A NEW OPTIMAL SUMMARY MEASURE FOR LIKERT SCALE STUDENT EVALUATIONS OF TEACHING

Thisaakhya Jayakody[1], B.M.S.G. Banneheka[2]

[1,2] Department of Statistics, Faculty of Applied Sciences, University of Sri Jayewardenepura, Nugegoda, Sri Lanka.

Abstract

Student evaluations of teaching (SET) is one of the oldest techniques used to assess lecturer performance in universities around the globe and it plays a crucial role in the Sri Lankan state university system as a critical quality assurance measure. However, universities continue to use incorrect summary measures to analyze student responses which in turn leads to incorrect decision making and other critical errors. The objective of this study was to identify a population parameter to represent the most deserving response to a question with 5-point Likert scale choices in a student feedback form and an optimal estimator for the same. The population mean and population median as parameters with corresponding estimators sample mean and sample median and the population mode as a parameter with sample mode and a newly introduced adjusted sample mode as estimators, were considered. The responses to a question with 5-point Likert scale choices were considered as the rounded integer values of a random variable that follows some continuous distribution on the support [1,5]. The bias and the root mean squared error (RMSE) of each estimator were estimated based on 10000 samples of size 50 from the truncated normal distribution truncated at 1 and 5. Simulations were done by fixing the value of each parameter at every possible student response. The population mean and median never take values 1 or 5. Hence they are not suitable when the deserving response is 1 or 5. Therefore, they were rejected even though their estimators perform well when the corresponding parameters are in the interval [2, 4]. The population mode can be any value in the interval [1, 5]. Hence it is suitable to represent the deserving response. Out of its two estimators, the adjusted sample mode was found to be an optimal estimator with a smaller bias and RMSE. These findings are discussed with the intention of developing a performance index to quantify students' perception regarding a lecturer in the future.

Keywords: SET, student feedback, simulations, optimal estimator, truncated normal

[1]*Thisaakhya Jayakody: thisaakhya@gmail.com*

[2]*B.M.S.G. Banneheka: sarath.banneheka@sjip.ac.lk*

A NEW OPTIMAL SUMMARY MEASURE FOR LIKERT SCALE STUDENT EVALUATIONS OF TEACHING

***Thisaakhya Jayakody*[1], *B.M.S.G. Banneheka*[2]**

[1, 2] *Department of Statistics, Faculty of Applied Sciences, University of Sri Jayewardenepura, Nugegoda, Sri Lanka.*

INTRODUCTION

Student Evaluations of Teaching (SET) was introduced in the 1920s, for providing information to lecturers regarding students' perception of their teaching. Since then, the application of SET has diversified and today almost all universities use it extensively (Zabaleta, 2007). Understanding student needs is primarily done via student feedback forms (Gruber et al., 2010) and many researchers emphasize on the importance of making student experience and needs the focal point of higher education systems (Douglas and Douglas, 2006; Hill et al., 2003). In other words, SET is utilized to treat the student as a customer and thoroughly understand their needs. Özcan (2013) states that education systems must provide lecturers with feedback regarding their professional performance to avoid teacher and system failure and this is done via SET. Similarly, Murray (2005) discusses its positive impacts on various levels while Johnson (2000), Rowley (2003) discuss in detail its significant role played in quality management systems and internal quality assurance practices within higher education institutes. Hence, SET plays a critical role in higher education. The state universities in Sri Lanka too use student feedback forms to fulfil numerous objectives. They operate under the guidance of quality assurance standards defined by the University Grants Commission (UGC) which has contributed to the fast-paced growth of the State university system. The handling of SET is one of the 11 elements of quality assurance standards defined by the UGC (UGC Sri Lanka, 2023). Furthermore, each university has its own set of guidelines on how to handle and use SET depicting the crucial role it plays in the current system. Additionally, it is more critical for Sri Lankan state universities as they operate under a free education system. Hence, maintaining the highest quality of education while awarding globally recognized degrees is critical.

