, we formulate an analysis of diversity components, and define parameters of interest. In Section 4, we derive estimates of the various parameters. Concluding remarks are made in Section 5.

2 Background Review

There exists a considerable amount of work in the sample survey literature concerning the modeling of the interviewer error; see Biemer and Stokes [2] and references therein. Since interviewers are typically thought of as being randomly chosen from a conceivably infinite population of interviewers, the interviewer error is considered as random measurement error. Traditionally, polytomous categorical survey responses have been analyzed component-wise, with each category treated as dichotomous for analysis. Then the interviewer error is considered as the random component in a random effects model that is directly analogous to the standard random effects ANOVA for quantitative variables; see, for example, Stokes and Mulry [9], Pannekoek [5].

Thus, in a random sample of n units let y_{ij} denote the response of the j-th unit interviewed by the i-th interviewer, where $i = 1, \ldots, I$; $j = 1, \ldots, n_i$. For a specific category of a polytomous categorical response, y_{ij} is a dichotomous response with the value of 1 if the unit j is classified into the category, and the value of 0 otherwise. For a given interviewer i, the y_{ij}'s are postulated as independent Bernoulli random variables with the same conditional probability $\pi_i = P(y_{ij} = 1|i) = E(y_{ij}|i)$ that interviewer i records a randomly chosen unit as belonging to the category. Since it is typically assumed that the interviewers are randomly selected from a infinite population of interviewers, the probabilities $\pi_i, i = 1, \ldots, I$, are assumed to be realizations of a random variable π, say, with $E(\pi) = \mu$ and $Var(\pi) = \sigma_\pi^2$.

The one-way random effects model applied to the dichotomous responses to describe the effect of the interviewer error is given by

$$y_{ij} = \mu + (\pi_i - \mu) + \epsilon_{ij}, \quad i = 1, ..., I; \quad j = 1, ..., n_i, \tag{1}$$

where $\pi_i - \mu$ is the random effect of the i-th interviewer on the responses. Note that since the possible values of y_{ij} are either 1 or 0, the values of ϵ_{ij} are restricted to $-\pi_i$ and $1 - \pi_i$, with conditional probabilities $1 - \pi_i$ and π_i respectively, for each i. The variance components for the model (1) are derived from the variance decomposition

$$\begin{aligned} Var(y_{ij}) &= EVar(y_{ij}|i) + VarE(y_{ij}|i) \\ &= E[(\pi_i(1 - \pi_i)] + Var(\pi_i). \end{aligned} \tag{2}$$

A more explicit representation of (2) in terms of the model parameters is

$$\mu(1 - \mu) = E[(\pi_i(1 - \pi_i)] + \sigma_\pi^2. \tag{3}$$

The variance component σ_π^2 is called interviewer variance, and represents what is called extra-multinomial variation. Note that $Cov(y_{ij}, y_{il}) = Var(\pi) = \sigma_\pi^2$, where y_{ij}, y_{il} are observations on respondents j and l recorded by the same interviewer i. Then, the intra-interviewer (intraclass) correlation is defined by

$$\rho = Cov(y_{ij}, y_{il})/Var(y_{ij}) = \sigma_\pi^2/\mu(1-\mu).$$

The intraclass correlation coefficient ρ is a measure of the relative magnitude of the interviewer variance. For a good discussion on the importance of ρ in survey data analysis see Stokes [7], and Biemer and Trewin [3].

The standard variance components estimation procedure gives unbiased estimates for all variance components and, for large number of interviewers, a consistent estimate of ρ. The estimation of σ_π^2 requires a randomization scheme known in survey methodology as interpenetration. In its most basic form, interpenetration consists of randomly dividing the sample into a number of subsamples (possibly of varying sizes), and assigning each subsample to a single interviewer. In view of this, the conditioning in (2) is on the random assignment of the n_i respondents to interviewer i as well as on the randomly selected interviewer i.

In estimating the variance components, the dichotomous nature of the responses is ignored and the analysis is carried out using the ANOVA method. It is not valid, however, to assume normality for either π_i or ϵ_{ij}. Furthermore, the variance components $Var(y_{ij}|i)$ are not constant across interviewers. Therefore, standard ANOVA methods of interval estimation and testing for σ_π^2 and ρ are not appropriate for dichotomous variables. For this reason, investigators have proposed alternative methods of inference regarding the interviewer variance and the intraclass correlation. Stokes and Hill [8], and Pannekoek [5] assume a beta distribution for the random effect π_i, and carry out likelihood based inference for σ_π^2 and ρ.

