J. Chosun Natural Sci.
Vol. 7, No. 3 (2014) pp. 208 – 213
http://dx.doi.org/10.13160/ricns.2014.7.3.208

Empirical Analysis on Rao-Scott First Order Adjustment for Two Population Homogeneity test Based on Stratified Three-Stage Cluster Sampling with PPS

Sunyeong Heo†

Abstract

National-wide and/or large scale sample surveys generally use complex sample design. Traditional Pearson chi-square test is not appropriate for the categorical complex sample data. Rao-Scott suggested an adjustment method for Pearson chi-square test, which uses the average of eigenvalues of design matrix of cell probabilities. This study is to compare the efficiency of Rao-Scott first order adjusted test to Wald test for homogeneity between two populations using 2009 Gyeongnam regional education offices's customer satisfaction survey (2009 GREOCSS) data. The 2009 GREOCSS data were collected based on stratified three-stage cluster sampling with probability proportional to size. The empirical results show that the Rao-Scott adjusted test statistic using only the variances of cell probabilities is very close to the Wald test statistic, which uses the covariance matrix of cell probabilities, under the 2009 GREOCSS data based. However it is necessary to be cautious to use the Rao-Scott first order adjusted test statistic in the place of Wald test because its efficiency is decreasing as the relative variance of eigenvalues of the design matrix of cell probabilities is increasing, specially more when the number of degrees of freedom is small.

Key words: J Categorical Data, Complex Sample Design, Design Effect, Rao-Scott First Order Adjustment, Wald Test

1. Introduction

National-wide and/or large scale surveys generally use several sampling methods together to select a representative sample, such as stratification, clustering, unequal probability sampling, multi-stage or multi-phase design, multi-frame sampling, and so on. This type of sampling is called complex sampling or complex sample design. Complex sample data do not satisfy the assumption of being independent and identically distributed (iid) which is generally required in traditional statistical theory.

Traditional Pearson chi-square tests are most often used for the tests of independence, homogeneity, and goodness-of-fit of categorical data. The Pearson chi-square tests require the conditions that data are iid with multi-normal distribution and the expected frequencies are so large that Pearson chi-square test statistics have asymptotically chi-square distribution. Therefore, when a sample data does not satisfy the iid condition and one uses Pearson chi-square test, the results of analyses can be severly distorted. Heo and Chung[1] showed the importance of analysis reflecting the sample design through empirical study comparing traditional Pearson chi-square test and Wald test for two sample homogeneity test based on complex sample design.

In the case of using the secondary data or some statistical software for analysis of categorical data, it is often only to be usable or known the estimates of cell probabilities and their variances but not usable or unknown the their covariance matrix. In the occasion, Wald test, an unbiased test, is not applicable method. Holt *et al.*[2], Rao and Scott[3-5] and Tomas and Rao[6] have suggested a adjustment method for Pearson test when one only knows the variances of cell probabilities but does not know their covariance matrix.

This study compares the efficiency of Rao-Scott first order adjusted test to Wald test for homogeneity between two populations using 2009 Gyeongnam regional education offices's customer satisfaction survey (2009 GREOCSS) data. The 2009 GREOCSS data were

Department of statistics, Changwon National University, Changwon 641-773, Korea

†Corresponding author : syheo@changwon.ac.kr
(Received : July 30, 2014, Revised : September 1, 2014,
Accepted : September 25, 2014)

collected based on stratified three-stage cluster sampling with probability proportional to size (pps) at the selection of primary sample units. In Chapter 2, it is reviewed the Rao-Scott first order adjusted test. In Chapter 3, empirical analysis is given for two population homogeneity using the 2009 GREOCSS data for both test: Rao-Scott first order adjusted test and Wald test. In Chapter 4, conclusions from numerical results are given.

2. Rao-Scott First Order Adjustment

Assume that two independent samples of size n_1 and n_2 are selected according to given sampling designs, and there are K categories $C_1, C_2, \cdots, C_K$ with probability p_{ij} of which a element belongs to jth category in ith population ($i=1,2;\quad j=1,2,\cdots,K$). The null hypothesis for test of homogeneity is

$$H_0: p_1 = p_2 \ (= p)$$

where $p_i = (p_{i1}, \cdots, p_{i,K-1})^T$ with $\sum_{j=1}^{K} p_{ij} = 1$, and $p = (p_1, \cdots, p_{K-1})^T$ with $\sum_{j=1}^{K} p_j = 1$.

