AN UNBIASED RESPONSE MODEL FOR ANALYSIS OF CATEGORICAL DATA

George E. Battese and Wayne A. Fuller, Iowa State University

1. INTRODUCTION

In this paper we consider the estimation of multinomial proportions when sample determinations contain classification or response errors. It is well known that if each individual in a simple random sample from a population is classified "correctly" then the sample proportions are unbiased estimators for the corresponding population proportions. However, in the presence of response errors the sample proportions are not necessarily unbiased for the population proportions. Given the response probabilities we present expressions for the expectations, variances and covariances for the sample proportions. In particular, we consider a response model that has the property that the sample proportions are unbiased estimators for the population proportions.

The effects of response errors on methods of analysis of categorical data have been considered by Bross (1954), Mote and Anderson (1965), Assakul and Proctor (1967), Koch (1969), McCarthy (1972), and others. Different response models have been discussed by Giesbrecht (1967), Bershad (1967), Koch (1968), Huang (1972), and Battese (1973).

In our discussion we assume that a simple random sample of size n is selected without replacement from a finite population of N individuals. Each individual is classified into one of r mutually exclusive and exhaustive classes. The sample classification is assumed to depend on the true class through parameters β_{ij}, i, $j = 1, 2, \ldots, r$, where $\sum_{j=1}^{r} \beta_{ij} = 1$, for all $i = 1, 2, \ldots, r$. The parameter, β_{ij}, is the probability that a randomly selected individual belonging to the i-th class is classified into the j-th class.

The proportion of the sample that is classified in class i is denoted by

$$p_i = \frac{1}{n} \sum_{k=1}^{N} \delta_k \gamma_k^i \tag{1.1}$$

where $\delta_k = 1$ if the k-th population individual is in the sample;

$= 0$ otherwise; and

$\gamma_k^i = 1$ if the k-th population individual is classified in class i;

$= 0$ otherwise.

It is readily verified that under these response hypotheses

$$E(p_i) = \sum_{m=1}^{r} P_m \beta_{mi} \equiv \bar{P}_i \tag{1.2}$$

$$\mathrm{Var}(p_i) = \frac{(N-n)}{n(N-1)} \bar{P}_i(1-\bar{P}_i) + \frac{(n-1)}{n(N-1)} \sum_{m=1}^{r} P_m \beta_{mi}(1-\beta_{mi}) \tag{1.3}$$

$$\mathrm{Cov}(p_i, p_j) = -\frac{(N-n)}{n(N-1)} \bar{P}_i \bar{P}_j - \frac{(n-1)}{n(N-1)} \sum_{m=1}^{r} P_m \beta_{mi} \beta_{mj} \tag{1.4}$$

where P_i, $i = 1, 2, \ldots, r$, denote the proportions of the population in the different classes.

The expression of (1.2) shows that the presence of response errors can result in the sample proportions being biased estimators of the true proportions.

2. AN UNBIASED RESPONSE MODEL

If the sample responses are such that the true response is reported a fraction, α, of the time and for the remaining fraction, $(1-\alpha)$, of the time the response is given with probabilities proportional to the population parameters P_i, $i = 1, 2, \ldots, r$, then the response probabilities are

$$\beta_{ii} = \alpha + (1-\alpha)P_i, \quad i = 1, 2, \ldots, r \tag{2.1}$$

$$\beta_{ij} = (1-\alpha)P_j, \quad i \neq j;\ i, j = 1, 2, \ldots, r \tag{2.2}$$

where α is a constant in the interval $[0, 1]$.

For this response model the sample proportion for any given class unbiasedly estimates the true proportion belonging to that class. That is, $\sum_{m=1}^{r} P_m \beta_{mi} = P_i$. The variances and covariances of the sample proportions for this response model are

$$\mathrm{Var}(p_i) = \frac{(N-n)}{n(N-1)} P_i(1-P_i) + \frac{(n-1)}{n(N-1)} (1-\alpha^2) P_i(1-P_i) \tag{2.3}$$

$$\mathrm{Cov}(p_i, p_j) = -\frac{(N-n)}{n(N-1)} P_i P_j - \frac{(n-1)}{n(N-1)} (1-\alpha^2) P_i P_j \; . \tag{2.4}$$

It is obvious that the response parameter, α, in the unbiased response model (2.1, 2.2) cannot be estimated without repeated responses from sample individuals. We consider the estimation of α and the (r-1) independent proportions, P_i, $i = 1, 2, \ldots, r-1$, from two independent responses on each sample individual. These two responses are assumed to be those obtained in " Trial 1" (an original interview) and " Trial 2" (a reinterview) of a survey. The proportion of the sample which responds in class i at Trial 1 and class j at Trial 2 is denoted by

$$p_{ij} = \frac{1}{n} \sum_{k=1}^{N} \delta_k \gamma_k^{ij} \qquad (2.5)$$

where γ_k^{ij} = 1 if the k-th individual is classified in class i at Trial 1 and class j at Trial 2;

= 0 otherwise.