Situations in which, in addition to the random interviewer effect, a fixed factor might affect the responses can be described by mixed effect models. In such mixed models, it is assumed that each interviewer is randomly selected from a different population identified with a level of a fixed factor, so that for level h, say, the π_{hi}'s are realizations of a random variable π_h with $E(\pi_h) = \mu_h$ and $Var(\pi_h) = \sigma_{\pi_h}^2$. The random factor may be nested within a fixed factor (such as area, or interviewing mode), or crossed with a fixed factor (such as interviewer shift or question type). For example, a nested survey design with interviewers within areas can be modeled by

$$y_{hij} = \mu + a_h + b_{h(i)} + \epsilon_{hij}, \quad h = 1, \ldots, H, \tag{4}$$

where $a_h = \mu_h - \mu$ is the fixed area effect, and $b_{h(i)} = \pi_{h_i} - \mu_h$ is the random effect of the i-th interviewer within area h. Then, in contrast to the usual set-up for quantitative responses, an intraclass correlation coefficient is defined at each level of the fixed factor, as $\rho_h = \sigma_{\pi_h}^2/\mu_h(1-\mu_h)$, and thus an overall measure of intraclass correlation needs to be defined to describe the intra-interviewer correlation of the responses over the levels of the fixed factor.

As in the case of the simple random effects model, ANOVA type inference procedures about the parameters of the mixed effects models, based on normality assumptions and constant variance components, are not valid for dichotomous variables. Anderson and Aitkin [1], and Stokes [7] proposed logit regression models for a hypothetical continuous latent variable that determines the values of the dichotomous response. They used likelihood based inference for the parameters of their model. Pannekoek [6] considered a generalized linear mixed model with the fixed effects measured at the interviewer level. For estimation of the parameters of his model he used two methods, maximum likelihood based on the assumption that the π_{hi}'s are realizations of beta distributed random variables, and the generalized estimating equation approach, which does not require such a distributional assumption.

All the methods for the analysis of interviewer effects cited above arose from the concern about the propriety of applying standard ANOVA methodology to dichotomous responses. These methods, however, are themselves problematic in various ways. The assumption of beta distribution for the random effect may be unsubstantiated. The distributional assumptions of Anderson and Aitkin [1] and Stokes [7] are open to the same criticism. Regarding mixed models, the beta binomial approach is limited in its scope of application to the nested model (4). The latent variable method uses a parameter measuring interviewer variability that is not interviewer variance. Furthermore, all these methods, particularly the latent variable method, are computationally quite complex. A common characteristic of the existing methodology for measuring interviewer effects is that a separate measure of interviewer variance is determined for each category of a polytomous response. This approach, born of analytical convenience, has the disadvantage of not defining a single measure of intraclass correlation that would provide a convenient summary of interviewer effect on a polytomous response. In mixed effects situations, overall intraclass correlation coefficients defined over the categories at a given level of the fixed factor, as well as over the levels of the fixed factor, would be of considerable interest. They could be used effectively as single summary of interviewer effects on hierarchical or cross classified polytomous categorical survey data, or as comparative measure of interviewer effects across fixed factor levels. For an exemplification of the need of such overall measures see studies of interviewer effects in Pannekoek [5, 6].

3 Analysis of Diversity Components

The analysis of interviewer effects on categorical survey data can be formulated in a general factor- response framework. In this framework, the interviewer effect is defined as the association of a categorical response with the explanatory categorical variable "interviewer". This approach is based on a proper measure of variation (diversity) for categorical variables.

Formally then, let Y be a categorical response variable with multinomial probability vector $\mu = (\mu_1, \ldots, \mu_r)$. Define a measure of difference $d(y, y')$ between Y and another categorical response Y' independent of Y, but identically distributed, to be zero if they agree and one otherwise. Then, as in Rao [10], a measure of variation D, called diversity, for the categorical variable Y is defined as

$$D(Y) = Ed(y, y') = \sum \mu_k(1 - \mu_k). \tag{5}$$

The diversity D is interpreted as the probability that two independent categorical responses with the same probability vector are different. A useful representation of D is $D(Y) = \mu'\Delta\mu$, where Δ is the matrix of differences between categories of Y. For nominal responses, as in the present context, the diagonal entries of Δ are zeros and the rest are ones. It is clear from (5) that $D(Y)$ is the sum of the diagonal entries in the covariance matrix Σ_μ, say, of the multinomial distribution with probability vector μ. Thus, we can also write $D(Y) = tr\Sigma_\mu$. In the special case of a dichotomous response, $D(Y)$ reduces to the variance of a binomial variable, as in (3). The diversity measure defined by (5) is known as the Gini-Simpson index.

