Journal of Korean
Data & Information Science Society
2005, Vol. 16, No. 3, pp. 609~620

Test of Homogeneity Baseon Complex Survey Data : Discussion Based on Power of Test[1)]

Sunyeong Heo[2)] · Sucheol Yi[3)]

Abstract

In the secondary data analysis for categorical data, situations often arise in which the estimated cell variances are available, but not the full matrix of variances. In this case, researchers are often inclined to use Pearson-type test statistics for homogeneity. However, for a complex sample observed cell proportions are not distributed as multinomial and Pearson-type test statistic generally is not distributed asymptotically as chi-square distribution. This paper evaluates powers for Wald test and Pearson-type test and the first order corrected test of Pearson-type test for homogeneity. The resulting power curves indicate that as the misspecification effect increases, the amount of inflation of significance level and the loss of power Pearson-type test are getting more severe.

Keywords : Complex sample design, Homogeneity test, Misspecification effect, Power of test, Wald test

1. Introduction

For data from a complex sample design, standard Pearson multinomial-based chi-square tests generally do not achieve nominal levels of type I error. See, e.g., Holt, Scott and Ewings (1980), Rao and Scott (1981, 1984, 1987) and Thomas and Rao (1987) for general background on the effect of a complex sample design on the Pearson chi-squared test. Holt, Scott and Ewings (1980) and Rao and Scott (1981) suggested misspecification-effect-based adjustments to the usual Pearson-type chi-square test statistics to give asymptotically valid homogeneity

---

1) This research is financially supported by Changwon National University in 2003.

2) First Author : Assistant Professor, Dept. of Statistics, Changwon National University, Changwon, 641-773, Korea.
Email: syheo@sarim.changwon.ac.kr.

3) Associate Professor, Dept. of Applied Mathematics, Changwon National University, Changwon, 641-773, Korea.

tests based on data from a complex sample design.

This paper investigates the efficiency of the Wald test, Pearson-type chi-square test and its first order adjusted test in terms of power. The proposed methods are applied to the data from the Dual Frame the U. S. National Health Interview Survey (NHIS)/ Random-Digit-Dialing (RDD) Methodology and Field Test Project conducted by Research Triangle Institute (RTI) in North Carolina.

In section 2, I review some test statistics for homogeneity. In section 3, I evaluate the power for Wald test, Pearson-type test, and the first order adjusted test for Pearson-type test. In section 4, the proposed method is applied to the Dual Frame NHIS/RDD data.

2. Homogeneity Test

Suppose we have two independent populations and take independent samples of size n_1 and n_2 from these populations; and suppose we are interested in a categorical variable with K mutually exclusive categories $C_1, C_2, \cdots, C_K$, each observation in the ith sample is classified into one of these categories. Let n_{ik} be the observed cell frequencies for category C_k in the ith sample, and p_{ik} be the proportion of units for category C_k in the ith population $(\sum_{k=1}^{K} p_{ik} = 1)$; and let $p_i = (p_{i1}, p_{i2}, \cdots, p_{i,K-1})'$ be the vector of proportions in the first $K-1$ categories for the ith population. The hypothesis of homogeneity of the two populations is $H_0 : p_1 = p_2 (= p)$ against $H_1 : p_1 \neq p_2$, where p is an unknown vector.

2.1. Simple Random Sampling with Replacement

For simple random sampling with replacement, the vector $(n_{i1}, n_{i2}, \cdots, n_{i,K-1})$ follows a multinomial (n_i, p_i) distribution, and a customary Pearson test statistic is

$$X^2 = \sum_{i=1}^{2}\sum_{k=1}^{K} n_i (\hat{p}_{ik} - \hat{p}_{0k})^2 / \hat{p}_{0k}$$

where $\hat{p}_{ik} = n_{ik}/n_i$, $\hat{p}_{0k} = (n_1\hat{p}_{1k} + n_2\hat{p}_{2k})/(n_1 + n_2)$. This test statistics equals

$$X^2 = \tilde{n}(\hat{p}_1 - \hat{p}_2)'\hat{P}_0^{-1}(\hat{p}_1 - \hat{p}_2) \qquad (1)$$

