J. of the Chosun Natural Science
Vol. 5, No. 4 (2012) pp. 241 – 245

Effect of Bias on the Pearson Chi-squared Test for Two Population Homogeneity Test

Sunyeong Heo†

Abstract

Categorical data collected based on complex sample design is not proper for the standard Pearson multinomial-based chi-squared test because the observations are not independent and identically distributed. This study investigates effects of bias of point estimator of population proportion and its variance estimator to the standard Pearson chi-squared test statistics when the sample is collected based on complex sampling scheme. This study examines the effect under two population homogeneity test. The standard Pearson test statistic can be partitioned into two parts; the first part is the weighted sum of χ_1^2 with eigenvalues of design matrix as their weights, and the additional second part which is added due to the biases of the point estimator and its variance estimator. Our empirical analysis shows that even though the bias of point estimator is small, Pearson test statistic is very much inflated due to underestimate the variance of point estimator. In the connection of design-based variance estimator and its design matrix, the bigger the average of eigenvalues of design matrix is, the larger relative size of which the first component part to Pearson test statistic is taking.

Key words: Bias, Complex Sampling Design, Pearson Test, Wald Test

1. Introduction

Goodness of fit test, homogeneity test, and independence test of categorical data with more than two categories, the standard Pearson chi-squared tests are used in general under the assumption of independency of observations. However, the survey data based on complex sample design is not satisfied with the condition of the standard Pearson multinomial-based chi-squared test. Complex sample design applies various sampling strategies together to select a representative sample, such as stratification, clustering, multi-stage or multi-phase, unequal probability, multi-frame and so on.[1] Today, many surveys are conducted based on complex sample design. Categorical survey data based on complex sample design does not fit the condition of the standard Pearson test and the use of it may cause wrong test results.

Holt et. al.[2] looked at the effect of correlations among elements on the performance of the standard Pearson chi-squared test. Through empirical analysis, they showed that the effect is sever for test of goodness-of-fit or homogeneity but less sever for test of independence. Rao and Scott[3-5] and Tomas and Rao[6] suggested test statistics adjusting the bias of the standard Pearson test statistics for the secondary data analysis collected based on complex sampling.

One of methods comparing more than two test statistics is to compare their power. Heo[7] showed through empirical power analysis that the Rao-Scott adjusted test to Pearson test has a tendency of moving upward comparing to unbiased Wald test. Recently, many surveys are conducted using complex sample design in the country. However, many survey reports are still published with results using the standard Pearson test.

This study investigates effects of bias of point estimator of population proportion and its variance estimator to the standard Pearson chi-squared test statistics for two population homogeneity test when the sample is collected based on complex sampling scheme. Chapter 2 reviews test statistics for two population homogeneity test and gives formula partitioned the standard Pearson test statistic into two parts. Chapter 3 gives the results of empirical analysis and Chapter 4 gives conclusions.

This study examines the effect under two population

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

†Corresponding author : syheo@changwon.ac.kr
(Received : September 18, 2012, Revised : December 15, 2012, Accepted : December 21, 2012)

homogeneity test. The standard Pearson test statistic can be partitioned into two parts; the first part is the weighted sum of χ_1^2 random variables with eigenvalues of design matrix as their weights, and the additional second part which is added due to the biases of the point estimator and its variance estimator.

Our empirical analysis shows that even though the bias of point estimator is small, Pearson test statistic is very much inflated due to underestimate the variance of point estimator. In the connection of design-based variance estimator and its design matrix, the bigger the average of eigenvalues of design matrix is, the larger relative size of which the first component part to Pearson test statistic is taking.

2. Homogeneity Test

2.1. Pearson Chi-squared Test

Suppose we have independent samples of size n_1 and n_2 from two populations, and each population is divided into K mutually exclusive categories with population proportions p_{ij} for ith population and jth category ($i = 1,2$, $j = 1,2,...,K$) and $\sum_{i=1}^{K} p_{ij} = 1$. The null hypothesis to be test for the homogeneity of population proportions is

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

where $p_i = (p_{i1},..., p_{i,K\text{-}1})^T$ and $p = (p_1,..., p_{K\text{-}1})^T$.