Now, let a categorical explanatory variable X, with I levels, represent the factor interviewer. Also let $\pi_i = (\pi_{i1}, \dots, \pi_{ir})'$ be the conditional probability vector of the response Y at the i-th level of X. Then, the conditional diversity of Y at the i-th level of X is $D(Y|\pi_i) = \pi_i'\Delta\pi_i = tr\Sigma_{\pi_i}$, where Σ_{π_i} is the covariance matrix of a multinomial distribution with fixed probability vector π_i. As indicated in the previous section, the probability vectors π_i, $i = 1, \dots, I$, associated with the levels of the random factor interviewer, are realizations of a random vector π, with expectation $E(\pi) = \mu$ and covariance matrix Σ_π. By relating the average conditional diversity $ED(Y|\pi)$ to the unconditional diversity $D(Y)$ it can be readily shown that

$$\mu'\Delta\mu = E(\pi'\Delta\pi) - E(\pi - \mu)'\Delta(\pi - \mu). \tag{6}$$

For the matrix Δ considered in this paper, we have $-E(\pi-\mu)'\Delta(\pi-\mu) = E(\pi-\mu)'(\pi-\mu)$. Noting further that $E(\pi - \mu)'(\pi - \mu)$ can be written as the expected half of the square Euclidean distance $(1/2)E(\pi_i - \pi_{i'})'(\pi_i - \pi_{i'})$ between any two realizations of the probability vector π, a diversity measure of π can be defined in analogy with (5) as $D(\pi) = E(\pi - \mu)'(\pi - \mu) = tr\Sigma_\pi$. It is to be noted that for a more general matrix Δ, as in the case of ordinal responses, we have $D(\pi) = -tr\Delta\Sigma_\pi$, and conditions have to be imposed on the pairwise differences between categories to ensure the non negativity of the terms in (6). Now, we may formally write (6) as

$$D(Y) = ED(Y|\pi) + D(\pi). \tag{7}$$

Equation (7) provides a decomposition of the total (unconditional) diversity of Y into two additive components, the average conditional diversity of Y given X, and the (extra-multinomial) diversity due to X. This decomposition is analogous to the usual variance decomposition for quantitative variables. It provides an analysis of diversity components for one-way classified categorical data (two-way contingency table), analogous to the one-way random effects ANOVA for quantitative data, as in (2).

Note that the formulation presented in this section is independent of quantifying the response variable Y, and of specifying a model for it. It is clear from (6), that the analysis of diversity components is entirely in terms of relating the unconditional and the conditional probability vectors of the variable Y to each other. Yet, for the matrix Δ defined above an interesting connection exists between the analysis of diversity components defined through (7) and the one-way multivariate analysis of variance (MANOVA) for a component of variance (or random effect) model. With quantification of the nominal response variable Y by an r-dimensional indicator variable (with 1 corresponding to the category of the response) a multivariate components of variance model can be set up as

$$Y = \mu + (\pi - \mu) + \epsilon, \tag{8}$$

where $\pi - \mu$ is the random effect component, and ϵ is the error vector, with its r entry values and its distribution determined by Y. For this model the MANOVA decomposition of variance components (multivariate analogue of (2)) is

$$\Sigma_\mu = \Sigma_\epsilon + \Sigma_\pi, \tag{9}$$

where $\Sigma_\epsilon = \Sigma_\mu - \Sigma_\pi$, and Σ_μ and Σ_π are as before. The MANOVA decomposition (9) can be viewed as a decomposition of a diversity measure defined as the variance-covariance functional, that is, the expected distance between two variables, say Z_1 ans Z_2, in $I\!R^r$ drawn randomly from the same population, with distance $d(Z_1, Z_2) = (Z_1 - Z_2)(Z_1 - Z_2)'$. Specifically, for this distance measure, $\Sigma = Ed(Y_1, Y_2)$ and $\Sigma_\pi = Ed(\pi_1, \pi_2)$.