Given that the Trial-1 and Trial-2 responses are independent, it can be verified that, for a general response model, the expectations, variances and covariances of the two-trial proportions are

$$E(p_{ij}) = \sum_{m=1}^{r} P_m \beta_{mi} \beta_{mj} \equiv P_{ij} \qquad (2.6)$$

$$\mathrm{Var}(p_{ij}) = \frac{(N-n)}{n(N-1)} P_{ij}(1-P_{ij}) + \frac{(n-1)}{n(N-1)} \sum_{m=1}^{r} P_m \beta_{mi} \beta_{mj} (1-\beta_{mi}\beta_{mj}) \qquad (2.7)$$

$$\mathrm{Cov}(p_{ij}, p_{i'j'}) = -\frac{(N-n)}{n(N-1)} P_{ij} P_{i'j'} - \frac{(n-1)}{n(N-1)} \sum_{m=1}^{r} P_m \beta_{mi} \beta_{mj} \beta_{mi'} \beta_{mj'}, \quad i \neq i' \text{ or } j \neq j' . \qquad (2.8)$$

Under the assumption of independence of the Trial-1 and Trial-2 responses, it is evident from (2.6) that the expected proportions P_{ij} and P_{ji} are equal. This implies that, at most, there exists $\frac{1}{2}(r+2)(r-1)$ independent expected proportions, P_{ij}. Further, the result of (2.6) suggests that large differences in the observed proportions p_{ij} and p_{ji}, $i \neq j$, may indicate lack of independence of the Trial-1 and Trial-2 survey responses. An approximate test for " lack of symmetry of the expected two-trial proportions, " or equivalently, " lack of independent classifications in two trials" is obtained with the statistic

$$X_S^2 = \sum_{i<}^{r} \sum_{j}^{r} (n_{ij} - n_{ji})^2 / (n_{ij} + n_{ji}) \qquad (2.9)$$

where n_{ij} denotes the number of the n sample individuals that are classified in the i-th class on Trial 1 and the j-th class on Trial 2 ($n_{ij} = np_{ij}$). This statistic converges to a central chi-square random variable with $\frac{1}{2}(r^2 - r)$ degrees of freedom under the hypothesis that $P_{ij} = P_{ji}$ for all i and j.

The expectations of the two-trial proportions for the unbiased response model (2.1, 2.2) are

$$P_{ii} = \alpha^2 P_i + (1-\alpha^2) P_i^2 \qquad (2.10)$$

and

$$P_{ij} = (1-\alpha^2) P_i P_j , \ i \neq j . \qquad (2.11)$$

The likelihood function for the Trial-1 and Trial-2 responses is that of the multinomial distribution with r^2 classes having probabilities, P_{ij}, $i, j = 1, 2, \ldots, r$, defined by (2.10) and (2.11). The maximum likelihood estimators for the independent parameters, P_i, $i = 1, 2, \ldots, r-1$, and α, are not readily obtainable from the likelihood equations [see Battese (1973)]. The Gauss-Newton estimators are, however, more easily obtained.

Given that the vector of the r^2-1 independent two-trial proportions is expressed by

$$Y = (p_{11}, p_{12}, \ldots, p_{1r}, \ldots, p_{r1}, p_{r2}, \ldots, p_{r,r-1})' \qquad (2.12)$$

we write the model

$$Y = P(\theta) + e \qquad (2.13)$$

where $P(\theta)$ denotes the vector of expected values of the sample proportions expressed as functions of the vector of independent parameters, θ; and e denotes the vector of the deviations of the observed proportions from the expected proportions. By expressing $P(\theta)$ as a Taylor expansion about an initial estimate for θ, denoted by $\tilde{\theta}$, we obtain the linear model

$$Y - P(\tilde{\theta}) = \frac{\partial P(\theta)}{\partial \tilde{\theta}} (\theta - \tilde{\theta}) + [R(\tilde{\theta}) + e] \qquad (2.14)$$