Under simple random sampling with replacement, the population proportion p_{ij} is estimated by $\hat{p}_{ij} = n_{ij}/n_i$, where n_{ij} is the observed cell counts for the jth category of the ith sample, and the ordinary Pearson test statistic is given by

$$X^2 = \sum_{i=1}^{2}\sum_{j=1}^{K} \frac{(n_{ij} - n_i\hat{p}_j)^2}{n_i\hat{p}_j}$$

where $\hat{p}_j = (n_{1j}+n_{2j})/n$ and $n = n_1+n_2$. Note that X^2 can be written in the quadratic form such as

$$X^2 = \tilde{n}(\hat{p}_1 - \hat{p}_2)^T \hat{P}^{-1}(\hat{p}_1 - \hat{p}_2) \qquad (2.1)$$

where $\tilde{n} = n_1n_2/n$ and $\hat{P} = \mathrm{diag}(\hat{p}) - \hat{p}\hat{p}^T$ with $p = (\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_{K-1})^T$. Under simple random sampling with replacement, the test statistic X^2 in (2.1) has asymptotically a chi-square distribution with K-1 degrees of freedom, χ_{K-1}^2, under H_0.

2.2. Wald Test

Now consider the more general sampling scheme. When the samples are selected according to complex sampling design, which uses a combination of stratification, clustering, multi-stage or multi-phase designs, unequal probability sampling, and so on, the asymptotic distribution of X^2 is no more χ_{K-1}^2 under H_0.

Suppose each population is stratified into L strata and $n_{hl}(\geq 2)$ clusters are selected with unequal probabilities for th stratum. Within lth first-stage cluster, $n_{hlm}(\geq 1)$ ultimate units are selected. This sampling scheme is slightly modified from the stratified multistage sampling design in Shao[8]. Then, the population proportion is estimated by

$$\hat{\pi}_{ij} = \hat{N}_{ij}/\hat{N}_i$$

where $\hat{N}_{ij} = \sum_{t \in s_i} w_{it}\delta_{tj}$ and $\hat{N}_i = \sum_{j=1}^{K} \hat{N}_{ij}$. Here, s_i is the set of ultimate units in the ith sample, $\delta_{tj} = 1$ if the tth ultimate unit in the ith sample belongs to the jth category and 0 otherwise, and w_{it} is the survey weights attached to the th unit in the th sample. In the above expression, a single subscript t is used instead of (hlm) for simplicity of the notation. Under some general conditions, $\hat{\pi}_{ij}$ is consistent estimator of p_{ij} (Shao, 1996). For $\hat{\pi}_i = (\hat{\pi}_{i1}, \ldots, \hat{\pi}_{i,K-1})^T$, if we can assume that $n_i^{-1/2}(\hat{\pi}_i - p_i)$ has asymptotically normal distribution $N_{K\text{-}1}(0, V_i)$ and $\hat{V}_i$ is consistent estimator of V_i, then the test statistic is

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

and X_w^2 is asymptotically χ_{K-1}^2 under H_0. The X_w^2 is called Wald test statistic.

2.3. Bias of Pearson Test

It follows from the quadratic form of Pearson test statistic in (2.1), that

$$X^2 = Q + E$$

where

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

and

$$E = 2\tilde{n}e^T\hat{P}^{-1}(\hat{\pi}_1 - \hat{\pi}_2) + \tilde{n}e^T\hat{P}^{-1}e,$$

where Z_j's are asymptotically independent $N(0,1)$, δ_j's

are eigenvalues of design effect matrix $D=(n_2D_1+n_1D_2)/n$ with $D_i=\hat{P}^{-1}\hat{V}_i$ (Holt et. al., 1980; Rao and Scott, 1981, 1984), and $e=(\hat{p}_1-\hat{\pi}_1)-(\hat{p}_2-\hat{\pi}_2)$. If $D_i=I$, then $d_j=1$ for all j and Q is asymptotically χ^2_{K-1}, and E becomes the estimator of bias of X. However, $\hat{P}\neq\hat{V}_i\,(i=1,2)$ in general in complex survey data, and the bias of Pearson test statistic X occurs under the above sampling design from both Q and E. From the following chapter, the amount of bias, which occurs due to Q and E, is examined through empirical analysis.

3. Empirical Analysis

3.1. Empirical Data

The empirical analysis was applied to the data of the 2009 education customers's satisfaction survey (2009 CSES) of Gyeongsangnam-do regional offices (GNE). The sampling design for the 2009 CSES of GNE is a stratified multistage sampling. The entire region of Gyeongsangnam-do was stratified into 40 strata according to 20 cities/counties, and level of school, elementary/middle school. Within each stratum, sample schools are selected proportional to the number of classes. Additional levels of sampling selected one class per grade level, and individual students. For the sample of teachers, 10 teachers are randomly selected from each sample school, and for the sample of parents the parents of the sampled students was surveyed. The numbers of respondents are 3,712 students, 3,456 parents, and 1,347 teachers.