The special case of analysis of diversity for a nominal response variable described in this paper corresponds to the additive variance component model (8) with the diversity measure in (5) equivalently defined as the expected Euclidean distance $E(Y_1 - Y_2)'(Y_1 - Y_2)$ between two realizations of the indicator variable Y. This leads to the decomposition $tr\Sigma = tr\Sigma_\epsilon + tr\Sigma_\pi$, instead of (9).

It is to be emphasized that the connection between MANOVA and the analysis of diversity for categorical data is restricted to the case of nominal categorical variables, with associated matrix Δ of differences among categories as described above. Unlike the rather arbitrary MANOVA variation measure for nominal variables, using the trace metric, the equivalent diversity measure $D(Y) = \mu'\Delta\mu$ is founded on the notion of quadratic entropy (Rao, [10]), which in its present special case of the Gini-Simpson index has the distinctive probabilistic interpretation noted above. Moreover, expressed only in terms of probability vectors the diversity $D(Y)$ lends itself to easy algebraic manipulation, as will be shown in the next section.

Returning to the diversity formulation, in the context of interviewer effects the first term in the right hand side of (7) is the average diversity within interviewers assignments, and the second term is the diversity between interviewers. The shortcut notation T, W and B for the total, within and between diversities will be used for convenience in Section 4. In analogy with the interviewer variance, we call $D(\pi)$ interviewer diversity. The ratio

$$\rho = \frac{D(\pi)}{D(Y)} = \frac{E(\pi - \mu)'(\pi - \mu)}{\mu'\Delta\mu} = \frac{tr\Sigma_\pi}{tr\Sigma_\mu} \tag{10}$$

is the proportion of diversity explained by the random interviewer factor X. It is a measure of association between X and Y, and defines an intraclass (intra-interviewer) correlation coefficient for the polytomous response Y. It follows from (10) that $\rho = 0$ if π is constant equal to μ, and $\rho = 1$ if $\pi_{ik} = 1$ for some k. Therefore, ρ represents the degree of homogeneity of responses within the levels of X. For each response category k of Y, treated as dichotomous, the coefficient ρ reduces to the coefficient $\rho_k = \sigma^2_{\pi_k}/\sigma^2_k$ of the previous section. Clearly then, ρ can be written as the weighted average

$$\rho = \frac{\sum \sigma_k^2 \rho_k}{\sum \sigma_k^2}$$

of the individual intra-interviewer coefficients, with weights the unconditional diversities (variances) associated with the categories of Y.

The analysis of diversity components described so far is an extension of the original analysis of diversity (ANODIV), (Rao [10]) — for the assessment of the effect of a fixed categorical factor on a categorical response — to the random effect case. To explain the development of an analysis of diversity for mixed effects, we present next the basic formulation of ANODIV.

In the fixed factor setting of ANODIV, let the conditional probability vectors π_h, $h = 1, \ldots, H$, be fixed (nonrandom). Let also λ_h, $(h = 1, \ldots, H;\ \sum \lambda_h = 1)$, be probabilities associated with the levels of the fixed factor X, so that λ_h is the probability that the response Y is at the h-th level of X. Then, the unconditional probability vector of Y is the mixture $\mu = E(\pi) = \sum \lambda_h \pi_h$. With conditional and unconditional diversities of Y as before, and with the same matrix Δ as in (6), the decomposition of the total diversity is

$$\mu'\Delta\mu = \sum \lambda_h \pi_h' \Delta \pi_h + \sum \lambda_h (\pi_h - \mu)'(\pi_h - \mu). \tag{11}$$

For a comprehensive exposition of ANODIV, see Rao [10].