Assume also that for an estimator $\hat{p}_i = (\hat{p}_{i1}, \cdots, \hat{p}_{i,K-1})^T$ of $p_i (i=1,2)$ based on a given sample design, $n_i^{-1/2}(\hat{p}_i - p_i)$ follows asymptotically normal distribution $N_{K-1}(0, V_i)$. In addition assume that $v_{ij} = Var(\hat{p}_{ij})$ is the jth diagonal element of the covariance matrix V_i of $\hat{p}_i$ and its consistent estimator $\hat{v}_{ij} = \widehat{Var}(\hat{p}_{ij})$ is available. Under these assumptions, the test statistic applying $\hat{p}_i$ to the traditional Pearson chi-square test statistic for H_0 is

$$\begin{aligned} Q_P^2 &= \sum_{i=1}^{2}\sum_{j=1}^{K} n_i \frac{(\hat{p}_{ij} - \hat{p}_j)^2}{\hat{p}_j} \\ &= (\hat{p}_1 - \hat{p}_2)^T \left(\frac{\hat{P}}{n_1} + \frac{\hat{P}}{n_2} \right)^{-1} (\hat{p}_1 - \hat{p}_2) \end{aligned}$$

where

$\hat{p}_j = (n_1\hat{p}_{1j} + n_2\hat{p}_{2j})/n$ and $\hat{P} = \mathrm{diag}(\hat{p}) - \hat{p}\,\hat{p}^T$ with $\hat{p} = (\hat{p}_1, \cdots, \hat{p}_{K-1})^T$. The test statistics Q_P^2's asymptotic distribution is

$$Q_P^2 \approx \sum_{j=1}^{K-1} \delta_j Z_j^2$$

where $Z_j \sim N(0,1)$. The δ_j $(j=1,2,\cdots,K-1)$ are eigenvalues of $\tilde{n} V^{1/2} P^{-1} V^{1/2}$, $V = Var(\hat{p}_1 + \hat{p}_2)$, $P = diag(p) - pp^T$, and $\tilde{n} = (n_1 + n_2)/n$.

The mean and variance of $\sum \delta_j Z_j^2$ are

$$E\left(\sum_{j=1}^{K-1} \delta_j Z_j^2 \right) = (K-1)\,\bar{\delta} \quad \text{and}$$

$$Var\left(\sum_{j=1}^{K-1} \delta_j Z_j^2 \right) = 2\sum_j (\delta_j - \bar{\delta})^2 + 2(K-1)\bar{\delta}^2.$$

Therefore, the power of Q_P^2 is affected by the size and dispersion of δ_js.

Based on the fact that $\bar{\delta}$ can be estimated by only using the jth diagonal elements of $\hat{V}_i$, $\hat{v}_{ij} = \widehat{Var}(\hat{p}_{ij})$, Holt ect.[2], Rao and Scott[3-5], and Tomas and Rao[6] have suggested adjustments of Pearson test statistic Q_P^2. The first order adjusted test statistic is

$$Q_{rs}^2 = Q_P^2 / \bar{\hat{\delta}}.$$

where $\bar{\hat{\delta}}$ is obtained by using $\hat{V}$와 $\hat{P}$ in the place of V and P. The $\bar{\hat{\delta}}$ is

$$\bar{\hat{\delta}} = (K-1)^{-1}\tilde{n}\sum_{i=1}^{2}\sum_{j=1}^{K} (\hat{v}_{ij} / \hat{p}_j),$$

which is calculated only using $\hat{v}_{ij} = \widehat{Var}(\hat{p}_{ij})$, the variances of cell probabilities.

The mean and variance of the asymptotic distribution of Q_{rs}^2 are

$$E(Q_P^2/\bar{\delta}) = (K-1) \quad \text{and}$$

$$V(Q_P^2/\bar{\delta}) = 2(K-1)\left(1 + \frac{s_\delta^2}{\bar{\delta}^2}\right)$$

where $s_\delta^2 = (K-1)^{-1}\sum_{j=1}^{K-1} (\delta_j - \bar{\delta})^2$. From the variance of the asymptotic distribution of Q_{rs}^2, it is obvious that the asymptotic distribution of Q_{rs}^2 generally has larger variance than χ_{K-1}^2. Consequently, the type I error of

Q_{rs}^2 can be larger than the nominal level of type I error.

