Pertanika J. Sci. & Technol. **19 (2)**: 285 – 291 (2011)
ISSN: 0128-7680

The Effects of Data Distribution on the Confidence Interval of the Number of Projected Items

Tan Yih Tyng*, Abdul Rahman Othman and Lai Choo Heng
School of Distance Education,
Universiti Sains Malaysia,
11800 Penang, Malaysia
E-mail: jessica.tan09@yahoo.com

ABSTRACT

Setting a question paper for test, quiz, and examination is one of the teachers' tasks. The factors that are usually taken into consideration in carrying out this particular task are the level of difficulty of the questions and the level of the students' ability. In addition, teachers will also have to consider the number of questions that have impact on the examination. This research describes a model-based test theory to study the confidence intervals for the projected number of items of a test, given the reliability of the test, the difficulty of the question, and the students' ability. Using the simulated data, the confidence intervals of the projected number of items were examined. The probability coverage and the length of the confidence interval were also used to evaluate the confidence intervals. The results showed that the data with a normal distribution, the ratio variance components of 4:1:5 and reliability equal to 0.80 gave the best confidence interval for the projected number of items.

Keywords: Confidence interval, reliability, normal distribution, chi-square distribution

INTRODUCTION

One of the teachers' tasks is to set question papers for tests, quizzes, and examination. However, several factors have to be taken into consideration when preparing the question papers. Among which are:

1. The difficulties of the questions
2. The total number of questions and time needed to answer them
3. Students' abilities

The level of the students' ability is taken into consideration, along with the difficulty level of the items as both will affect the reliability of the test. In addition to this, the number of items in a test paper must be appropriate as it is also capable of causing mental fatigue and lost of concentration. Hence, the main purpose of this research work was to study the confidence interval of the projected number of items (questions), n_b. This study used Generalizability Theory (G-theory) established by Cronbach, Rajaratnam & Gleser (1963) to assess the reliability of the test items. G-theory is made up of two parts or studies. The first part is known as the *generalizability study* (G study), while the second in called the *decision study* (D study). In the G study, the magnitudes of the factors affecting a particular test were estimated as variance components. These are the same variances from the random effects model of the multifactor analysis of variance (ANOVA). From these variance components the reliability of the test can be calculated. In contrast, the D study uses these variance components to improve the design of the test. One aspect of this is the number

Received: 29 January 2010
Accepted: 10 June 2010
*Corresponding Author

of the projected items, n_b, given certain reliability and magnitudes of the variance components of the items' difficulty and students' ability. In more specific, the effects of data distribution on the confidence interval of n_b were also investigated.

MATERIALS AND METHODS

According to Bernnan (2001), the G-theory can be explained using the statistical factorial design. It emphasizes on the estimation of variance components. The model used in this study was the random model with two effects crossed design without interaction.

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \tag{1}$$

Where, Y_{ij} is score obtained by the i^{th} student answering the j^{th} question, τ_i is the effect i^{th} student, β_j is the effect of the j^{th} item and ε_{ij} is the error term. Since each student was evaluated once, the student-question interaction effect was confounded in the error term. The effects in Equation (1) were random, hence their distributions were $\tau_i \sim N(0, \sigma_\tau^2)$, $\beta_j \sim N(0, \sigma_\beta^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_e^2)$. Based on the definition of the model, the variance for the observed score was partitioned into:

$$Var(Y_{ij}) = \sigma_\tau^2 + \sigma_\beta^2 + \sigma_e^2 \tag{2}$$

where, σ_τ^2, σ_β^2 and σ_e^2 are known as variance components. In this study, these variance components were estimated using the expected values for the mean sum of square for the effects, as shown below:

$$E(MSstudent) = \sigma_e^2 + b\sigma_\tau^2 \tag{3}$$

$$E(MSitem) = \sigma_e^2 + a\sigma_\beta^2 \tag{4}$$

$$E(MSE) = \sigma_e^2 \tag{5}$$

Based on Equations (3), (4) and (5), the variance components can be estimated using the following:

$$\hat{\sigma}_e^2 = MSE \tag{6}$$

$$\hat{\sigma}_\beta^2 = \frac{MSitem - MSE}{a} \tag{7}$$

$$\hat{\sigma}_\tau^2 = \frac{MSstudent - MSE}{b} \tag{8}$$

where, a is the number of students and b is the number of the original items. The generalizability coefficient $\rho_{n_b}^2$, which is based on n_b questions, is given by:

$$\rho_{n_b}^2 = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \dfrac{\sigma_e^2}{n_b}} \tag{9}$$

From Equation (9), n_b is written as:

$$n_b = \frac{\rho_{n_b}^2 \sigma_e^2}{(1 - \rho_{n_b}^2)\sigma_\tau^2} \tag{10}$$

The value of n_b can be obtained if the values of σ_τ^2, σ_e^2 and $\rho_{n_b}^2$ are known. Let this value be known as the target, n_b. The target n_b can be used to monitor the effectiveness of the confidence intervals of n_b. With reference to Burdick and Graybill (1992), the 100 (1-2α) % confidence interval for $\frac{\sigma_\tau^2}{\sigma_e^2}$ is given by:

$$\frac{L^* - 1}{b} < \frac{\sigma_\tau^2}{\sigma_e^2} < \frac{U^* - 1}{b} \tag{11}$$