where $\hat{p}_i = (\hat{p}_{i1}, \hat{p}_{i2}, \cdots, \hat{p}_{i,K-1})'$ for $i = 1, 2$ and $\hat{P}_0 = diag(\hat{p}_0) - \hat{p}_0\hat{p}_0'$ with $\hat{p}_0 = (\hat{p}_{01}, \hat{p}_{02}, \cdots, \hat{p}_{0,K-1})'$ and $\tilde{n} = n_1 n_2 / n$, $n = n_1 + n_2$. When n_1 and n_2 are large, X^2 is approximately distributed as a χ^2_{K-1} random variable under H_0 (Scott and Rao, 1981).

2.2. Complex Sampling

(1) A Wald-type test

Under a complex sample design, $n_i\hat{p}_i$ does not follow multinomial distribution, and under H_0 the asymptotic distribution of X^2 in expression (1) is not χ^2_{K-1}, chi-square distribution with $K-1$ degrees of freedom.

For a complex sample, let $\hat{\pi}_i = (\hat{\pi}_{i1}, \hat{\pi}_{i2}, \cdots, \hat{\pi}_{i,K-1})$ be a consistent estimator of p_i, and assume that $n_i^{1/2}(\hat{\pi}_i - p_i)$ converges in distribution to $N_{K-1}(0, V_i)$ as n_i increases $(i = 1, 2)$. When a consistent estimator $\hat{V}_i$ of V_i is available, the Wald test statistic

$$X^2_W = (\hat{\pi}_1 - \hat{\pi}_2)'(n_1^{-1}\hat{V}_1 + n_2^{-1}\hat{V}_2)^{-1}(\hat{\pi}_1 - \hat{\pi}_2) \tag{2}$$

is asymptotically distributed as χ^2_{K-1} under H_0 for sufficiently large n_i, $i = 1, 2$ (Holt, Scott and Ewings, 1980). When $K = 2$, X^2_W is expressed as

$$X^2_W = (\hat{v}_{1,11} + \hat{v}_{2,11})^{-1}(\hat{\pi}_{11} - \hat{\pi}_{21})^2 \tag{3}$$

where $\hat{\pi}_{i1}$ and $\hat{v}_{i,11} = \widehat{Var}(\hat{\pi}_{i1})$ are scalars.

For methods to obtain design-based consistent estimators of variances, see e.g., Krewski and Rao (1981) and Shao (1996).

(2) The first order correction for Pearson-type test statistics

In secondary data analysis, situations often arise in which the estimates of V_i are not available. In this cases, researchers often use a Pearson-type chi-squared test statistic for testing $H_0 : p_1 = p_2$,

$$X^2_P = \tilde{n}(\hat{\pi}_1 - \hat{\pi}_2)'\tilde{P}_0^{-1}(\hat{\pi}_1 - \hat{\pi}_2) \tag{4}$$

where $\tilde{P}_0 = diag(\hat{\pi}_0) - \hat{\pi}_0 \hat{\pi}_0{}'$ with $\hat{\pi}_0 = (\hat{\pi}_{01}, \hat{\pi}_{02}, \cdots, \hat{\pi}_{0,K-1})'$ and $\hat{\pi}_{0k} = (n_1 \hat{\pi}_{1k} + n_2 \hat{\pi}_{2k})/n$. When $K=2$, X_P^2 is expressed as

$$X_P^2 = \tilde{n} (\hat{\pi}_{11} - \hat{\pi}_{21})^2 / \{\hat{\pi}_{01}(1 - \hat{\pi}_{01})\}$$

where $\hat{\pi}_{i1}$ and $\hat{\pi}_{01}$ are scalars. Under a complex sample design, the distribution of $n_i \hat{\pi}_i$ is not multinomial and X_P^2 generally is not distributed asymptotically as χ^2_{K-1} under H_0.

