XXI IMEKO World Congress "Measurement in Research and Industry"
August 30 – September 4, 2015, Prague, Czech Republic

PRECISION EVALUATION FOR NON-QUANTITATIVE MEASUREMENTS ~ BINARY AND ORDINAL CATEGORICAL CASES ~

Tomomichi Suzuki [1], *Akimasa Katakura* [2], *Rei Miyazawa* [3], *Natsuki Sano* [4]

[1] 1 Tokyo University of Science, Noda, Japan, szk@rs.tus.ac.jp
[2] Tokyo University of Science, Noda, Japan, j7410040@ed.tus.ac.jp
[3] Tokyo University of Science, Noda, Japan, j7411121@ed.tus.ac.jp
[4] Tokyo University of Science, Noda, Japan, nsano@rs.tus.ac.jp

Abstract – Precision evaluation in quantitative measurements is a thoroughly discussed topic and the established methods are in use. Many methods are proposed for qualitative data, but their effectiveness and statistical properties are not so clear. This paper examines their properties by applying some of the major methods and then by comparing the applied results. The comparisons are made to binary data and also to ordinal categorical data. Some of the similarities and the differences among the methods became clear after examining the results.

Keywords: repeatability, reproducibility, qualitative data

1. INTRODUCTION

Precision evaluation in quantitative measurements is a thoroughly discussed topic and the established methods are in use. ISO published ISO 5725 *accuracy* (*trueness and precision*) *of measurement methods and results* which deals with quantitative measurements. But there are is standard equivalent for non-quantitative measurements.

Many methods are proposed for qualitative data including binary data, but their effectiveness and statistical properties are not so clear. This paper examines their properties by applying some of the major methods and then by comparing the applied results. The comparisons are made to binary data and also to ordinal categorical data.

2. PRECISION FOR QUANTITATIVE DATA

ISO 5725 [1] uses two terms trueness and precision to describe the accuracy of a measurement method. Trueness refers to the closeness of agreement between the arithmetic mean of a large number of test results and the true or accepted reference value. Precision refers to the closeness of agreement between test results.

ISO 5725 defines the precision using the model shown in (1).

$$y = m + B + e \tag{1}$$

In (1), y is the measurement result, m is the general mean (expectation), B is the laboratory component of bias under repeatability conditions, and e is the random error in every measurement under repeatability conditions. The term B is considered a random variable whose expectation equals zero and whose variance is expressed as σ_L^2. The term e is a random variable whose expectation equals zero and whose variance is expressed as σ_e^2.

ISO 5725 [1] also introduces two components of precision: repeatability and reproducibility

Repeatability is the precision under repeatability conditions. Repeatability conditions are defined as "conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time." Repeatability indicates the smallest variation for a particular measurement method.

Reproducibility is the precision under reproducibility conditions. Reproducibility conditions are defined as "conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment." Reproducibility indicates the largest variation for a particular measurement method.

Using the notation shown in (1), the repeatability variance σ_r^2 and reproducibility variance σ_R^2 can be expressed as in (2) and (3).

$$\sigma_r^2 = \sigma_e^2 \tag{2}$$

$$\sigma_R^2 = \sigma_L^2 + \sigma_r^2 = \sigma_L^2 + \sigma_e^2 \tag{3}$$

The estimates of repeatability variance and reproducibility variance are calculated from interlaboratory studies or collaborative assessment experiments [2].

3. QUALITATIVE DATA

There are many types of non-quantitative measurements. The type of data varies accordingly, for example, binary data, ordinal categorical data, pure categorical data, etc. In this paper, the methods to evaluate precision for binary data are considered.

The data format of qualitative data is shown in Table 1. Every laboratory measures the identical test item number of times. The number of laboratories is usually denoted as p, and the number of replication is usually denoted as n. If the study is conducted on multiple items or levels, the data for each item or level can be expressed as Table 1.