Despite the significance of SET in the Sri Lankan higher education system, universities tend to use inappropriate measures to summarize student responses for a given question which leads to incorrect decision making and complete failure of the purpose served by SET. Many universities use mean scores to rank lecturers based on student feedback (McCullough and Radson, 2011). However, student responses for Likert scale type questions are in ordinal scale which is classified based on a natural order. Numbers assigned to ordinal data are an indication of its rank or level. More importantly it does not follow assumptions of normality and consistent spacing. Hence using the arithmetic mean as a summary measure is statistically meaningless. (Jenkins, 2021; Sermeus and Delesie, 1996). Furthermore, student response distributions are not always normally distributed. Supposing the mean is used to summarize responses that follow a skewed distribution, the true perception of students with respect to a given question/item will be misrepresented because of the influence of a few low responses on the mean (Linse, 2017). The incorrect choices regarding summary measures made by universities lead to the categorization of lectures who receive low ratings as ineffective, which is vehemently disregarded by Berk (2013). As a solution for this problem Altman (1991), Sermeus and Delesie (1996), Wilcoxon (1956) propose the median as a suitable measure of central tendency for SET data. Yet, these ideas have not been taken seriously and universities tend to inappropriately summarize student responses. Therefore, the objective of this study was to select the population characteristic and corresponding estimator that optimally summarizes the student responses for a given question. Three population characteristics; namely population mean, population median and population mode and their corresponding estimators are compared to achieve this objective.

METHODOLOGY

This study was based on the following assumptions:

1. Questions of student feedback forms are based on a 5-point Likert scale.
2. The response Y from a student for a given question is a rounded integer value of a random variable X that follows some continuous distribution on the support [1,5]. We use the truncated normal distribution, truncated at 1 and 5, in our illustration.

Since student responses are not always normally distributed, theoretical distributions that are flexible in terms of shape must be considered for modelling. The truncated normal distribution is highly flexible in terms of shape and can accurately capture the distribution of most student responses. Furthermore, the truncated normal distribution is used for quality improvement processes and SET is an instrument used to measure quality of education (Jeang, 1997; Kaupr and Cho, 1996). Hence, this study uses the truncated normal distribution truncated at 1 and 5 for simulations.

Sample mean and sample median are estimators of population mean and population median respectively while sample mode (based on frequency of responses) and adjusted sample mode (based on density of responses) are estimators of population mode.

The adjusted sample mode is a new estimator introduced by us. It is an adjustment made to the commonly used mode which is the observation with the highest frequency. The adjusted sample mode is the observation corresponding to the highest frequency after adjusting the frequencies at the two ends of the Likert scale. The idea behind it, is as follows: As mentioned above, the responses 1,2,3,4 and 5 are considered as rounded integer values of some continuous random variable X with support [1,5]. Accordingly, the response 1 represents the interval [1,1.5) of width 0.5 while responses 2,3,4 represent the intervals [1.5,2.5), [2.5, 3.5) and [3.5,4.5) respectively. Each interval has a width of 1. The response 5 represents the interval [4.5,5] of width 0.5. Therefore, if the frequencies in the above intervals are f1, f2, f3, f4 and f5 respectively, the corresponding densities are given by 2f1, f2, f3, f4 and 2f5. Consequently, we arrive at the adjusted sample mode by selecting the response with the highest density or the adjusted frequency. The following example illustrates the idea.'

Table 1: Illustration of an application of adjusted sample mode

Response	**1**	**2**	**3**	**4**	**5**
Class interval	[1,1.5)	[1.5,2.5)	[2.5, 3.5)	[3.5,4.5)	[4.5,5]
Frequency	20	36	32	9	3
Adjusted frequency (density)	40	36	32	9	6

According to Table 1 the sample mode and adjusted sample mode are 2 and 1 respectively.