In the secondary data analysis, there are often the cases of which the estimated cell variances, $\hat{v}_{i,kk}$, are avaliable. Holt, Scott and Ewings (1980) noted that under general design conditions and H_0, the test statistics X_P^2 is distributed asymptotically as a weighted sum, $\sum_{k=1}^{K-1} \lambda_k W_k$, of independent χ_1^2 random variable W_k. The weights λ_k are the singular values of $(n_2 D_1 + n_1 D_2)/n$ where $D_i = P_0^{-1} V_i$ for the ith population; and $P_0 = diag(\pi_0) - \pi_0 \pi_0{}'$ and the kth element of π_0 is $\pi_{0k} = (n_1 p_{ik} + n_2 p_{2k})/n$. They called D_i the design effect matrix and used it to develop a first-order adjustment for X_P^2. However, consistent with Skinner et al. (1989, p. 28) I will use term misspecification effect matrix for D_i.

Holt, Scott and Ewings (1980) suggested the simple correction of the test statistic X_P^2 such as

$$X_c^2 = X_P^2 / \bar{\hat{\lambda}} \tag{5}$$

where $\bar{\hat{\lambda}}$ is the average of $\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_{K-1}$, and the estimators, $\hat{\lambda}_k$, of λ_k are obtained by $V_i = \hat{V}_i$ and $P_0 = \tilde{P}_0$. Scott and Rao (1981) notes that $\bar{\hat{\lambda}}$ is calculated by

$$\begin{aligned} \bar{\hat{\lambda}} &= (K-1)^{-1} \tilde{n} \sum_{i=1}^{2} \sum_{k=1}^{K} (1 - \hat{\pi}_{0k})\ \hat{d}_{ik} / n_i \\ &= (K-1)^{-1} \tilde{n} \sum_{i=1}^{2} \sum_{k=1}^{K} \hat{v}_{i,kk} / \ \hat{\pi}_{0k} \end{aligned}$$

where $\hat{d}_{ik} = n_i \hat{v}_{i,kk} / \{\hat{\pi}_{0k}(1-\hat{\pi}_{0k})\}$ and $\hat{v}_{i,kk}$ is the kth diagonal element of $\widehat{Var}(\hat{\pi}_i)$. The $\overline{\hat{\lambda}}$ only requires a knowledge of cell variances $\hat{v}_{i,kk}$. Through empirical analysis, Holt, Scott and Ewings (1980) showed that X_c^2 is a very good approximation in terms of the nominal type I error rate. They also noted that the inflation of the type I error is getting bigger as the variabilities of λ_k's or K increase.

3. Power of Homogeneity Test

Under a complex sample design and for a consistent estimator, $\hat{\pi}_i$, of p_i, assume the following conditions.

(C.1) The asymptotic distribution of $\{n_1^{1/2}(\hat{\pi}_1 - p_1), n_2^{1/2}(\hat{\pi}_2 - p_2)\}'$ is $N_{2(K-1)}(0, V_p)$ where $V_p = blockdiag(V_1, V_2)$.

(C.2) There exist a consistent estimator, $\hat{V}_i$, of V_i.

3.1. Power of the Wald Test

Under conditions (C.1) and (C.2), by the Slutsky theorem (Casella and Berger, 1990, p. 220, Theorem 5.3.5)

$$\hat{V}_D^{-1/2}(\hat{\pi}_1 - \hat{\pi}_2) \xrightarrow{L} N_{K-1}(D, I)$$

where $\hat{V}_D = n_1^{-1}\hat{V}_1 + n_2^{-1}\hat{V}_2$, and $D = p_1 - p_2$. By Graybill (1976, p. 127, Corollary 4.2.1.4), the asymptotic distribution of the Wald test statistic X_W^2 in (2) is a non-central chi-square distribution $\chi_{K-1}^2(\delta)$ with $\delta = D'V_D^{-1}D/2$. Thus, the power of the Wald test is

$$\begin{aligned} 1-\beta_W &= \Pr\{X_W^2 > \chi_{K-1,\alpha}^2 \mid D\} \\ &\approx \Pr\{W_{K-1} > \chi_{K-1,\alpha}^2 \mid D\} \end{aligned}$$

where W_{K-1} is a non-central chi-square random variable $K-1$ degrees of freedom with non-centrality parameter $\delta = D'V_D^{-1}D/2$. When $K=2$, the power is

$$\begin{aligned} 1-\beta_W &= \Pr\{X_W^2 > \chi^2_{1,\alpha} \mid D\} \\ &\approx \Pr\{W_1 > \chi^2_{1,\alpha} \mid D\} \end{aligned} \tag{6}$$

where W_1 is a non-central chi-square random variable, $\chi^2_1(\delta)$, on one degree of freedom with $\delta = V_D^{-1}D^2/2 = (p_{11}-p_{21})^2/\{2(v_{1,11}+v_{2,11})\}$. When $H_0: p_1 = p_2$ is true, the W_1 and W_{K-1} have the central chi-square distributions and the Wald test statistic achieves asymptotically the nominal type I error rate α.