Table 1. Data format for qualitative data.

laboratory	run 1	…	run k	…	run n
1	y_{11}	…	y_{1k}	…	y_{1n}
2	y_{21}	…	y_{2k}	…	y_{2n}
:					
:					
i	y_{i1}	…	y_{ik}	…	y_{in}
:					
:					
p	y_{p1}	…	y_{pk}	…	y_{pn}

In the binary data case, the value of y_{ik} takes only two values, for example (positive, negative), (detect, non-detect), (pass, fail). Values 1 and 0 are often used to identify the measurement result. In the ordinal categorical data case, the value y_{ik} takes various values, for example (good, fair, poor), (A, B, C, D). Numerical values, for example(1,2,3,4), are often used to identify the measurement result.

4. PRECISION FOR QUALITATIVE DATA

The following precision evaluation methods for qualitative data are examined in this study.

4.1. Methods examined for binary data

Three methods to evaluate precision for binary data are examined in this study.

4.1.1. ISO 5725 based method

This is the method proposed by Wilrich [3] and discussed in Horie et al [4]. This method is based on the approach of ISO 5725 part 2 [2] and treats the measurement values (zero and one) in the same manner as in ISO 5725 in analysis (ANOVA etc.).

4.1.2. Kappa statistic

The method uses the Fleiss' Kappa statistic [5]. The method is also known as Attribute Agreement Analysis(AAA) [6]. Kappa statistic, or kappa coefficient can be calculated for between appraisers (between laboratories) and also for within appraisers (within laboratories).

4.1.3. Likelihood based method

This method is proposed by van Wieringen et al. [7]. The idea is based on sensitivity and specificity. It applies latent class model, where unknown true value (pass or fail) is assumed. The probability of a product being 'pass' will be estimated. The theory is based on maximum likelihood functions method. Repeatability and reproducibility are defined using deviances which are based on likelihood. The estimate is obtained by EM algorithm. This method is intended to integrate sensitivity and specificity with repeatability and reproducibility.

4.2. Methods examined for ordinal categorical data

Three methods to evaluate precision for ordinal categorical data are examined in this study.

4.2.1. Kappa statistic

The method uses the Fleiss' Kappa statistic [5]. The method is also known as Attribute Agreement Analysis [6]. Kappa statistic, or kappa coefficient can be calculated for between appraisers (between laboratories) and also for within appraisers (within laboratories).

4.2.2. ORDANOVA

This is the method proposed by Gadrich et al [8] and discussed in Bashkansky et al [9]. This method is based on the essential characteristics of ordinal categorical data.

4.2.3.Kendall rank correlation coefficient

This is the method proposed by Kendall [10]. This popular statistic can be used to compare two ordering systems.

Since there will be many ties in this study, the formulae considering ties are used.

5. SIMULATION FOR COMPARISON

5.1. Simulation procedure

The comparisons of the methods are performed in the following procedure.

1) The parameters of the models are set. The values are selected to reproduce actual collaborative studies as much as possible.
2) Set of data is obtained through simulation.
3) Precision measure is calculated for the data using the precision evaluation methods.
4).Repeat procedure 2) and 3) for number of simulations.
5) Repeat procedure 2) to 4) under different values of parameters.

After all the precision measures are calculated, the results are compared.

5.2 Parameters for binary data case

The beta binomial distribution is assumed behind y_{ik}. The data for laboratory i follows binomial distribution $\mathrm{Bin}(n, \pi_i)$, where π_i follows beta distribution $\mathrm{Beta}(\alpha, \beta)$. The following parameters are considered.

number of laboratories p: 6, 12

number of measurement replications n: 8, 12, 24, 48

parameters (α, β):

(15,15), (40,40), (125,125) for $\bar{\pi}_i$ = 0.5
large repeatability (no.1 to 3)

(15,5), (45,15), (141,47) for $\bar{\pi}_i$ = 0.75
medium repeatability (no.4 to 6)

(9,1), (27,3), (90,9) for $\bar{\pi}_i$ = 0.9
small repeatability (no.7 to 9)

(In each line of the above, they are in decreasing order in terms of between laboratory variance.)

number of simulations for each parameter set: 100

5.3 Parameters for ordinal categorical data case

In this study, the number of ordinal categories is set to five. The categories are named as one, two, three, four and five. In actual ordinal categorical data, only the order of values can be evaluated. In this study, a latent (or hidden) continuous variable *Z* is assumed, and the value (hence category) of y_{ik} is the closest integer of *Z*. *Z* is sum of three (the centre of the categories) plus random uniform variate with expectation $\sqrt{3}\sigma_L$ and between-laboratory variance σ_L^2 plus random normal variate with expectation zero and within-laboratory variance σ_e^2. The following parameters are considered.

