ASSESSING SOME ESTIMATION CRITERIA OF
MEASUREMENT ERROR FOR CATEGORICAL DATA

Roshanak Alimohammadi

Statistics Department
Mathematical Sciences Faculty
Alzahra University, Vanak, Tehran, IRAN

Abstract: The measurement error is an essential and unavoidable part of nonsampling error in surveys. Therefore, estimation of the measurement error to compute precision of the survey results is necessary. In the previous investigations, modeling of measurement error for continues and categorical data is studied. In this paper, some criteria to quantify measurement error of categorical data are proposed. Besides, in a simulation study, the effect of sample size and number of categories on performance of the criteria is assessed. Furthermore, based on the simulation results, the standard error of the criteria is compared and the proper criterion is proposed for each of the cases of categorical data.

1. Introduction

The measurement error is a major component of nonsampling error in surveys. The modeling of the measurement error may be considered as a way to quantify the measurement error. Kish [7], Biemer and Stocks [3] and Biemer and Trewin [4] applied ANOVA for measurement error modeling. Alimohammadi and Navvabpour [1] proposed a model of the error in face to face surveys for continues data. Modeling of the measurement error for categorical data is different

from continuous data. Alimohammadi [2] proposed a model of measurement error for multilevel categorical data specially for nominal data in surveys.

In this paper, some agreement criteria are applied to estimate measurement error of categorical data. Besides, in a simulation study, the standard error of the estimators of these criteria to quantify measurement error of categorical data is compared.

2. Some Criteria to Quantify Measurement Error of Categorical Data

Cohen [5] introduced the kappa coefficient, and Flies and Cohen [6] introduced the weighted kappa to assess agreement between two raters. The weighted kappa coefficient is a generalization of the simple kappa coefficient (named Cohen's Kappa), using weights to consider the relative difference between categories. Consider a $l \times l$ ($i = 1, 2, \ldots, l$ and $j = 1, 2, \ldots, l$) probability table, where π_{ij} is the probability of agreement between two raters that recognized i'th level of variable by rater 1 and j'th level by rater 2 respectively. $\pi_{i.}$ and $\pi_{.j}$ are the marginal probabilities of the $l \times l$ agreement probability table.

In this $l \times l$ table, the orthogonal elements indicate agreement probability between rater 1 and 2. The weights w_{ij} are constructed so that for all i and j, $w_{ii} = 1$, and $w_{ij} = w_{ji}$. The weighted kappa coefficient is defined as:

$$\kappa_w = \frac{\theta_{0w} - \theta_{ew}}{(1 - \theta_{ew})}, \tag{1}$$

where $\theta_{0w} = \Sigma_i \Sigma_j w_{ij} \pi_{ij}$ and $\theta_{ew} = \Sigma_i \Sigma_j w_{ij} \pi_{i.} \pi_{.j}$. In the relation (1), if all the weights $w_{ij} = 0$ except of the main diagonal, and $w_{ii} = 1$, the usual unweighted kappa, named Cohen's kappa may be obtained as:

$$\kappa = \frac{\theta_0 - \theta_e}{(1 - \theta_e)}, \tag{2}$$

where $\theta_0 = \Sigma_i \pi_{ii}$ is probability of agreement and $\theta_e = \Sigma_i \Sigma_j \pi_{i.} \pi_{.j}$ shows the proportion of units for which agreement is expected by chance. $\pi_{i.}$ and $\pi_{.j}$ indicate sum of rows and columns in the probability table respectively. By considering absolute value $|i - j|$ for the amount of measurement error, it can be shown that $\kappa = 1$, if and only if $\theta_0 = 1$. Therefore, in this case there is complete agreement between the observed and true values. In the other words, measurement error does not occur for this amount of the criterion. $\kappa = 0$ if and only if $\theta_0 = \theta_e$. This equality is satisfied when $\pi_{ij} = \pi_{i.} \pi_{.j}$. Therefore,

the observed values and the true values are independent. This criterion may be negative, when the probability of agreement is less than the probability of agreement by chance. This case indicates large amount of measurement error. For binary variables, that is 2×2 probability table, it can be easily shown that the minimum value of the criterion is -1. Generally, the results of this section imply that criteria (1) and (2) can be applied for interpreting the amount of measurement error for categorical data. In fact, the kappa and weighted kappa are applied to quantify measurement error of categorical data. If the categorical data are nominal, Cohen's kappa is appropriate criterion to estimate measurement error. But, when ordinal data are considered, the weighted kappa is more appropriate. In the weighted kappa, the two major kinds of weights are linear and quadratic weights. The linear weight is defined as $w_{ij}^L = 1 - \frac{|i-j|}{l-1}$ and the quadratic weight is $w_{ij}^Q = 1 - \frac{(i-j)^2}{(l-1)^2}$, where l is the number of variable categories. It can be shown easily that the amount of the unweighted kappa is less than the weighted kappa and the weighted kappa with linear weights is less than the weighted kappa with quadratic weights.