With $L^* = \frac{MSstudent}{MSE(F_{\alpha;(a-1);(a-1)(b-1)})}$, $U^* = \frac{MSstudent}{MSE(F_{1-\alpha;(a-1);(a-1)(b-1)})}$

Hence, Equation (10) can be rewritten as:

$$\frac{\sigma_\tau^2}{\sigma_e^2} = \frac{\rho_{n_b}^2}{n_b(1 - \rho_{n_b}^2)} \tag{12}$$

By $\frac{\sigma_\tau^2}{\sigma_e^2}$ substituting from Equation (12) into Equation (11), the following confidence is obtained.

$$\frac{(L^* - 1)}{b} < \frac{\rho_{n_b}^2}{n_b(1 - \rho_{n_b}^2)} < \frac{(U^* - 1)}{b} \tag{13}$$

From Equation (13), the confidence interval 100(1-2α) % of n_b is obtained as shown in Equation (14):

$$\frac{\rho_{n_b}^2}{(1 - \rho_{n_b}^2)(U^* - 1)} < n_b < \frac{\rho_{n_b}^2}{(1 - \rho_{n_b}^2)(L^* - 1)} \tag{14}$$

In this study, the investigation on whether the confidence intervals of n_b remained constant at certain levels of $\rho_{n_b}^2$, σ_τ^2, σ_β^2 and σ_e^2 was carried out when the distribution of the data was varied. Since the variance components were supposed to be obtained from a testing scenario, the number of students, a = 30, and the number of items in their test, b = 6, were fixed in this study. The rationales for these choices include:

1. Normally, the class size in a Malaysian school is 30; and
2. The usual number of questions in Paper 2 of the Mathematics subject for the Unified Examination of Chinese (UEC) schools is 6, and the usual time allocated is two and half hours.

It is important to note that the UEC paper was chosen over Penilaian Menengah Rendah (PMR) or Sijil Pelajaran Malaysia (SPM) for the sake of convenience. This study can certainly be extended to Paper 2 of the Mathematics subject of these examinations which comprises of 20 and 15 items, respectively. Data in the form of 30 students by 6 items were simulated. Two types of data distributions were considered. They were the normal and chi-square distributions. These distributions were selected because they are additive in nature. Thus, the distribution type in preserved throughout Equation 1. The ratios of $\sigma_\tau^2 : \sigma_\beta^2 : \sigma_e^2$ were set at 1:1:8, 2:1:7, 3:1:6 and 4:1:5, respectively. The ratio for the variance component σ_τ^2 was gradually changed from 1 to 4 to reflect a shift from the homogeneous to the heterogeneous ability of the students. The variance component representing the item's difficulty σ_β^2 was set at the same level, i.e. at 1. A test is considered reliable when its calculated reliability value is greater than or equivalent to 0.80. In this study, the values of ρ^2= 0.80, 0.85, 0.90, and 0.95 were therefore considered.

Due to the unavailability of the real data, a student by item data matrix was simulated with the size 30 x 6. Each question had a maximum score of 10. The data matrix was generated in such a manner that $\mu = 5$ and the $\mathrm{var}(Y_{ij}) = 2.25$.

One of the approaches used to monitor confidence intervals is probability coverage. Efron & Tibshirani (1993) illustrated this by constructing many confidence intervals of a parameter and determined the probability coverage by counting the number of times the target parameter fell within the interval. Similarly, the same calculation was done for the confidence intervals of n_b. A count is consider to be *count left* if the value of target n_b falls to the left or less than the lower bound. If the target n_b falls within the interval, it is considered as *count in*. If the target n_b falls to the right or greater than the upper bound, it is taken as *count right*. In addition, there are possibilities of negative values for the interval. Therefore, a *count impossible* was allocated for the intervals with negative bounds. As for a 95% confidence interval, a count of 950 was expected if 1000 data sets were generated.

Another approach of evaluating the confidence interval of n_b is the length of the confidence interval, n_b. The shorter the length of the confidence interval, the better they become.

RESULTS AND DISCUSSION

Table 1 shows the probability coverage of 95% confidence intervals of n_b using generated from the normal and chi-square distributions.

Table 1 shows that the values of reliability coefficient ρ^2 in both the normal and chi-square distributions do not have any effect on probability coverage. In conditions where the ratios of component variance are kept constant, the numbers of *count in*, *count left* and *count right* are not affected by ρ^2. On the other hand, the ratios of variance components do affect probability coverage. The number of *count i*n would increase when the magnitude of the student variance component was increased at the expense of the error variance. Both distributions gave better results when the ratios of variance components were at 3:1:6 and 4:1:5, respectively. However, only normal data achieved 95% of the confidence interval at the ratio of 4:1:5, while the chi-square data only achieved about 87% of the confidence interval.