where $\frac{\partial P(\theta)}{\partial \tilde{\theta}}$ denotes the $(r^2 - 1) \times r$ matrix of partial derivatives of $P(\theta)$ with respect to the r elements of θ, evaluated at $\tilde{\theta}$; and $R(\tilde{\theta})$ denotes the vector of remainder terms in the Taylor expansion of $P(\theta)$ about the value of $\tilde{\theta}$. Possible initial estimations for the elements of θ are

$$\tilde{P}_i = \sum_{j=1}^{r} (p_{ij}+p_{ji})/2, \; i=1,2,\ldots,r-1 \qquad (2.15)$$

$$\tilde{\alpha} = \{ \sum_{j=1}^{r} (p_{jj}-\tilde{P}_j^2)/[\tilde{P}_j(1-\tilde{P}_j)r]\}^{1/2}. \qquad (2.16)$$

The estimator $\tilde{P}_i$ is unbiased for P_i under the assumptions of the unbiased response model. The initial estimator (2.16) for α is suggested because the quantities, $(P_{ii}-P_i^2)/P_i(1-P_i)$, for all $i=1,2,\ldots,r$, are equal to α^2 for the unbiased response model (2.1, 2.2).

We estimate the vector $\theta - \tilde{\theta} = \epsilon$ by

$$\hat{\epsilon} = (F' \tilde{V}^{-1} F)^{-1} F' \tilde{V}^{-1} W \qquad (2.17)$$

where $W = Y - P(\tilde{\theta})$; $F = \frac{\partial P(\theta)}{\partial \tilde{\theta}}$; and

$$\tilde{V} = \frac{1}{n} \{\mathrm{Diag}[P(\tilde{\theta})] - P(\tilde{\theta})[P(\tilde{\theta})]'\} . \qquad (2.18)$$

We consider the improved estimator

$$\hat{\theta} = \tilde{\theta} + \hat{\epsilon} \qquad (2.19)$$

and estimate its covariance matrix by

$$\hat{\mathrm{Cov}}(\hat{\theta}) = (F' \tilde{V}^{-1} F)^{-1} . \qquad (2.20)$$

An approximate test for the hypotheses (2.1, 2.2) of the unbiased response model is obtained with the statistic

$$X_U^2 = \sum_{i=1}^{r} \frac{[n_{ii} - nP_{ii}(\hat{\theta})]^2}{nP_{ii}(\hat{\theta})} + \sum_{i<j}^{r} \sum^{r} \frac{[n_{ij}+n_{ji}) - 2nP_{ij}(\hat{\theta})]^2}{2nP_{ij}(\hat{\theta})} \qquad (2.21)$$

where the $P_{ii}(\hat{\theta})$ and $P_{ij}(\hat{\theta})$, $i \neq j$, denote the estimates for the expected proportions (2.10, 2.11) obtained with the parameter estimates of (2.19). It can be shown [see Battese (1973)] that the statistic, X_U^2, converges to a central chi-square random variable with $\frac{1}{2}(r-2)(r+1)$ degrees of freedom under the hypothesis of the unbiased response model (2.1, 2.2).

3. EMPIRICAL EXAMPLE

During September and October of 1970 the Statistical Laboratory of Iowa State University conducted a survey of 262 Iowa farm operators. Each farm operator was personally visited in September and asked questions about his farming operations. About one month later each farm operator was personally visited by another interviewer. The questionnaire used for the second interview was constructed so that some of the questions were exactly the same as in the first interview. One of the purposes of the survey was to estimate the relative magnitude of the variance of response errors for several variates important in farm surveys. An analysis of the survey is presented in Battese, Fuller and Hickman (1972).

One of the questions that was asked farm operators in this study was: " In terms of total value of sales, what was the most important agricultural product sold from the land you operated in 1969? " Not all farm operators gave the same answer in the two interviews. We consider the data obtained in coding the responses into three categories of " most important product" : hogs, cattle, and not hogs or cattle. The distribution of the survey responses in the two different interviews is shown in Table 1.