Let θ_1, θ_2, θ_3 be the population mean, population median and population mode respectively and m_1, m_2, m_3, m_4 be the sample mean, sample median, sample mode and adjusted sample mode respectively. We explored the suitability of the above population characteristics and the corresponding estimators as summary measures of student responses, assuming that the responses are the rounded integer values of a truncated normal distribution with parameters μ and σ which represent the mean and the standard deviation of the corresponding un-truncated normal distribution. For the entire study $\sigma = 1$ was considered fixed and the μ values were determined when θ_1=1, 2, 3, 4, 5. Similarly μ values were determined when θ_2=1, 2, 3, 4, 5 and θ_3=1, 2, 3, 4, 5. Due to the non-linearity of the standard formulas corresponding to the mean and median of a truncated normal distribution, the Newton-

Raphson method was utilized to determine the μ values when θ_1 and θ_2 were controlled. The μ values thus determined can be seen in Table 2.
Student responses are anonymous and if a class consists of a small number of students, they may feel insecure, and their responses might be biased. The class size should be large enough for the responses to be authentic. This study selected the class size 50 as a practical minimum since there are more than 50 students in most of the general degree classes at present.

Table 2: μ values determined by controlling θ_1, θ_2 and θ_3

Population characteristic	Value				
	1	2	3	4	5
θ_1	NA	1.4832	3	4.5168	NA
θ_2	NA	1.6873	3	4.3127	NA
θ_3	1	2	3	4	5

Consider the case μ=1.4832| θ_1=2. Assuming that a class consists of 50 students, a sample of 50 observations were generated from a truncated normal distribution with μ = 1.4832 and σ = 1. Thereafter, m_1 the estimate of θ_1 was calculated. This was repeated with different samples of the same size, 10000 times. Using the 10000 simulated estimates, the Bias and the Root mean squared error (RMSE) of m_1 were estimated. This was repeated for μ values obtained when θ_1 = 1,3,4,5. The same process was repeated for μ values obtained when θ_2 = 1, 2, 3, 4, 5 and the Bias and RMSE of m_2 for each scenario was estimated. Similarly, μ values were obtained for θ_3 = 1, 2, 3, 4, 5 and the Bias and RMSE of m_3 and m_4 were estimated. Estimators are adopted because of their unbiasedness (Ezell and Land, 2005). Although, this study does not wish to introduce an unbiased estimator for summarizing SET, it strives to select an estimator that is relatively less biased in comparison to the interested estimators. The RMSE is a “general purpose error metric for numeric predictions” and it accounts for the variance of an estimator (Christie and Neil, 2021). Hence, the bias and RMSE estimates were used to evaluate estimator performance. All calculations and simulations were done using R software packages (R Core Team, 2022).

RESULTS AND DISCUSSION

Table 3 and Figure 1 show the results generated from the simulations obtained by controlling the values of θ_1, θ_2 and θ_3.

Table 3: RMSE and Bias estimates (correct to 4 decimal places) obtained via simulations

Population Characteristic	Estimator	Response				
		1	2	3	4	5
θ_1	Bias (m_1)	NA	-0.0200	0.0000	0.0200	NA
θ_2	Bias (m_2)	NA	0.0000	0.0000	0.0000	NA
θ_3	Bias (m_3)	0.7800	0.0800	0.0000	-0.0800	-0.7800
θ_3	Bias (m_4)	0.0700	-0.2000	0.0000	0.1900	-0.0700
θ_1	RMSE (m_1)	NA	0.0217	0.0011	0.0222	NA
θ_2	RMSE (m_2)	NA	0.0003	0.0007	0.0006	NA
θ_3	RMSE (m_3)	0.7801	0.0781	0.0014	0.0786	0.7801
θ_3	RMSE (m_4)	0.0738	0.1953	0.0011	0.1872	0.0738