3.2. Power of the First Order Correction of X_P^2

Theorem: Assume conditions (C.1) and (C.2). For any non-zero $D = p_1 - p_2$, the Pearson-type test statistic X_P in expression (4) is asymptotically distributed as $\sum_{k=1}^{K-1}\lambda_k Y_k^2$ where Y_k^2 are independent non-central chi-square random variable with one degree of freedom and non-centrality parameter δ_k, the kth element of $D'V_D^{-1}D/2$; and $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_{K-1}$ are the ordered singular values of $\tilde{n}(V_D^{1/2})'P_0^{-1}V_D^{1/2}$.

From the above theorem, the power of X_P^2 is obtained by

$$\begin{aligned} 1-\beta_P &= \Pr\{X_P^2 > \chi^2_{K-1,\alpha} \mid D\} \\ &\approx \Pr\{\sum_{k=1}^{K-1}\lambda_k Y_k^2 > \chi^2_{K-1,\alpha} \mid D\}. \end{aligned}$$

When $K=2$, the power is

$$\begin{aligned} 1-\beta_P &= \Pr\{X_P^2 > \chi^2_{1,\alpha} \mid D\} \\ &\approx \Pr\{\lambda Y^2 > \chi^2_{1,\alpha} \mid D\} \end{aligned} \tag{7}$$

where $\lambda = \tilde{n}\{\pi_0(1-\pi_0)\}^{-1}(v_{1,11}+v_{2,11})$ and $\pi_{01} = (n_1 p_{11} + n_2 p_{21})/n$; and Y^2 is a non-central chi-square random variable on one degree of freedom with non-centrality parameter $\delta = V_D^{-1}D^2/2 = (p_{11}-p_{21})^2/\{2(v_{1,11}+v_{2,11})\}$.

Corollary: Under the conditions (C. 1) and (C. 2), $X_c^2 = X_P^2 / \bar{\hat{\lambda}}$ is distributed asymptotically as $\sum_{k=1}^{K-1} \left(\frac{\lambda_k}{\bar{\lambda}} \right) Y_k^2$.

Corollary is driven directly from the above Theorem. The $\bar{\lambda}$ is the average of $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_{K-1}$.

From the above corollary, the power of the first order correction of X_P is obtained by

$$\begin{aligned} 1 - \beta_c &= \Pr\{X_c^2 > \chi^2_{K-1,\alpha} \mid D\} \\ &\approx \Pr\{\sum_{k=1}^{K-1} \left(\frac{\lambda_k}{\bar{\lambda}} \right) {Y_k}^2 > \chi^2_{K-1,\alpha} \mid D\}. \end{aligned}$$

When $K = 2$,

$$\begin{aligned} 1 - \beta_c &= \Pr\{X_c^2 > \chi^2_{1,\alpha} \mid D\} \\ &\approx \Pr\{Y^2 > \chi^2_{1,\alpha} \mid D\}. \end{aligned} \tag{8}$$

Note that the random variable W_1 in expression (6) has the same distribution as Y^2 in expressions (7) and (8). Therefore, the asymptotic power of X_c^2 is the same as X_W^2 when $K = 2$.

4. Application to the U. S. Health Survey Data

4.1. Dual Frame NHIS/RDD Data

The data used for empirical analysis is Dual Frame NHIS/RDD data. This data collected from two states of the U. S. by Research Triangle Institute (RTI) in North Carolina. This data is composed of two groups. The first group is the persons who lived in both states and were interviewed face-to-face in NHIS and then re-interviewed by telephone (call it NHIS group). The second group is the persons selected using random digit dialing (RDD) from the persons not interviewed in the NHIS, and interviewed by telephone for both interviews (call it RDD group).