The size of increment of the variance depends on the size of relative variance of $\hat{\delta}_i$s, $s_{\hat{\delta}}^2 / \bar{\hat{\delta}}^2$, and the amount of inflation of type I error also depends on it.

Wald test statistic for test of homogeneity of two populations is

$$Q_W^2 = (\hat{p}_1 - \hat{p}_2)^T \left(\frac{\hat{V}_1}{n_1} + \frac{\hat{V}_2}{n_2} \right)^{-1} (\hat{p}_1 - \hat{p}_2).$$

The asymptotic distribution of Q_W^2 is chi-square distribution with $K-1$ degrees of freedom under H_0.

3. Empirical Analysis

Empirical analysis was conducted by applying to the 2009 Gyeongnam regional education offices's customer satisfaction survey (2009 GREOCSS) data. The 2009 GREOCSS was conducted to measure the level of satisfaction of students, parents and teachers about educational service offered by 20 Gyengnam regional education offices.

The 2009 GREOCSS data were collected by stratified three-stage cluster sampling with probability proportional to size. Gyeongnam was first stratified into 20 regions (10 cities and 10 counties) and the types of schools. For each stratum, sample schools were selected by probability proportional to the number of classes. On the second stage of sampling, one class per grade was randomly selected within each sample school. On the final stage, sample students within each sample class were randomly surveyed. Parents of sample students were surveyed, and about 10 teachers per sampled school were surveyed. One can refer to Heo and Chang[7] and Chung *et al.*[8] for details of the sample design of the 2009 GREOCSS. In the 2009 GREOCSS, the level of satisfaction of students and parents were measured for 6 interested domains and the teachers's satisfaction level for 5 domains. Each item in a domain was measured by 5-point scale. In this study, one item selected per each interested domain for analysis.

For two population homogeneity test, Gyeongnam was divided into two subpopulation; one is consisted of 10 cities and the other is 10 counties. The first subpopulation is called city and the second is called county from the subsequence. The test was conducted for homogeneity between city and county.

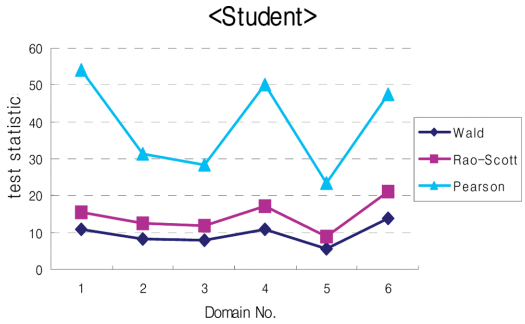

Fig. 1. Test statistics for two population homogeneity test between city and county by satisfaction domain for students.

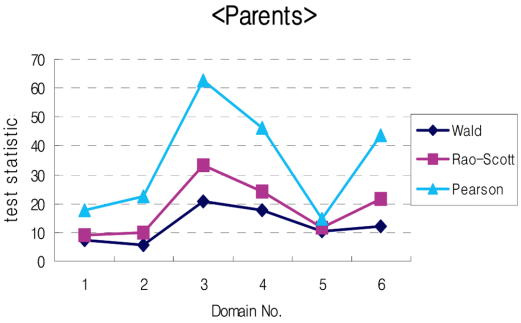

Fig. 2. Test statistics for two population homogeneity test between city and county by satisfaction domain for parents.

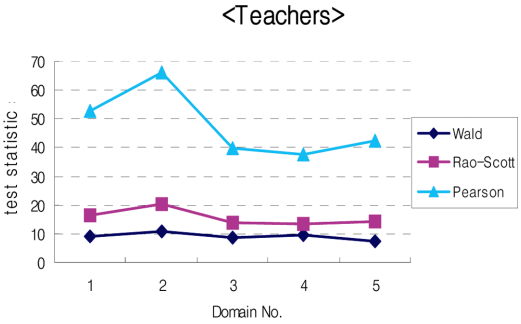

Fig. 3. Test statistics for two population homogeneity test between city and county by satisfaction domain for teachers.

Fig. 1~Fig. 3 show that three test statistics for homogeneity between city and county for students,

parents and teachers. Wald is the value of Wald test statistic Q_W^2, Rao-Scott first order adjusted Q_{rs}^2, and Pearson Q_P^2.