number of laboratories *p*: 10, 20, 40, 80

number of measurement replications *n*: 10, 20, 40, 80

parameters (σ_L^2, σ_e^2):

$(1/12, 0.75^2)$, $(1/48, 0.75^2)$, $(1/192, 0.75^2)$
large repeatability

$(1/12, 0.50^2)$, $(1/48, 0.50^2)$, $(1/192, 0.50^2)$
medium repeatability

$(1/12, 0.25^2)$, $(1/48, 0.25^2)$, $(1/192, 0.25^2)$
small repeatability

(In each line of the above, they are in decreasing order in terms of between laboratory variance.)

number of simulations for each parameter set: 1000

6. RESULTS AND DISCUSSION

6.1. Binary data case

The excerpts from the results of the binary data cases are shown in Fig. 1 to Fig. 3. Although the same datasets are used in calculating the precision measures, the histograms of the precision measures and relations are different among the methods.

Comparing Fig. 2 with other figures, it is apparent that the dispersion of the precision measure for AAA method is relatively large compared to other methods. From Fig. 3, we can see that three levels of set parameters are not reproduced when using likelihood based method, especially when number of measurement replication is small.

Largely speaking, the ISO 5725 based method was the best among the methods, especially for evaluating reproducibility variances. Regarding the repeatability variances, the AAA method may the best. Regarding the measurement replications, we can say that replication of 24 or more is necessary to obtain 'accurate' precision measures.

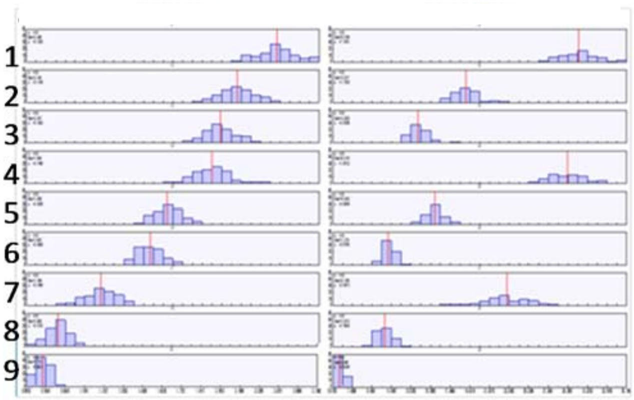

Fig. 1. Reproduciblity variance of ISO 5725 based method

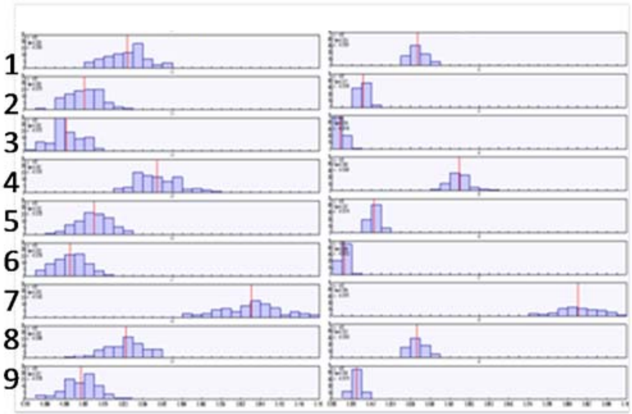

Fig. 2. Repeatability variance of AAA

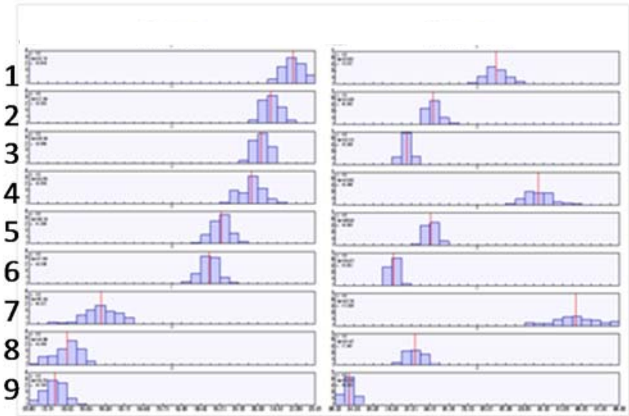

Fig. 3. Repeatability variance of likelihood based method

6.2. Ordinal categorical data case

The excerpts from the results of the ordinal categorical data cases are shown in Fig. 4 and Fig. 5. Although the same datasets are used in calculating the precision measures, the histograms of the precision measures and relations are also different among the methods.