Education satisfaction of students and parents are surveyed with 7 indices and job satisfaction of teachers are surveyed with 6 indices. Table 1 gives the number of items per index by student and teacher. For this study, the average per index was calculated and the average was categorized into 3 to 5 categories, which is given in Table 1.

3.2. Empirical Analysis

For two population homogeneity test, the population was divided into two sub-populations, 10 cities and 10 counties. The numbers of respondents were 2,647 students in cities and 1,065 students in counties, and 834 teachers in cities and 513 teachers in counties. For this study, only student data and teacher data were analyzed. Table 2 and Table 4 show Pearson test statistics and Wald test statistics for students and teachers. They also show the components of Pearson test statistics, Q and E. Table 3 and Table 5 show the averages and standard deviations of the absolute differences between biased estimator $\hat{p}_{ij}$ and the consistent estimator $\hat{\pi}_{ij}$ of the population proportion p_{ij}, and the averages and standard deviations δ_{ij} of eigenvalues of design effect matrix.

From Table 2 and Table 4, the Pearson Test statistics, which suppose simple random sampling and multinomial distribution, show much larger values than Wald test statistics, which reflects the complex sampling scheme. From Table 2 and Table 3, and Table 4 and Table 5, they show that the relative size of Q to X^2 is bigger when the averages of eigenvalues of design matrix is large than they are small. From Table 3 and Table 5, the averages and standard deviations of $|\hat{p}_{ij}-\hat{\pi}_{ij}|$ were obtained by calculating $|\hat{p}_{ij}-\hat{\pi}_{ij}|$ for th category and th population and taking average or calculating standard deviation for each index. From Table 3 and Table 5, they show that the average bias of point estimator $\hat{p}_{ij}$,

Table 1. The numbers of items and categories per index by students and teachers

Index No.	Education Satisfaction Index	Students		Teachers	
		No. of item	No. of cat.	No. of item	No. of cat.
1	School operation	8	4	11	3
2	Diversity of school education	10	4	10	3
3	Advancement of curriculum	8	4	10	4
4	Assistance for health and further continuing education	4	4	4	4
5	Building infrastructure for improving teacher's ability	3	5	3	4
6	Region's specialized business	2	5	3	4
7	Ability obtained from school education	5	5	-	-
	Total	40		41	

Table 2. Pearson and wald test statistics (students)

Index No	No. of cat.	Pearson Test			X_w^2
		X^2	Q	E	
1	4	34.87	31.55	3.28	7.26
2	4	71.42	68.80	2.52	16.53
3	4	28.23	29.16	-0.90	4.46
4	4	62.24	50.12	12.20	9.38
5	5	25.06	14.33	10.65	3.31
6	5	75.71	52.77	22.91	9.14
7	5	47.62	46.27	1.40	6.52

Table 3. Means and standard deviations of absolute values of $\hat{p}_{ij}-\hat{\pi}_{ij}$ and eigenvalues of design effect matrix (students)

Index No.	No. of cat.	$\lvert\hat{p}_{1j}-\hat{\pi}_{1j}\rvert$		$\lvert\hat{p}_{2j}-\hat{\pi}_{2j}\rvert$		Design Effect	
		Mean	S.D.	Mean	S.D.	$\bar{\delta}$	s_δ
1	4	0.0048	0.0037	0.0071	0.0059	3.065	2.538
2	4	0.0025	0.0006	0.0054	0.0040	3.997	2.961
3	4	0.0054	0.0037	0.0066	0.0055	4.023	2.672
4	4	0.0035	0.0028	0.0132	0.0100	3.793	2.926
5	5	0.0034	0.0031	0.0051	0.0053	2.512	2.367
6	5	0.0040	0.0038	0.0117	0.0111	3.441	2.991
7	5	0.0023	0.0007	0.0054	0.0034	4.962	3.713

Table 4. Pearson and Wald test statistics (teachers)

Index No	No. of cat.	Pearson Test			X_w^2
		X^2	Q	E	
1	3	34.19	55.48	-21.30	7.83
2	3	34.37	54.35	-19.99	7.69
3	4	21.51	37.04	-15.53	6.97
4	4	37.11	35.11	2.00	11.45
5	4	28.92	47.07	-18.15	7.21
6	4	45.93	63.44	-17.51	8.07

Table 5. Means and standard deviations of absolute values of $\hat{p}_{ij}-\hat{\pi}_{ij}$ and eigenvalues of design effect matrix (teachers)