We can now consider more general classifications involving fixed and random factors. For an hierarchical classification with the levels of a random factor nested within the levels of a fixed factor, the analysis of diversity is as follows. Let π_{hi} be the random probability vector associated with the h-th level of the fixed factor and the i-th level of the random factor, so that $\mu_h (= E(\pi_{hi}))$ is the probability vector associated with the h-th level of the fixed factor, and $\mu (= E(\mu_h) = \sum \lambda_h \mu_h)$ is the probability vector of Y. Let also $D(Y) = \mu'\Delta\mu$, $\Delta(Y|h) = \mu_h'\Delta\mu_h$ and $D(Y|\pi_{hi}) = \pi_{hi}'\Delta\pi_{hi}$ be the diversities within various levels of classification. Then, in view of (6) and (11), the decomposition of the total diversity of Y, in two stages, is

$$\begin{aligned} D(Y) &= \sum \lambda_h D(Y|h) + \sum \lambda_h (\mu_h - \mu)'(\mu_h - \mu) \\ &= \sum \lambda_h E D(Y|\pi_h) + \sum \lambda_h D(\pi_h) + \sum \lambda_h (\mu_h - \mu)'(\mu_h - \mu) \end{aligned} \tag{12}$$

or more explicitly

$$\begin{aligned} \mu'\Delta\mu &= \sum \lambda_h \mu_h' \Delta \mu_h + \sum \lambda_h (\mu_h - \mu)'(\mu_h - \mu) \\ &= \sum \lambda_h E(\pi_h'\Delta\pi_h) + \sum \lambda_h E(\pi_h - \mu_h)'(\pi_h - \mu_h) + \sum \lambda_h (\mu_h - \mu)'(\mu_h - \mu) \end{aligned} \tag{13}$$

In the context of interviewer effects, the diversity components in the right hand side of (12) are, in order, the average (overall) diversity within interviewer assignments, the average (over the levels of the fixed factor) diversity between interviewers, and the average diversity between the levels of the fixed factor. Here, the λ_h's are to be interpreted as the mixing proportions of respondents corresponding to the levels of the fixed factor. The main parameters of interest are the interviewer diversity within the h-th level of the fixed factor, $D(\pi_h) = E(\pi_h - \mu_h)'(\pi_h - \mu_h)$, and the overall (average) interviewer diversity $\sum \lambda_h D(\pi_h)$. We can also define useful measures of relative diversities as follows.

The proportion of total diversity explained by the fixed factor, given by

$$\rho(F) = \frac{D(Y) - \sum \lambda_h D(Y|h)}{D(Y)} = \frac{\sum \lambda_h (\mu_h - \mu)'(\mu_h - \mu)}{\mu'\Delta\mu}$$

is a measure of association between the response and the fixed factor.

The proportion of diversity explained by the random factor at the h-th level of the fixed factor, given by

$$\rho(R|F=h) = \frac{D(\pi_h)}{D(Y|h)} = \frac{E(\pi_h - \mu_h)'(\pi_h - \mu_h)}{\mu_h' \Delta \mu_h}$$

defines an intraclass correlation coefficient at the h-th level of the fixed factor. It can be used to compare interviewer effects at different levels of the fixed factor, e.g., in different areas or interviewing modes.

The overall intraclass correlation coefficient is defined by

$$\rho(R|F) = \frac{\sum \lambda_h D(\pi_h)}{\sum \lambda_h D(Y|h)} = \frac{\sum \lambda_h D(Y|h) - \sum \lambda_h ED(Y|\pi_h)}{\sum \lambda_h D(Y|h)}.$$

This is a measure of partial association between the response and the random factor after eliminating the effect of the fixed factor. It has the form of the weighted average

$$\rho(R|F) = \frac{\sum \lambda_h D(Y|h) \rho(R|F=h)}{\sum \lambda_h D(Y|h)}$$

A measure of multiple association between the response and the two factors is defined as

$$\rho(F, R) = \frac{D(Y) - \sum \lambda_h ED(Y|\pi_h)}{D(Y)}$$

4 Parameter Estimation

The various diversity components in (6) and (13) are quadratic functions of population probability vectors that can be estimated from sample observations. The resulting sample diversities can then be used to estimate the population diversities. We illustrate first this estimation procedure for the diversity components in (6), corresponding to the simple random effects ANODIV.

Let n_{ik}, $i = 1, \ldots, I$; $k = 1, \ldots, r$, denote the number of responses in the k-th category for the i-th level of X. Let also $n_i = \sum_k n_{ik}$ and $n = \sum_i n_i$. We will also use the notations $v_i = (n_{i1}, \ldots, n_{ir})'$, for the vector of frequencies in the i-th level of X; $p_i = (1/n_i)v_i$, for the vector of observed proportions in the i-th level of X; $\bar{p} = \sum_i (n_i/n) p_i = (1/n) \sum_i v_i$, for the vector of observed proportions in the combined sample. For a vector $a = (a_1, \ldots, a_r)'$, we will use Λ_a to denote the diagonal matrix with elements $a_1, \ldots, a_r$.