Table 1. Frequency of farmers reporting the " most important product"

Trial-1 class	Trial-2 class			
	Hogs	Cattle	Other	Totals
Hogs	85	9	2	96
Cattle	12	77	4	93
Other	9	8	56	73
Totals	106	94	62	262

With these sample frequencies, the statistic (2.9) to test for " lack of independent classifications in the two trials" has the value

$$X_S^2 = (9-12)^2/21 + (2-9)^2/11 + (4-8)^2/12 = 6.22 .$$

The five percent critical value for a chi-square distribution with three degrees of freedom $[\frac{1}{2}(r^2 - r) = 3$ for $r = 3]$ is 7.81. At this level we do not reject the hypothesis of independent classifications in the two trials of the survey.

The initial estimators (2.15, 2.16) for the parameters in the unbiased model have values $\tilde{P}_1 = 0.385$, $\tilde{P}_2 = 0.357$ and $\tilde{\alpha} = 0.864$. From these initial estimates for the parameters in the model, the estimates for the P_{ij} in (2.10, 2.11) are $\tilde{P}_{11} = 0.325$, $\tilde{P}_{12} = 0.035$, $\tilde{P}_{13} = 0.025$, $\tilde{P}_{22} = 0.299$ and $\tilde{P}_{23} = 0.023$. The variables involved in the estimator $\hat{\epsilon}$ of (2.17) are thus

$$W = \begin{bmatrix} 0.324 \\ 0.034 \\ 0.008 \\ 0.046 \\ 0.294 \\ 0.015 \\ 0.034 \\ 0.030 \end{bmatrix} - \begin{bmatrix} 0.325 \\ 0.035 \\ 0.025 \\ 0.035 \\ 0.299 \\ 0.023 \\ 0.025 \\ 0.023 \end{bmatrix} = \begin{bmatrix} -0.001 \\ -0.001 \\ -0.018 \\ 0.011 \\ -0.005 \\ -0.008 \\ 0.009 \\ 0.007 \end{bmatrix}$$

and

$$F = \begin{bmatrix} 0.942 & 0.000 & 0.409 \\ 0.091 & 0.098 & -0.238 \\ -0.032 & -0.098 & -0.172 \\ 0.091 & 0.098 & -0.238 \\ 0.000 & 0.927 & 0.397 \\ -0.091 & -0.025 & -0.159 \\ -0.032 & -0.098 & -0.172 \\ -0.091 & -0.025 & -0.159 \end{bmatrix}.$$

The estimated covariance matrix (2.18) is obtained with the values of the vector $P(\tilde{\theta})$. From these data the elements of the estimator $\hat{\epsilon}$, defined by (2.17), are calculated to be 0.0003, 0.0030 and -0.0009 with standard errors 0.028, 0.027 and 0.020, respectively. The new estimates for the parameters in the model are thus

$$\hat{P}_1 = 0.386, \quad \hat{P}_2 = 0.360 \text{ and } \hat{\alpha} = 0.863 .$$

With these parameter estimates the expected frequencies for the two interviews are estimated by $n\hat{P}_{11} = 85.25$, $n\hat{P}_{12} = 78.97$, $n\hat{P}_{13} = 53.92$, $n\hat{P}_{22} = 18.55$, $n\hat{P}_{23} = 13.10$, and $n\hat{P}_{33} = 12.21$. The statistic (2.21) for testing the response model has the value

$$\begin{aligned} X_U^2 &= (85-85.25)^2/85.25 + (77-78.97)^2/78.97 \\ &\quad + (56-53.92)^2/53.92 + (21-18.55)^2/18.55 \\ &\quad + (11-13.10)^2/13.10 + (12-12.21)^2/12.21 \\ &= 0.79 . \end{aligned}$$

The five-percent critical value for the chi-square distribution with two degrees of freedom $[\frac{1}{2}(r-2)(r+1) = 2$ for $r = 3]$ is 5.99. We therefore conclude that the unbiased response model (2.1, 2.2) is consistent with the observed frequencies for the "most important agricultural product in 1969."

4. EXISTENCE OF GENERAL UNBIASED RESPONSE MODELS

The unbiased response model, defined by (2.1, 2.2), consists of (r-1) independent population proportions and one response parameter, α. The model has the property that the probabilities of incorrectly reporting a given class are the same. We seek to determine if there exist more general response models that satisfy the unbiasedness conditions;

$$\sum_{m=1}^{r} P_m \beta_{mi} = P_i , \quad i = 1, 2, \ldots, r .$$

We assume that for a survey sample, in which each individual reports his classification in an interview and a re-interview, the expected proportions are (conceptually) known and satisfy the conditions

$$P_{ij} = P_{ji} , \quad i, j = 1, 2, \ldots, r \tag{4.1}$$

and

$$\sum_{j=1}^{r} P_{ij} = P_i , \quad i = 1, 2, \ldots, r . \tag{4.2}$$

We seek to determine conditions under which it is possible to "recover" the response probabilities that generated the expected proportions, P_{ij}.