Index No.	No. of cat.	$\lvert\hat{\pi}_{1j}-\hat{p}_{1j}\rvert$		$\lvert\hat{\pi}_{2j}-\hat{p}_{2j}\rvert$		Design Effect	
		Mean	S.D.	Mean	S.D.	$\bar{\delta}$	s_δ
1	3	0.0080	0.0056	0.0145	0.0072	4.264	3.857
2	3	0.0047	0.0034	0.0170	0.0140	4.430	3.808
3	4	0.0035	0.0034	0.0146	0.0058	4.443	2.699
4	4	0.0037	0.0031	0.0136	0.0105	2.693	2.732
5	4	0.0046	0.0030	0.0144	0.0100	3.961	3.632
6	4	0.0035	0.0032	0.0163	0.0069	3.723	3.634

when sample was selected based on complex sampling design, are relatively smaller effect on X^2 and does not have large effect on it than the design based variance estimator. Nevertheless, the effect of bias of $\hat{p}_{ij}$ to X^2 may not be ignored. Index 6 of $\lvert\hat{p}_{ij}-\hat{\pi}_{ij}\rvert$ students has relatively lager average of from the second sample and the inflation rate of X^2 relative to X_w^2 is almost 828% which gives the largest inflation rate.

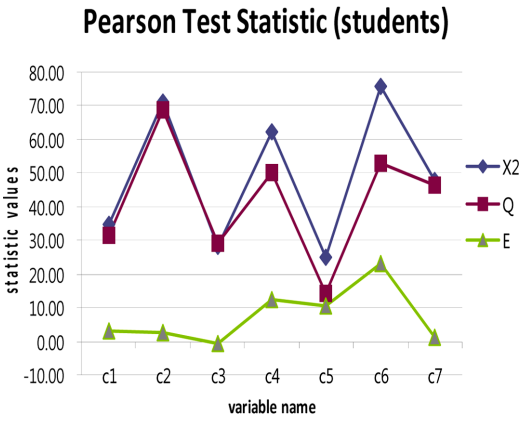

Fig. 1. Components of X^2 (students).

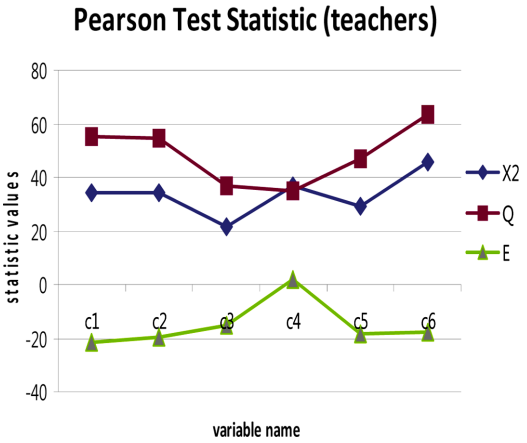

Fig. 2. Components of X^2 (teachers)

Fig. 1 and Fig. 2 show the pearson test statistics X^2, which is marked with diamond symbol, and the components of X^2, Q marked with sqaure and E marked with triangle.

4. Conclusion

Categorical data collected based on complex sample design is not proper for the standard Pearson multinomial-based chi-squared test because the observations are not independent and identically distributed. This study investigated the effect of biases of point estimator of population proportion and its variance estimator to the standard Pearson chi-squared test statistics for two population homogeneity test when the sample was collected by complex sampling scheme. the standard Pearson test statistic can be partitioned two parts; the first part is the weighted sum of χ_1^2 with eigenvalues of design matrix as weights, and the additional second part which is added due to the biases of the point estimator and its variance estimator. Our empirical analysis shows that even though the bias of point estimator is small, Pearson test statistic is very much inflated due to underestimate the variance. In the connection of design-based variance estimator and its design matrix, the bigger the average of eigenvalues of design matrix is, the larger relative size of which the first component part to Pearson test statistic is taking. Finally, when one analyses categorical data collected by complex sampling scheme, it needs to caution to use the Pearson chi-squared tests.

References

[1] P. J. Lavrakas, “Encyclopedia of survey research methods”, Sage, London, Vol. 2, p. 113, 2008.
[2] D. Holt, A. J. Scott, and P. D. Ewings, “Chi-squared tests with survey data”, J. the R. Stat. Soc. A, 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”, The Annals of Statistics, 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”, The Annals of Statistics, 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, “Power analysis of the Rao-Scott first-order adjustment to the Pearson test for homogeneity”, Joint Statistical Meetings Proceedings, Seattle, U.S.A., pp. 3126-3129, 2006.
[8] J. Shao, “Resampling methods in sample surveys (with discussion)”, Statistics, Vol. 27, 203-254, 1996.