For estimation, we assume that the responses at different levels of X are stochastically independent, and that conditionally on the distribution of π_i the vector v_i follows multinomial law with parameters n_i and $\pi_i = (\pi_{i1}, \ldots, \pi_{ir})'$. Thus the conditional expectation and covariance matrix of p_i are π_i and $(1/n_i)\Sigma_{\pi_i}$, where $\Sigma_{\pi_i} = \Lambda_{\pi_i} - \pi_i \pi_i'$.

With the above notations, the sample versions of the diversities T, W and B are

$$\hat{T} = \bar{p}' \Delta \bar{p}\ ; \quad \hat{W} = \frac{1}{n} \sum_i n_i p_i' \Delta p_i\ ; \quad \hat{B} = -\frac{1}{n} \sum_i n_i (p_i - \bar{p})' \Delta (p_i - \bar{p}),$$

satisfying $\hat{T} = \hat{W} + \hat{B}$.

It is easy to verify that for the matrix Δ considered in this paper $tr\Delta\Lambda_{\pi_i} = 0$. Then, conditionally on the π_i's, standard calculations involving expectations of quadratic forms yield

$$E(\hat{W}) = \frac{1}{n}\sum_i (n_i - 1)\pi_i'\Delta\pi_i \tag{14}$$

$$E(\hat{B}) = \frac{1}{n^2}\sum_i (n - n_i)\pi_i'\Delta\pi_i - \frac{1}{n}\sum_i n_i(\pi_i - \bar{\pi})'\Delta(\pi_i - \bar{\pi}) \tag{15}$$

$$E(\hat{T}) = \bar{\pi}'\Delta\bar{\pi} - \frac{1}{n^2}\sum_i n_i\pi_i'\Delta\pi_i, \tag{16}$$

where $\bar{\pi} = \sum_i (n_i/n)\pi_i$. The expressions (14), (15) and (16) above are random quantities, as functions of the π_i's. Taking expectations with respect to the distribution of π on both sides of (14), (15) and (16) we get after simple matrix algebra

$$EE(\hat{W}) = \frac{n-I}{n}E(\pi'\Delta\pi) = \frac{n-I}{n}[tr\Sigma_\pi + \mu'\Delta\mu] = \frac{n-I}{n}[tr\Sigma_\mu - tr\Sigma_\pi] = \frac{n-I}{n}W,$$

$$EE(\hat{B}) = \frac{I-1}{n}[tr\Sigma_\mu - tr\Sigma_\pi] + \left[1 - \sum_i\left(\frac{n_i}{n}\right)^2\right]tr\Sigma_\pi = \frac{I-1}{n}W + \left[1 - \sum_i\left(\frac{n_i}{n}\right)^2\right]B,$$

$$EE(\hat{T}) = \frac{n-1}{n}\mu'\Delta\mu + \left[\frac{1}{n} - \sum_i\left(\frac{n_i}{n}\right)^2\right]tr\Sigma_\pi = \frac{n-1}{n}T + \left[\frac{1}{n} - \sum_i\left(\frac{n_i}{n}\right)^2\right]B.$$

Letting

$$\tilde{W} = \frac{n}{n-I}\hat{W}\ ;\quad \tilde{B} = \frac{n}{I-1}\hat{B}\ ;\quad n_\circ = \frac{1}{n(I-1)}[n^2 - \sum_i n_i^2]$$

we obtain unbiased estimators

$$\tilde{W}\ ;\quad \frac{1}{n_\circ}[\tilde{B} - \tilde{W}]\ ;\quad \frac{1}{n_\circ}[\tilde{B} + (n_\circ - 1)\tilde{W}]$$

of W, B and T, respectively. Then a consistent estimator of the intraclass correlation $\rho = B/T$ is

$$\hat{\rho} = \frac{\tilde{B} - \tilde{W}}{\tilde{B} + (n_\circ - 1)\tilde{W}}.$$