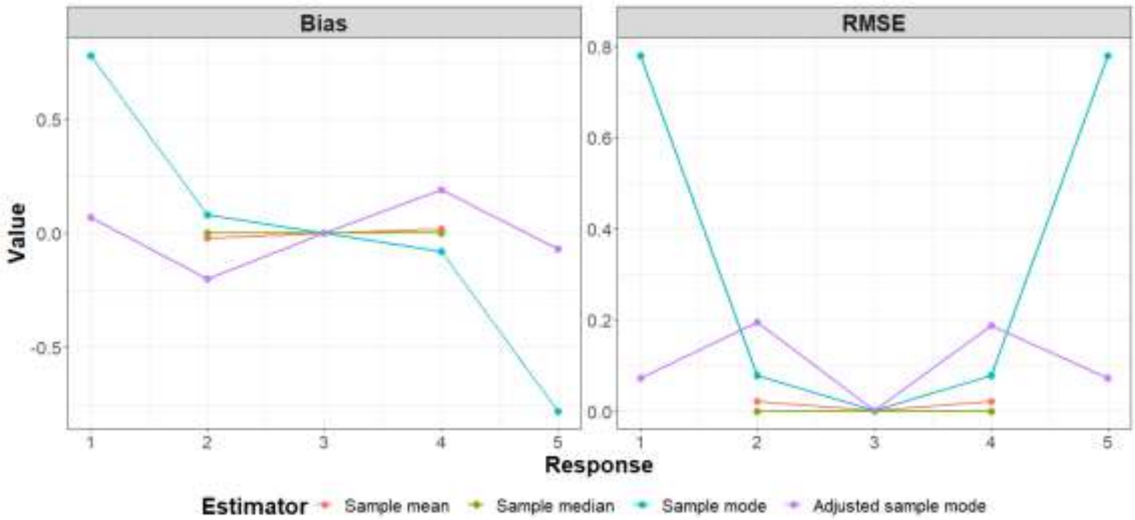

Figure 1: Change in Bias and RMSE by estimator

When the response Y is regarded as the rounded integer value of a continuous random variable X with support [1,5], the “average” of X may be considered as the “deserving value” for the item. The mean, median and mode are commonly used representations of the “average”. Out of these, which characteristic of X is most suitable for the representation of the “deserving value” and which estimator is the most suitable for estimating this average? The selected characteristic should be able to represent any value from 1 to 5.

Consider θ_1, the mean of X. The values 1 and 5 are impossible for θ_1 when X follows a truncated normal distribution, truncated at 1 and 5. Consider a lecturer with a deserving value of 1 for a given question. For θ_1 to be 1 the population must consist of responses less than and greater than 1. Similarly, for θ_1 to be 5 the population must consist of responses less than and greater than 5. However, this is practically impossible because a student only has 5 choices. Furthermore, it is theoretically impossible because the truncated normal distribution is truncated at 1 and 5. This, however, does not mean that a lecturer does not truly deserve a 1 or 5 for a given question. This implies that θ_1 cannot represent all possibilities. Therefore, we eliminate θ_1 even though its estimator, m_1 of the responses Y has the smallest bias and RMSE when $\theta_1 \neq 1,5$.

Consider θ_2, the median of X. The values 1 and 5 are impossible for θ_2 as well when X follows a truncated normal distribution, truncated at 1 and 5. As in the case of θ_1 for θ_2 to be 1, 50% of the responses must be less than 1 and 50% of the responses must be greater than 1. Similarly, for θ_2 to be 5, 50% of the responses must be less than 5 and 50% of the responses must be greater than 5. This is impossible practically and theoretically for the same reasons explained in the case of θ_1. Therefore, we eliminate the option θ_2 as well.

Consider θ_3, the mode of X. The values 1 and 5 are possible for θ_3 when X follows a truncated normal distribution, truncated at 1 and 5. Its estimator m_3 has the highest bias and RMSE when θ_3 = 1,5 even though it performs well when θ_3 = 2, 3, 4. The estimator m_4 of θ_3 is optimal since it works well (not too large and reasonably small for all values of θ_3 in the range [1,5]). Furthermore, m_4 provides more accurate estimates of θ_3 with higher probability than those provided by m_3. More importantly θ_3 can take any value from 1 to 5 unlike the θ_1 and θ_2 where values 1 and 5 are impossible or very unlikely.

Therefore, we recommend the use the mode θ_3 of X as the characteristic that can be estimated optimally and the adjusted sample mode m_4 as its estimator. Any value in the range [1,5] is possible for θ_3 and its estimator m_4 has a reasonably small bias and RMSE at all values of θ_3.