We consider the equations (4.1, 4.2) and

$$P_{ij} = \sum_{m=1}^{r} P_m \beta_{mi} \beta_{mj} , \quad i,j = 1,2,\ldots,r , \tag{4.3}$$

and seek to solve for the parameters β_{ij}, $i, j = 1, 2, \ldots, r$, such that they are nonnegative and satisfy the conditions $\sum_{j=1}^{r} \beta_{ij} = 1$, for all $i = 1, 2, \ldots, r$.

We first investigate the solution for the case of two classes (r=2). For the two-category case the conditions, $\sum_{j=1}^{2} \beta_{ij} = 1$ and $\sum_{m=1}^{2} P_m \beta_{mi} = P_i$, $i = 1, 2$, imply that the response parameters, β_{12}, β_{21} and β_{22} are expressed by

$$\beta_{12} = 1 - \beta_{11} \tag{4.4}$$

$$\beta_{21} = (1-\beta_{11}) P_1 / (1-P_1) \tag{4.5}$$

and

$$\beta_{22} = 1 - (1-\beta_{11}) P_1 / (1-P_1) \tag{4.6}$$

The expected proportion P_{11}, expressed in terms of β_{11}, is thus

$$P_{11} = \beta_{11}^2 P_1 + (1-\beta_{11})^2 P_1^2 / (1-P_1). \tag{4.7}$$

By expressing this equation as a quadratic in

β_{11} we obtain

$$0 = P_1(\beta_{11}-P_1)^2-(1-P_1)(P_{11}-P_1^2).$$

There exists a real solution for β_{11} if

$$P_{11}-P_1^2 \geq 0 . \qquad (4.8)$$

Given that this condition is satisfied a solution for the response parameters is

$$\beta_{ii} = |c|+(1-|c|)P_i , \quad i = 1, 2 \qquad (4.9)$$

$$\beta_{ij} = (1-|c|)P_j \quad , \ i \neq j \qquad (4.10)$$

where

$$c^2 = (P_{11}-P_1^2)/P_1(1-P_1)$$
$$= (P_{22}-P_2^2)/P_2(1-P_2) . \qquad (4.11)$$

Further, if the expected proportions satisfy the conditions

$$P_i^2 \leq P_{ii} \leq P_i^2/(1-P_i), \quad i = 1, 2, \qquad (4.12)$$

then the solutions for the response parameters are

$$\beta_{ii} = c+(1-c)P_i \quad , \ i = 1, 2 \qquad (4.13)$$

$$\beta_{ij} = (1-c)P_j \quad , \ i \neq j , \qquad (4.14)$$

where c^2 is defined by (4.11). It follows from (4.13, 4.14) that if the negative root of (4.11) is taken when the conditions of (4.12) are satisfied, then β_{ii} is less than β_{ji}, $j \neq i$, $i = 1, 2$. This is an unlikely situation in practice so that the solution defined by (4.9, 4.10) gives the appropriate response probabilities for an unbiased response model in the two-category case. The response model defined by (4.9, 4.10) is obviously a member of the class of unbiased response models defined by (2.1, 2.2).

It is readily seen that when there are more than two categories for the responses, solutions of the Equations (4.1, 4.2, 4.3) for the response parameters cannot be obtained in closed form without additional assumptions. We assume that for the r-category case ($r \geq 3$) Condition 1 is satisfied.

Condition 1: The probabilities of correct classification for class i, $i = 1, 2, \ldots, r$, are those that would be obtained from the 2 x 2 interview-reinterview problem considering only the two classes, "class i" and "not class i."

Given Condition 1 it can be shown [see Battese (1973)] that the solution for the probability of a correct response for class i is

$$\beta_{ii} = P_i+[(1-P_i)(P_{ii}-P_i^2)/P_i]^{1/2} \qquad (4.15)$$

provided that P_{ii} is no smaller than P_i^2. Further, if $r = 3$ and Condition 1 is satisfied, then the solution for the probability of incorrectly reporting class j when class i is the true class is

$$\beta_{ij} = (1-\beta_{jj})P_j/(1-P_j), \quad i \neq j . \qquad (4.16)$$

For the case when $r > 3$ we assume that Condition 2 is satisfied.