Fig. 1~Fig. 3 show that Rao-Scott first order adjusted test statistics give values very close to Wald test statistics, but the values of Pearson test statistic are much larger than both statistics. It is more clear when they are compared with numerical results in Table 1~Table 3.

Table 1~Table 3 give the numerical results of Wald test and Rao-Shao first order adjusted test, and summary of eigenvalues of design effect matrix. As reviewed in chapter 2, the variance of asymptotic distribution of Rao-Scott first order adjusted test statistic Q_{rs}^2 is $1+s_{\hat{\delta}^2}/\bar{\hat{\delta}}^2$ times grater than the variance of chi-square distribution with $K-1$ degrees of freedom, and hence it has longer tail than χ_{K-1}^2. From Table 1. and Table 2. which have the same number of categories, the ratio of p-value of Q_W^2 to p-value of Q_{rs}^2 is the largest when $1+s_{\hat{\delta}^2}/\bar{\hat{\delta}}^2$ is 1.799, the largest value. And the ratio is the smallest when $1+s_{\hat{\delta}^2}/\bar{\hat{\delta}}^2$ is 1.295, the second smallest value, but at which the ratio of test statistic Q_W^2 to Q_{rs}^2 is the smallest. From Table 1~Table 3 the ratio of two

Table 1. Values of Wald and Rao-Scott adjusted test statistics and summary of eigenvalues of design matrix for two population homogeneity test between city and county for student

Domain	No. of Cat.(K)	Q_W^2		Q_{rs}^2		Design effect		
		Test statistics	p-value	Test statistics	p-value	$\bar{\hat{d}}$	$s_{\hat{d}}$	$1+s_{\hat{\delta}}^2/\bar{\hat{\delta}}^2$
1	5	11.01	0.0265	15.35	0.0040	3.528	2.894	1.673
2	5	8.08	0.0887	12.54	0.0138	2.506	1.770	1.499
3	5	7.88	0.0961	11.86	0.0184	2.385	1.547	1.421
4	5	10.84	0.0284	17.08	0.0019	2.937	2.160	1.541
5	5	5.50	0.2397	8.87	0.0644	2.629	2.017	1.589
6	5	13.78	0.0080	21.24	0.0003	2.234	1.847	1.684

Table 2. Values of Wald and Rao-Scott adjusted test statistics and summary of eigenvalues of design matrix for two population homogeneity test between city and county for parents

Domain	No. of Cat.(K)	Q_W^2		Q_{rs}^2		Design effect		
		Test statistics	p-value	Test statistics	p-value	$\bar{\hat{d}}$	$s_{\hat{d}}$	$1+s_{\hat{\delta}}^2/\bar{\hat{\delta}}^2$
1	5	7.17	0.1272	9.14	0.0577	1.914	0.825	1.186
2	5	5.69	0.2235	9.99	0.0406	2.255	1.367	1.367
3	5	20.95	0.0003	33.46	0.0000	1.875	1.115	1.354
4	5	17.89	0.0013	24.29	0.0001	1.903	0.899	1.223
5	5	10.30	0.0357	11.47	0.0218	1.298	0.705	1.295
6	5	12.07	0.0168	21.68	0.0002	2.022	1.807	1.799

Table 3. Values of Wald and Rao-Scott adjusted test statistics and summary of eigenvalues of design matrix for two population homogeneity test between city and county for teachers

Domain	No. of Cat.(K)	Q_W^2		Q_{rs}^2		Design effect		
		Test statistic	p-value	Test statistics	p-value	$\bar{\hat{d}}$	$s_{\hat{d}}$	$1+s_{\hat{\delta}}^2/\bar{\hat{\delta}}^2$
1	4	8.92	0.0304	16.23	0.0010	3.239	2.510	1.601
2	4	10.74	0.0132	20.28	0.0001	3.251	2.767	1.724
3	5	8.81	0.0660	13.89	0.0077	2.858	2.387	1.698
4	4	9.32	0.0253	13.60	0.0035	2.776	2.172	1.612
5	5	7.44	0.1144	14.29	0.0064	2.957	2.604	1.775

p-values is the largest for $K = 4$ and $1 + s_{\hat{\delta}^2} / \bar{\hat{\delta}}^2 = 1.724$. From the Tables, it is shown that the efficiency of Rao-Scott first adjusted test to Wald test depends on both the number of categories and the relative variance of eigenvalues of design matrix of cell probabilities. In general, the p-value of Rao-scott first order adjusted test becomes smaller as the number of categories become smaller and the relative variances become larger.