Table 2 displays the lengths of the confidence intervals of n_b for the data generated from the normal and chi-square distributions. The results showed that the length of the confidence interval is affected by both the ratio of variance component and the value of ρ^2. Both distributions gave shorter expected length of confidence intervals when the magnitude of student variance components increased against the error. When the value of ρ^2 was heightened, the expected length of the confident intervals also increased. The shortest expected length of confidence intervals was generated from the normal distribution. The shortest length was 10.37, which was evaluated at the ratio of 4:1:5 and ρ^2 equivalent to 0.80. For the chi-square distribution data, the shortest length was observed at 18.55 when the ratio was 4:1:5 and ρ^2 equivalent to 0.80.

CONCLUSIONS

Comparatively, the data generated from the normal distribution had higher probability coverage than those from the chi square distribution. This study has shown that when the variance components ratio is 4:1:5, the probability coverage for both the distributions reached the maximum values. Despite this, n_b will give a realistic value only when $\rho = 0.80$, i.e. a targeted number of items of 5. The length of the confidence interval of n_b is the shortest when the variance components ratio is 4:1:5 and $\rho = 0.80$ for both the distributions with the data from normal distribution being shorter. Conclusively, both the approaches in monitoring the confidence interval of n_b produced coherent results. The appropriate number of items in the existing format of test paper is adequate as it gives a high reliability.

TABLE 1
Results from the normal distribution data and chi-square distribution data for the probability of coverage

			Normal distribution				Chi-square distribution			
Ratio of component variance	ρ^2	Target $n_b = \frac{\rho_{nb}}{1-\rho_{nb}}\left(\frac{\sigma_e^2}{\sigma_\tau^2}\right)$	Count left	Count in	Count right	Count impossible	Count left	Count in	Count right	Count impossible
1:1:8	0.80	32	0	483	517	48	0	452	548	99
1:1:8	0.85	45	0	483	517	48	0	453	547	99
1:1:8	0.90	72	0	483	517	48	0	452	548	99
1:1:8	0.95	152	0	483	517	48	0	452	548	99
2:1:7	0.80	14	39	896	65	2	62	790	148	2
2:1:7	0.85	20	38	896	66	2	60	789	151	2
2:1:7	0.90	32	38	896	66	2	58	788	154	2
2:1:7	0.95	67	38	896	66	2	60	789	151	2
3:1:6	0.80	8	43	942	15	0	69	874	57	0
3:1:6	0.85	11	47	941	12	0	77	870	53	0
3:1:6	0.90	18	43	942	15	0	69	874	57	0
3:1:6	0.95	38	43	942	15	0	69	874	57	0
4:1:5	0.80	5	38	952	10	0	90	871	39	0
4:1:5	0.85	7	39	951	10	0	99	863	38	0
4:1:5	0.90	11	42	949	9	0	102	861	37	0
4:1:5	0.95	24	37	953	10	0	86	873	41	0

TABLE 2
Results from the normal distribution data and chi-square distribution data for the length of confidence interval

Ratio of component variance	ρ^2	Target $n_b = \frac{\rho_{nb}}{1-\rho_{nb}}\left(\frac{\sigma_e^2}{\sigma_\tau^2}\right)$	Normal distribution			Chi-square distribution		
			Estimated lower bound	Estimated upper bound	Estimated length of confidence interval	Estimated lower bound	Estimated upper bound	Estimated length of confidence interval
1:1:8	0.80	32	12.18	2428.93	2420.53	12.256	276.58	268.62
1:1:8	0.85	45	17.26	3440.98	3429.09	17.36	391.82	380.54
1:1:8	0.90	72	27.41	5465.08	5446.20	27.57	622.30	604.39
1:1:8	0.95	152	57.87	11537.40	11497.52	58.21	1313.74	1275.93
2:1:7	0.80	14	7.13	99.61	93.25	7.41	140.82	134.66
2:1:7	0.85	20	10.10	141.11	132.11	10.50	199.50	190.77
2:1:7	0.90	32	16.04	224.11	209.81	16.67	316.85	302.99
2:1:7	0.95	67	33.87	473.13	442.94	35.19	668.90	639.65
3:1:6	0.80	8	4.42	29.65	25.31	4.60	39.30	34.88
3:1:6	0.85	11	6.26	42.01	35.86	6.52	55.68	49.41
3:1:6	0.90	18	9.94	66.72	56.95	10.36	88.43	78.47
3:1:6	0.95	38	20.99	140.84	120.24	21.87	186.69	165.66
4:1:5	0.80	5	2.90	13.26	10.37	3.07	21.62	18.55
4:1:5	0.85	7	4.10	18.79	14.70	4.35	30.63	26.28
4:1:5	0.90	11	6.51	29.84	23.34	6.91	48.65	41.75
4:1:5	0.95	24	13.75	62.99	49.28	14.58	102.71	88.13

REFERENCES

Brennan, R.L. (2001). *Generalizability theory*. New York: Springer-Verlag.

Burdick, R.K., & Graybill, F.A. (1992). *Confidence intervals on variance components*. New York: Marcel Dekker, Inc.

Cronbach, L.J., Rajaratnam, N., & Gleser, G.C. (1963). Theory of generalizability: A liberalization of reliability theory. *The British Journal of Statistical Psychology*, *XVI* (2), 137-162.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.