From the questionnaire used for the data, I selected question G1, "Are any firearms now kept in or around your home?", with possible responses "yes", or "no". The hypothesis in which I am interested $H_0: P(G1 = yes| State1) - P(G1 = yes| State2) = 0$. To evaluate power for the hypothesis, I used the original face-to-face NHIS responses from the NHIS group, and the first responses from the combined NHIS and RDD groups. I call the first data set as the NHIS data and the second as the combined data. I draw power curves only for the NHIS data, since the NHIS and combined data show very similar patterns.

[Table 1] summarize sample size, point estimates, their estimated standard errors, variances, and misspecification effects for the NHIS and combined data. The misspecification effects of the NHIS and combined data are equal to $\hat{\lambda} = 1.32$.

4.2. Power Evaluation

The power of the Wald test is estimated by

$$1 - \hat{\beta}_W = Pr\{ \widehat{W}_1 > \chi^2_{1,\alpha} | \, D = p_{11} - p_{21} \}$$

where $\widehat{W}_1 \sim \chi^2_1(\hat{\delta})$ and $\hat{\delta} = D^2/2(\hat{v}_{1,11} + \hat{v}_{2,11})$. The power of the Pearson-type test is estimated by

$$\begin{aligned} 1 - \hat{\beta}_P &= Pr\{\hat{\lambda}\widehat{W}_1 > \chi^2_{1,\alpha} | \, D = p_{11} - p_{21} \} \\ &= Pr\{ \widehat{W}_1 > \hat{\lambda}^{-1}\chi^2_{1,\alpha} | \, D = p_{11} - p_{21} \} \end{aligned}$$

where $\hat{\lambda} = \tilde{n}(\hat{v}_{1,11} + \hat{v}_{2,11})/\{\hat{\pi}_{01}(1 - \hat{\pi}_{01})\}$.

[Figure 1] shows powers of the Wald test and the Pearson-type test for the NHIS data. [Figure 1] shows inflation of type I error rate for the Pearson-type test (0.088) with the nominal significance level $\alpha = 0.05$.

For a given λ, the non-centrality parameter is estimated by

$$\tilde{\delta} = \frac{D^2}{2(\hat{v}_{1,11} + \hat{v}_{2,11})} = \frac{\tilde{n}D^2}{2\lambda\{\hat{\pi}_{01}(1 - \hat{\pi}_{01})\}}$$

and power is evaluated by

$$1 - \hat{\beta}_P = Pr\{ \widehat{W}_1 > \lambda^{-1}\chi^2_{1,\alpha} | \, D = p_{11} - p_{21} \}$$

where $\widehat{W}_1 \sim \chi_1^2(\tilde{\delta})$.

[Figure 2] shows powers for $\lambda = 0.8, 1.0, 2.0, 3.0$ from the Pearson-type test for the NHIS. When $\lambda = 0.8$, the Pearson-type test is conservative and has better power for all range of D. However, the bigger λ is, the more the significance level is inflated and the more the power is reduced, especially for area of large $|D|$.

5. Conclusion

In the secondary data analysis for categorical variable, situations often arise when the full matrix of variances of the estimates of proportions are not available, but the cell variance estimates are available. In this case researchers are inclined to use Pearson-type test statistics. However, for a complex sample observed cell proportions are not distributed as multinomial and Pearson-type test statistic generally is not distributed asymptotically as chi-square distribution. In previous researches, researchers suggest a method to correct Pearson-type test statistics for categorical data under a complex sample, e.g., Holt, Scott and Ewings (1980), Scott and Rao (1981) and Rao and Scott (1981).