These results are analogous to those obtained by the standard ANOVA method applied to an one-way random effects model for a quantitative response variable Y. This is explained by the connection between the diversity and MANOVA components of a nominal response variable, established in the previous section. A similar connection exists between the corresponding sample quantities. Noting that the diversity components $\hat{W}$, $\hat{B}$ and $\hat{T}$ can be written as

$$\hat{W} = \frac{1}{n}\sum_i n_i p_i'\Delta p_i = \frac{1}{n}\sum_i n_i \sum_k p_{ik}(1 - p_{ik}), \tag{17}$$

$$\hat{B} = -\frac{1}{n}\sum_i n_i(p_i - \bar{p})'\Delta(p_i - \bar{p}) = \frac{1}{n}\sum_i n_i \sum_k (p_{ik} - \bar{p}_k)^2, \tag{18}$$

$$\hat{T} = \bar{p}'\Delta\bar{p} = \sum_k \bar{p}_k(1-\bar{p}_k), \tag{19}$$

it is easy to see that the traces of the MANOVA within-group, between-group and total-group sum of squares matrices corresponding to the vectors of the observed proportions p_i and $\bar{p}$ are simply n times the quantities in (17), (18) and (19), respectively. Hence, standard MANOVA carried out using the trace metric will produce the estimates for the diversity components W, B, T, and for the intraclass correlation $\rho = B/T$.

The estimation procedure described above extends readily to the mixed effect ANODIV with the hierarchical classification discussed in the previous section. A brief outline follows.

Let then n_{hik}, $h=1,\ldots,H$; $i=1,\ldots,I_h$; $k=1,\ldots,r$, denote the number of responses in the k-th category for the i-th interviewer at the h-th level of the fixed factor. Let also $n_{hi} = \sum_k n_{hik}$, $n_h = \sum_i n_{hi}$, $n = \sum_h n_h$, and $\lambda_h = n_h/n$. Then, for $v_{hi} = (n_{hi1},\ldots,n_{hir})'$, $p_{hi} = (1/n_{hi})v_{hi}$ is the vector of observed proportions in the (h,i) class of the hierarchical classification; $\bar{p}_h = \sum_i (n_{hi}/n_h)p_{hi}$ is the vector of observed proportions in the combined sample at the h-th level of the fixed factor; $\bar{p} = \sum_h \lambda_h \bar{p}_h$ is the vector of observed proportions in the combined sample over the levels of the fixed factor.

In this setting we assume that the responses are stochastically independent across the various levels of the hierarchical classification, and that conditionally on the distribution of π_{hi} the vector v_{hi} follows multinomial law with parameters n_{hi} and π_{hi}. Thus the conditional expectation and covariance matrix of p_{hi} are π_{hi} and $(1/n_{hi})\Sigma_{\pi_{hi}}$, from which the expectations and covariance matrices of $\bar{p}_h$ and $\bar{p}$ can be readily obtained. The sample versions of the diversity components in (13) are

$$\hat{T} = \bar{p}'\Delta\bar{p}\ ; \quad \hat{W} = \sum_h \lambda_h \frac{1}{n_h}\sum_i n_{hi}p'_{hi}\Delta p_{hi}$$

$$\hat{B}(R) = -\sum_h \lambda_h \frac{1}{n_h}\sum_i n_{hi}(p_{hi}-\bar{p}_h)'\Delta(p_{hi}-\bar{p}_h)\ ; \quad \hat{B}(F) = -\sum_h \lambda_h(\bar{p}_h-\bar{p})'\Delta(\bar{p}_h-\bar{p}),$$

where $\hat{B}(R)$ and $\hat{B}(F)$ denote diversity among levels of the random factor and between levels of the fixed factor, respectively. These sample diversities satisfy $\hat{T} = \hat{W} + \hat{B}(R) + \hat{B}(F)$. The derivation of unbiased estimators of the diversity components in (13) is similar, though more involved, to that described above for the simple random effects ANODIV. Thus writing $\hat{W} = \sum_h \lambda_h \hat{W}_h$, $\hat{B}(R) = \sum_h \lambda_h \hat{B}_h(R)$, and letting

$$\tilde{W}_h = \frac{n_h}{n_h - I_h}\hat{W}_h\ ; \quad \tilde{B}_h = \frac{n_h}{I_h-1}\hat{B}_h(R)\ ; \quad n_h^\circ = \frac{1}{n_h(I_h-1)}[n_h^2 - \sum_i n_{hi}^2]$$

we obtain unbiased estimators

$$\sum_h \lambda_h \tilde{W}_h\ ; \quad \sum_h \lambda_h \frac{1}{n_h^\circ}[\tilde{B}_h - \tilde{W}_h]\ ;$$