Both the AAA method and the ORDANOVA method are able to express the precision of the measurements. The larger the number of laboratories is, the better the evaluation of the precision. When the number of replication was small, Kendall rank correlation coefficient was not efficient. Overall, the ORDANOVA method seems the best among the methods.

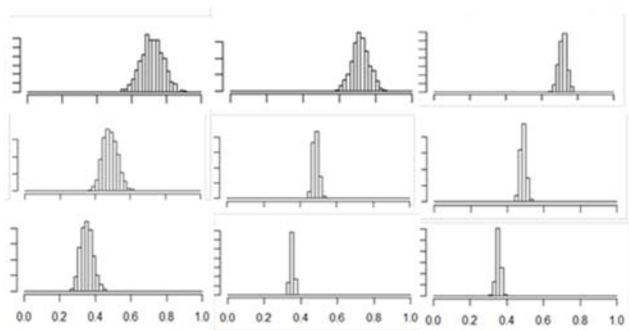

Fig. 4. Repeatability variance of AAA

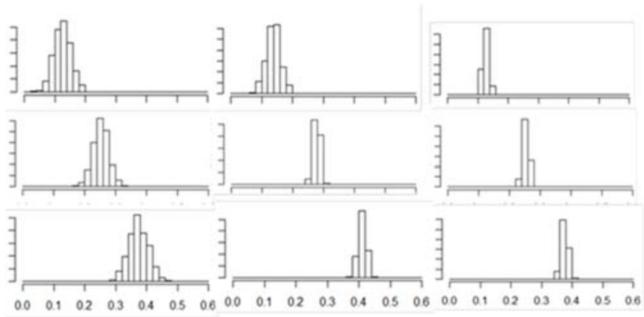

Fig. 5. Repeatability variance of ORDANOVA

4. CONCLUSIONS

The method for evaluating precision for binary data and for ordinal categorical data are compared and examined. We also proposed how to evaluate the precision measures. This procedure can be applied in evaluating more methods.

Through examination of the results, some statistical aspects of the methods became clear, especially on evaluating repeatability and reproducibility variances.

REFERENCES

[1] ISO 5725, *Accuracy (trueness and precision) of measurement methods and result – Part 1 :General principles and definitions*, ISO, 1994.

[2] ISO 5725, *Accuracy (trueness and precision) of measurement methods and result – Part 2 : Basic methods for the determination of repeatability and reproducibility of a standard measurement methods*, ISO, 1994.

[3] P.-Th. Wilrich, "The determination of precision of qualitative measurement methods by interlaboratory experiments", *Accreditation and Quality Assurance*, vol. 15, pp. 439-444, 2010.

[4] K. Horie, Y. Tsutsumi, and T. Suzuki, "Calculation of Repeatability and Reproducibility for Qualitative Data", *6th ANQ Congress*, CDROM, Bangkok, Thailand, 2008.

[5] J. L. Fleiss, *Statistical Methods for Rates and Proportions 2nd edition*, John Wiley &Sons, 1981.

[6] ISO/TR 14468, *Selected illustrations of attribute agreement analysis*, ISO, 2010.

[7] W. N. van Wieringen, J. De Mast, "Measurement System Analysis for Binary Data", *Technometrics*, vol. 50, pp. 468-478, 2008.

[8] T. Gadrich, and E. Bashkansky, "ORDANOVA: Analysis of Ordinal Variation", *Journal of Statistical Planning and Inference*, 142, 3174-3188, 2012.

[9] E. Bashkansky, T. Gadrich, and I. Kuselman, "Interlaboratory comparison of measurement results of an ordinal or nominal binary property", *Accreditation and Quality Assurance*, vol. 17, pp. 239-243, 2012.

[10] M. Kendall, "A New Measure of Rank Correlation", *Biometrika*, vol.30, pp. 81-89, 1938.