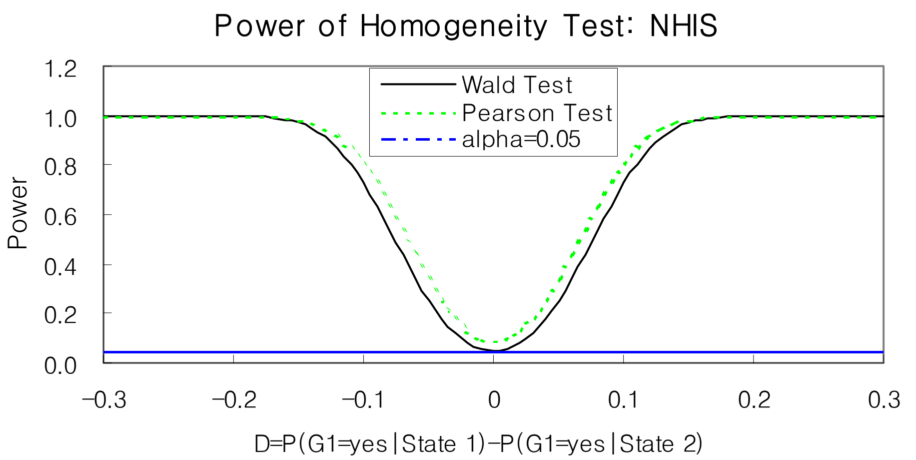
I searched for the way of evaluation of power for Wald test statistics and Pearson-type test statistics and its first order corrected test statistics for homogeneity test. The proposed methods were applied to data from the Dual Frame NHIS/RDD Methodology and Field Test Project conducted by Research Triangle Institute (RTI).

I presented power curves only for the NHIS data, since the power of the combined data shows very similar pattern as the NHIS data. The resulting power curves show that as the misspecification effect is getting bigger, the amount of inflation of significance level of Person-type test is getting more and the power is getting worse.

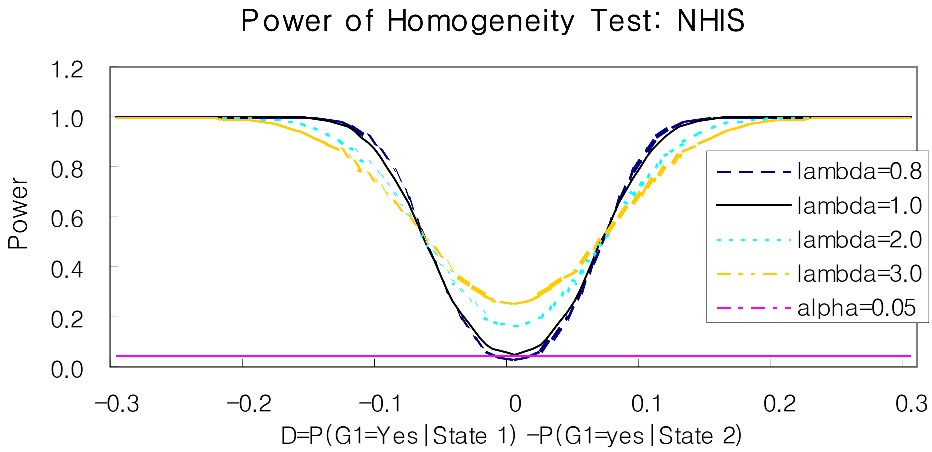
In this discussion, I assumed two independent populations. For power curves, I drew power curves for variable with only two categories. One can extend this idea and method for more than two populations and more than two categories.

[Table 1] Sample size, estimates of cell proportions, their standard errors and variances, and misspecification effects.

	state	n_i	$\hat{\pi}_{i1}$	$s.e.(\hat{\pi}_{i1})$	$\hat{v}_{i,11}$	$\hat{\lambda}$
NHIS	State 1	650	0.46927	0.02258	0.00051	1.32
	State 2	1,019	0.27980	0.01581	0.00025	
Combined	State 1	1,295	0.44733	0.01581	0.00025	1.32
	State 2	1,623	0.28003	0.01304	0.00017	



[Figure 1] Power of Wald test statistic and Pearson-type test statistic for test of homogeneity for NHIS data



[Figure 2] Power of Pearson-type test statistic when $\lambda = 0.8, 1.0, 2.0, 3.0$ for test of homogeneity for NHIS data

Appendix

Proof of Theorem: Under conditions (C.1) and (C.2), $\hat{\pi}_1 - \hat{\pi}_2$ is distributed asymptotically as $N_{K-1}(D, V_D)$ where $D = p_1 - p_2$; and $V_D = n_1^{-1} V_1 + n_2^{-1} V_2$ is a symmetric nonsingular matrix.