Condition 2: The probabilities of misclassification, β_{ij}, $i \neq j$, $i, j = 1, 2, \ldots, r$, are those that would be obtained from the 3 x 3 interview-reinterview problem considering only the three classes, "class i", "class j", and "not class i or class j."

Given Conditions 1 and 2 the solutions for the response probabilities are given by (4.15) and (4.16). These solutions and the conditions $\sum_{j=1}^{r} \beta_{ij} = 1$, for all $i = 1, 2, \ldots, r$, imply that the relationship

$$(P_{ii}-P_i^2)/P_i(1-P_i) = (P_{jj}-P_j^2)/P_j(1-P_j) = c^2 \qquad (4.17)$$

holds for all i and j. Equations (4.15)-(4.17) imply that if the response probabilities satisfy Conditions 1 and 2 then the solution of the response probabilities in terms of the expected proportions is given by

$$\beta_{ii} = |c|+(1-|c|)P_i, \quad i = 1,2,\ldots,r \qquad (4.18)$$

and

$$\beta_{ij} = (1-|c|)P_j, \quad i \neq j, \ i,j = 1,2,\ldots,r \qquad (4.19)$$

where c^2 is defined by (4.17)

Conditions 1 and 2 are strong conditons and it is easy to think of situations where we would expect them to be violated. For example, we would not expect the situation to hold for individuals placed into classes on the basis of a free response to a continuous variable. However, in situations where sample responses are obtained to open-end questions, such as, "What is the most important source of your income?", it may be reasonable to assume that the response hypotheses satisfy Conditions 1 and 2.

5. CONCLUSIONS

In this paper we present a class of unbiased response models that is defined by (r-1) independent proportions and a single response parameter. The response probabilities are a function of the true class proportions and the probabilities of misclassification in a given class are the same. Models that define the response probabilities independently of the population proportions generally will not satisfy the unbiasedness condition. For example, Mote and Anderson (1965), considered two simple response models in their investigation of the effect of misclassification on the size and power of chi-square goodness-of-fit tests for categorical data. The first model assumed equal probabilities of misclassification and the second model assumed that the only misclassifications were into classes adjoining the true classes. In both of these models the sample proportions are not, in general, unbiased estimators of the corresponding population proportions.

ACKNOWLEDGEMENT

The work reported in this paper was partially supported by the Bureau of the Census through the Joint Statistical Agreement, J.S.A. 72-4.

REFERENCES

[1] Assakul, K. and Proctor, C.H. (1967), "Testing Independence in Two-Way Contingency Tables With Data Subject to Misclassification," Psychometrika, 32, 67-76.

[2] Battese, G.E. (1973), " Parametric Models for Response Errors in Survey Sampling," Ph.D. thesis, 127 p., Iowa State University, Ames.

[3] Battese, G. E., Fuller, W. A. and Hickman, R. D. (1972), " Interviewer Effects and Response Errors in a Replicated Survey of Farm Operators in Selected Iowa Counties," Unpublished Report to the USDA's Statistical Reporting Service, Statistical Laboratory, Iowa State University, Ames.

[4] Bershad, M.A. (1967), " Gross Changes in the Presence of Response Errors," Unpublished memorandum, U.S. Bureau of the Census, Washington, D.C.

[5] Bross, I. (1954), " Misclassification in 2x2 Tables," Biometrics, 10, 478-86.

[6] Giesbrecht, F. G. (1967), " Classification Errors and Measures of Association in Contingency Tables," Proceedings of the Social Statistics Section of the American Statistical Association, 1967, 271-76.

[7] Huang, H. T. (1972), " Combining Multiple Responses in Sample Surveys," Ph.D. thesis, Iowa State University, Ames.

[8] Koch, G.G. (1968), " A Simple Model for Misclassification Errors in 2x2 Contingency Tables," Research Triangle Institute Technical Report SU-363, 24-28.

[9] Koch, G.G. (1969), " The Effect of Nonsampling Errors on Measures of Association in 2x2 Contingency Tables," Journal of the American Statistical Association, 64, 852-63.

[10] McCarthy, P.J. (1972), " Effects of Discarding Inliars When Binomial Data Are Subject to Classification Errors," Journal of the American Statistical Association, 67, 515-29.

[11] Mote, V.L. and Anderson, R.L. (1965), " An Investigation of the Effect of Misclassification on the Properties of χ^2-tests in the Analysis of Categorical Data," Biometrika, 52, 95-109.