CONCLUSIONS/RECOMMENDATIONS

With the rejection of the population mean and population median we conclude that the population mode summarizes Likert scale SET for a given question optimally. Furthermore, the adjusted sample mode is selected as the optimal estimator for summarizing SET for a given question as it estimates all possibilities of student responses with a reasonably small bias and RMSE. Sample mean and sample median are rejected even though they have a smaller bias and RMSE amongst all estimators as they do not represent all possibilities of student responses. SET research is a highly researched study area with multiple schools of thought regarding its application. The use of inappropriate methods to analyze SET has been one of the root causes of such a division. However, it continues to stay relevant and is extensively used by many universities. Furthermore, it is the first time that such a study has been conducted with the objective of selecting the most optimal summary measure for SET data via data simulation from a known probability distribution. Given the significant role played by SET in the Sri Lankan state university system, we hope to introduce a better method to summarize student responses for a whole questionnaire. Currently we are in the process of continuing the same methodology for other skewed continuous probability distributions such as the beta distribution. Furthermore, we hope to conduct the same process for varying class sizes. Consequently, we will use the results to derive an overall performance index to quantify students' perception regarding a lecturer.

CONFLICT OF INTEREST STATEMENT

The authors do not have affiliations with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

REFERENCES

Altman, D. G. (1991). Practical statistics for medical research Chapman and Hall. *London and New York*.

Berk, R. A. (2013). *Top 10 flashpoints in student ratings and the evaluation of teaching: What faculty and administrators must know to protect themselves in employment decisions*. Stylus Publishing, LLC.

Christie, D., & Neill, S. P. (2021). Measuring and observing the ocean renewable energy resource. *Reference Module in Earth Systems and Environmental Sciences; Elsevier: Amsterdam, The Netherlands*.

Douglas, J., & Douglas, A. (2006). Evaluating teaching quality. *Quality in Higher education*, *12*(1), 3-13.

Gruber, T., Reppel, A., & Voss, R. (2010). Understanding the characteristics of effective professors: The student's perspective. *Journal of Marketing for Higher Education*, *20*(2), 175-190.

Hill, Y., Lomas, L., & MacGregor, J. (2003). Students' perceptions of quality in higher education. *Quality assurance in education*, *11*(1), 15-20.

Jenkins, S. P. (2021). Inequality comparisons with ordinal data. *Review of Income and Wealth*, *67*(3), 547-563.

Jeang, A. (1997). An approach of tolerance design for quality improvement and cost reduction. *International journal of production research*, *35*(5), 1193-1211.

Johnson, R. (2000). The authority of the student evaluation questionnaire. *Teaching in Higher Education*, *5*(4), 419-434.

Kapur, K. C., & Cho, B. R. (1996). Economic design of the specification region for multiple quality characteristics. *IIE transactions*, *28*(3), 237-248.

Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in educational Evaluation*, *54*, 94-106.

McCullough, B. D., & Radson, D. (2011). Analysing student evaluations of teaching: Comparing means and proportions. *Evaluation & Research in Education*, *24*(3), 183-202.

Murray, H. G. (2005, June). Student evaluation of teaching: Has it made a difference. In *Annual meeting of the society for teaching and learning in higher education* (pp. 1-15).

Ezell, M. E., & Land, K. C. (2005). Ordinary least squares (OLS).

Özcan, K. (2013). Student evaluation of lecture and teaching effectiveness in higher education. *Educational Research and Reviews*, *8*(8), 378.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rowley, J. (2003). Designing student feedback questionnaires. *Quality assurance in education*, *11*(3), 142-149.

Sermeus, W., & Delesie, L. (1996). Ridit analysis on ordinal data. *Western Journal of Nursing Research*, *18*(3), 351-359.

University Grants Commission of Sri Lanka. (2023). *Codes of Practice*. Retrieved from https://www.eugc.ac.lk/qac/codes-of-practise.html

Wilcoxon, F. (1956). Nonparametric Statistics for Behavioral Sciences.

Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in higher education*, *12*(1), 55-76.