On the other hand, it needs to be carful before making decision based on the p-value of Q_{rs}^2 which is near to the nominal level of significance α. For example, the domain 2 and 3 of students (Table 1) has $1 + s_{\hat{\delta}^2} / \bar{\hat{\delta}}^2$ equal to 1.499 and 1.421 respectively, and these values are not that large comparing to others. Their corresponding p-values of Q_{rs}^2 are 0.0136 and 0.0184 respectively, and one can make decision of reject H_0 under $\alpha = 0.05$. However, the p-values of Q_W^2 on these two domains are 0.0887 and 0.0961, and the null hypotheses are not rejected under $\alpha = 0.05$. It is similar at the domain 2 of parents (Table 2).

4. Conclusions

Today, it is in general using complex sample design. The categorical complex sample data generally do not satisfy the iid condition required for the traditional Pearson chi-square tests. If one uses the Pearson chi-square test for complex sample data, the results of analysis can be severly distorted.

For test of homogeneity, it is needed to know full covariance matrix of cell probabilities. However, when one uses the secondary data or some statistical softwares for analysis of categorical data, it is often the case only to be usable or known the estimates of cell probabilities and their variances, but not usable or unknown the their covariance matrix.

Holt *et al.*[2], Rao and Scott[3-5] and Tomas and Rao[6] have suggested an adjustment method of Pearson test for the situation, which is generally called Rao-Scott first order adjustment. The adjusted test statistic corrects Pearson test statistic using the average of eigenvalues of design matrix of cell probabilities, and the average of eigenvalues can be calculated by only using the variances of cell probabilities. The asymptotic distribution of the adjusted test statistic has the same mean as χ_{K-1}^2, but has the variance $1 + s_{\hat{\delta}^2} / \bar{\hat{\delta}}^2$ times grater than the variance of χ_{K-1}^2.

In this study, the efficiency of Rao-Scott first order adjusted test to Wald test for two population homogeneity is examined based on 2009 Gyeongnam regional education offices's customer satisfaction survey (2009 GREOCSS) data. The 2009 GREOCSS data was collected based on stratified three-stage cluster sampling with probability proportional to size (pps) at the selection of primary sample units.

The numerical analysis shows that the p-value of Rao-scott first order adjusted test generally becomes smaller as the number of categories becomes smaller and the relative variance larger. However, it needs to be carful before making final decision based on the p-value of Q_{rs}^2 even though the relative variance is not big if its p-value is a little smaller than the nominal level of significance α.

Acknowledgements

This research is financially supported by Changwon National University in 2013 ~ 2014.

References

[1] S. Heo and Y. Chung, “Effect of complex sample design on Pearson test statistic for homogeneity”, J. Korean Data Inform. Sci. Soc., Vol. 23, pp. 757-764, 2012.

[2] D. Holt, A. J. Scott, and P. D. Ewings, “Chi-squared tests with survey data”, J. R. Stat. Soc. A Stat., Vol. 143, pp. 302-320, 1980.

[3] J. N. K. Rao and A. J. Scott, “The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit the independence in two-way tables”, J. Am. Stat. Assoc., Vol. 76, pp. 221-230, 1981.

[4] J. N. K. Rao and A. J. Scott, “On chi-squared test for multiway contingency tables with cell proportions estimated from survey data”, Ann. Stat., Vol 12, pp. 46-60, 1984.

[5] J. N. K. Rao and A. J. Scott, “On simple adjustments to chi-square tests with sample survey

data", Ann. Stat., Vol. 15, pp. 385-397, 1987.

[6] D. R. Thomas and J. N. K. Rao, "Small-sample comparisons of level and power for simple goodness-of-fit statistics under cluster sampling", J. Am. Stat. Assoc., Vol. 82, pp. 630-636, 1987.

[7] S. Heo and D. Chang, "A sample survey design for service satisfaction evaluation of regional education offices". J. Korean Data Inform. Sci. Soc., Vol 21, pp. 671-678, 2010.

[8] Y. Chung, D, Jung, and S. Heo, "2009 Customer satisfaction evaluation survey to the service from Gyeongsangnam-do regional offices of education", Changwon National University, 2009.