$$\hat{T} + \frac{1}{n}\sum_h \lambda_h \frac{1}{n_h^\circ}\left[[n_h - n_h^\circ(I_h-1) - n]\tilde{B}_h + [n_h^\circ(I_h-n) + (n-n_h)]\tilde{W}_h\right]\ ;$$

$$\hat{T} + \frac{1}{n}\sum_h \lambda_h \frac{1}{n_h^\circ}\left[[n_h - n_h^\circ(I_h-1)]\tilde{B}_h + [n_h^\circ(I_h-n_h)]\tilde{W}_h\right]$$

of W, $B(R)$, $B(F)$ and T, respectively. Then consistent estimators of the various measures of relative diversities defined in the previous section can be derived. In particular, a consistent estimator of the overall intraclass (intra-interviewer) correlation coefficient $\rho(R|F)$ is

$$\hat{\rho}(R|F) = \frac{\sum_h \frac{\lambda_h}{n_h^\circ}[\tilde{B}_h - \tilde{W}_h]}{\sum_h \frac{\lambda_h}{n_h^\circ}[\tilde{B}_h + (n_h^\circ - 1)\tilde{W}_h]}.$$

The estimation technique employed for the random effect ANODIV for hierarchical classification of responses is based on the familiar ANOVA method of moments. As with ANOVA, for mixed effect ANODIV with cross classification, or for more general mixed effect ANODIV situations, the various diversity components can be estimated by adopting the usual Henderson's method of fitting constants.

5 Concluding Remarks

In this article we have outlined an alternative methodology for the analysis of interviewer effects on categorical responses. The generality and computational simplicity of the ANODIV procedure makes it readily applicable to a wide range of problems involving the assessment of the effect of interviewers, or other factors defined at the interviewer level, on polytomous categorical survey data. Inferential aspects have been discussed in the context of estimation of parameters of interest. It is important to note that no distributional assumptions are made beyond the second moments. Relevant non-parametric hypothesis testing techniques may be developed based on the asymptotic behaviour of the estimated parameters. It is interesting to note that testing the hypothesis of interviewer effect amounts to a homogeneity test in a random set of multinomial distributions identified with the random levels of the interviewer factor. Since the probability vectors of these multinomial vectors are random, such a homogeneity test is equivalent to testing that the frequencies of the responses associated with the interviewers are multinomial, against a general alternative of mixture multinomials. Results on hypothesis testing will be discussed elsewhere.

References

[1] Anderson, D.A. and Aitkin, M. (1985). Variance components models with binary response: Interviewer variability. *J. R. Statist. Soc.* B, 47, 203- 210.

[2] Biemer, P. and Stokes, S.L. (1991). Approaches to the modeling of measurement error. In: Biemer, Groves, Lyberg, Mathiowetz and Sudman (eds.), *Measurement Errors in Surveys*. Wiley, 488-516.

[3] Biemer, P. and Trewin, D. (1997). A review of measurement error effects on the analysis of survey data. In: Lyberg, Biemer, Collins, Leeuw, Dippo, Schwarz, Trewin (eds.), *Survey Measurement and Process Quality*. Wiley, 603-631.

[4] Mak, T.K. (1988). Analyzing intraclass correlation for dichotomous variables. *J. R. Statist. Soc.* B, 37, 344-352.

[5] Pannekoek, J. (1988). Interviewer variance in a telephone survey. *J. Off. Statist.*, 4, 375-384.

[6] Pannekoek, J. (1991). A mixed model for analyzing measurement errors for dichotomous variables. In: Biemer, Groves, Lyberg, Mathiowetz and Sudman (eds.), *Measurement Errors in Surveys.* Wiley, 488-517

[7] Stokes, S.L. (1988). Estimation of interviewer effects for categorical items in a random digit dial telephone survey. *J. Am. Statist. Assoc.*, 83, 623-630.

[8] Stokes, S.L. and Hill, J.R. (1985). Modelling interviewer variability for dichotomous variables. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 344-348.

[9] Stokes, S.L. and Mulry, M.H. (1987). On the design of interpenetration experiments for categorical data items. *J. Off. Statist.*, 3, 389-401.

[10] Rao, C.R. (1982). Diversity: its measurement, decomposition, apportionment and analysis. *Sankhya: Ind. J. Stat.*, Ser. A, 44, 1-22.

c.3