The Pearson-type test statistic in (5) can be expressed as

$$X_P^2 = [(\hat{\pi}_1 - \hat{\pi}_2)'(V_D^{-1/2})'][(V_D^{1/2})'\tilde{n}\tilde{P}_0^{-1}V_D^{1/2}][V_D^{-1/2}(\hat{\pi}_1 - \hat{\pi}_2)].$$

and $V_D^{-1/2}(\hat{\pi}_1 - \hat{\pi}_2)$ is distributed asymptotically as $N_{K-1}(V_D^{-1/2}(p_1 - p_2), I)$.

Let $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_{K-1}$ be the ordered singular values of $\tilde{n}(V_D^{1/2})'P_0^{-1}V_D^{1/2}$; and Q be a orthogonal matrix such as $\tilde{n}(V_D^{1/2})'P_0^{-1}V_D^{1/2} = Q\Lambda Q'$ and $\Lambda = diag(\lambda_1, \cdots, \lambda_{K-1})$. Then,

$$X_P^2 = Y'\Lambda Y = (\Lambda^{1/2}Y)'(\Lambda^{1/2}Y) \quad (*)$$

where $Y = Q'\{V_D^{-1/2}(\hat{\pi}_1 - \hat{\pi}_2)\}$; and Y is asymptotically distributed as $N_{K-1}(\mu_Y, I)$ with $\mu_Y = Q^{-1}\{V_\pi^{-1/2}(p_1 - p_2)\}$; and the asymptotic distribution of $\Lambda^{1/2}Y$ is $N_{K-1}(\Lambda^{1/2}\mu_Y, \Lambda)$. By Graybill (1976, p. 127, corollary 4.2.1.4) and Slutsky Theorem (Casella and Berger, 1990, p. 220, Theorem 5.3.5), X_P^2 is distributed asymptotically as $\sum_{k=1}^{K-1}\lambda_k Y_k^2$ where the Y_k^2's are asymptotically independent non-central chi-square random variables on one degree of freedom with non-centrality parameters δ_k's; and δ_k is the kth element of $D'V_D^{-1}D/2$.

References

1. Casella, G. and Berger, R. L. (1990). *Statistical Inference*. Pacific Grove, CA: Wadsworth & Brooks/Cole.
2. Graybill, F. A. (1976). *Theory and Application of the Linear Model*. Belmont, CA: Wadsworth.
3. Holt, D., Scott, A. J., and Ewings, P. D. (1980). Chi-squared tests with survey data. *Journal of the Royal Statistical Society, Series A*, 143, 302-320.
4. Krewski, D. and Rao, J. N. K. (1981). Inference from stratified samples : properties of the linearization, jackknife and balanced repeated replication methods. *The Annals of Statistics* 9, 1010-1019.
5. Rao, J. N. K. and Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit the independence in two-way tables. *Journal of the American Statistical Association* 76, 221-230.

6. Rao, J. N. K. and Scott, A. J. (1984). On chi-squared test for multiway contingency tables with cell proportions estimated from survey data. *The Annals of Statistics* 12, 46-60.
7. Rao, J. N. K. and Scott, A. J. (1987). On simple adjustments to chi-square tests with sample survey data. *The Annals of Statistics* 15, 385-397.
8. Scott, A. J. and Rao, J. N. K. (1981). Chi-squared tests for contingency tables with proportions estimated from survey data. In D. Krewski, R. Platek, and J. N. K. Rao (eds.), *Current Topics in Survey Sampling, pp. 247-265*. New York: Academic Press.
9. Shao, J. (1996). Resampling methods in sample surveys (with discussion). *Statistics* 27, 203-254.
10. Skinner, C. J., Holt, D., and Smith, T. M. F. (1989). *Analysis of Complex Surveys.* New York: John Wiley & Sons Ltd.
11. Thomas, D. R. and Rao, J. N. K. (1987). Small-sample comparisons of level and power for simple goodness-of-fit statistics under cluster sampling. *Journal of the American Statistical Association* 82, 630-636.

[received date : Apr. 2005, accepted date : May. 2005]