In the next section, the performance of the considered criteria is compared in a simulation study by computing standard errors for these criteria.

3. Simulation Study to Compare Standard Errors of the Quantifying Criteria of Measurement Error

In the following simulations, 1000 repeats of each of the cases is conducted and Average of Standard Error (ASE) for the criteria is computed.

In Table 1, the effect of number of categories is computed for Cohen's kappa (κ), linear weighted kappa (κ_W^L) and quadratic weighted kappa (κ_W^Q) criteria.

Number of levels	ASE(κ)	ASE(κ_W^L)	ASE(κ_W^Q)
3	0.073	0.111	0.107
4	0.066	0.128	0.110
5	0.0632	0.181	0.112

Table 1: Effect of categories number on ASE

Based on the results of the simulation study in Table 1, for each of number of variable categories, Cohen's kappa has the least standard error in average (ASE) and quadratic weighted kappa has less standard error than linear weighted kappa. Then, Cohen's kappa is a proper and reasonable criterion

to estimate measurement error for nominal data and the weighted kappa with quadratic weights has a better performance than linear weighted kappa to estimate measurement error for ordinal categorical data. On the other hand, ASE of Cohen's kappa reduces by increasing categories numbers, but ASE of weighted kappa (both linear and quadratic weights) increases by increasing categories number.

Simulation study for investigating the effect of sample size on Average of Standard Error (ASE) for the criteria is shown in Table 2.

Sample size	ASE(κ)	ASE(κ_W^L)	ASE(κ_W^Q)
30	0.120	0.235	0.208
50	0.087	0.186	0.181
75	0.071	0.150	0.139
100	0.062	0.132	0.111

Table 2: Effect of level sample size on ASE

The results of simulation study in Table 2 shows that increasing of sample sizes decreases average of standard error and this result is reasonable. On the other hand, comparison of the columns of Table 2, confirms that quadratic weighted kappa is more appropriate than linear weighted kappa between two considered weights. Therefore, applying weighted kappa with quadratic weights is confirmed to assess measurement error for ordinal categorical data specially. If nominal categorical data is considered, Cohen's kappa may be more appropriate because of the less standard error and the best performance.

4. Conclusion

The estimation of the measurement error is essential to study data quality in surveys. In this paper, some criteria to estimate of measurement error for categorical data are proposed based on the previous works. Furthermore, based on a simulation study, Cohen's kappa, the weighted kappa with linear and quadratic weights are compared.

The results illustrate that Cohen's kappa has the best performance to quantify the measurement error of nominal data and the weighted kappa with quadratic weights is more appropriate than that of linear weights to estimate measurement error. Therefore, applying Cohen's kappa is proposed to quantify the measurement error of nominal data and the weighted kappa with quadratic

weights is more appropriate to estimate of the measurement error for ordinal data.

Acknowledgements

References

[1] R. Alimohammadi and H.R. Navvabpour, Response error modeling in face to face surveys, *Research Journal of Science of Isfahan University*, **4**, No 33 (2008), 1-14.

[2] R. Alimohammadi, Modeling of measurement error for categorical data in surveys, *International Mathematical Forum*, **7**, No 57 (2012), 2833-2837.

[3] P. Biemer and S.L. Stokes, Approaches of modeling of measurement error, In: P. Biemer et al., *Measurement Error in Surveys*, Wiley, New York (1991), 487-516.

[4] P. Biemer and D. Trewin, A review of measurement error effects on the analysis of survey data, In: L. Lyberg, P. Biemer, M. Collins, E. Deleeuw, C. Dippo, N. Schwartz and D. Trewin (Eds.), *Survey Measurement and Process Quality*, Wiley, New York (1997).

[5] J. Cohen, A coefficient of agreement for nominal scales, *Educational and Psycological Measurement*, **10**, No 1 (1960), 37-46.

[6] J. Fleiss and J. Cohen, The equivalence of weighted kappa and the intraclass correlation coefficient as measure of reliability, *Educational and Psycological Measurement*, **33** (1973), 613-619.

[7] L. Kish, Studies of interviewer variance for attitudinal variables, *Journal of the American Statistical Association*, **57** (1962), 92-115.