The Franklin's Randomized Response Model for Two Sensitive Attributes

Cheng C. Chen and Sarjinder Singh
Department of Mathematics, Texas A&M University-Kingsville, Kingsville, TX 78363, USA
E-mail: kfccc01@tamuk.edu

Abstract

In this paper, we suggest a new method to estimate the proportions of two non-overlapping sensitive attributes using the famous method of Franklin (1989: Communications in Statistics-Theory and Methods). The extension is based on the recent work of Singh and Chen (2009: Journal of Statistical Planning and Inference) where they suggest to utilize higher order moments of the scrambling variable to estimate single proportion of a sensitive attribute in survey sampling. The most important point is that here we use information from one sample to estimate proportions of two sensitive attributes. An attempt has been made to extend the proposed estimators to Bayesian and Empirical Bayes estimation techniques. Simulation studies are also performed to see the performance of the proposed estimators with their competitors.

Key Words: Franklin's randomized response model, Estimation of proportion, Sensitive attributes, Higher order moments.

1. Introduction

The problem of estimation of proportion of a potentially sensitive attribute in the literature of survey sampling has been very well addressed following the pioneer work of Warner (1965). The use of randomized response sampling in social, medical and environmental sciences has been well respected as one could refer to a monograph composed by Waltz *et al.* (2004) and a thesis by Blank (2008). Singh and Chen (2009) have introduced the idea of use of higher order moments of the scrambling variables to improve the estimate of single proportion without effecting the respondents' cooperation.

The problem of estimation of trinomial proportions has been found to be very useful especially during the election period in the United States of America. The voters in the US can be divided into three mutually exclusive groups: Democrat, Republican and Other. Although at this stage it does not appear to be a very serious problem if one discloses his/her privacy if he/she will prefer either of these three groups; however, it seems with time the competition for the presidential position is becoming more difficult due to several unavoidable reasons. There may be a situation that people will not feel safe while disclosing their preferences to vote in the US, and that day is probably not far. Assuming that the partition of voters will still remain trinomial because it may not be easy to establish a new competitor as strong as are democrats, republicans and others who are presently functioning. Here we develop a new method that could remain useful in such circumstances to organizations which currently conduct surveys about the prediction of future president for the US in the forthcoming elections. If required, the presently proposed model can easily be extended to the case of multinomial distribution as discussed at the end of this paper.

In the US and in Canada, a new gender, gay, is being introduced in the government sector. Thus, it will not take long until a person can be classified into three categories based on gender: Male, Female and Gay. The problem of estimation of proportion of true males, true females and true gays will remain a hot issue until gays do not feel comfortable reporting to others. It is obvious that some males and females could be considered as part of the 'gay' category, but we may be interested to estimate the proportions of true males, true females and true gays without disclosing their privacy in a city, in a province, or in a country. The proposed method could also be used in such situations. There are many other personal questions where three categories are feasible where the answer is: exactly known, exactly unknown and not sure, and hence leads to problem of trinomial proportions estimation.

In a careful examination of the literature in randomized response sampling as cited in Tracy and Mangat (1996) and later on, it has been observed that not much attention has been paid to estimate sensitive multinomial proportions. Abul-Ela *et al.* (1967) extended Warner (1965) design to multichotomous case when a population can be considered to be divided into t disjoint classes C_j with unknown proportions π_j ($j = 1,2,...,t$, $0 < \pi_j < 1$, $\sum \pi_j = 1$). It is assumed that at least one of the classes carries sigma and at least one carries no stigma. They suggested to take s $(= t-1)$ independent simple random with replacement samples of sizes n_i ($i = 1,2,..., s$, $\sum n_i = n$), and then a randomized response device is employed to each one of the samples. They examined, in detail, the extent of bias and the mean square error of the estimators for $t = 3$. Bourke and Dalenious (1973, 1974) proposed latin square measurement design to extend Warner's model to the multinomial case. Their design uses t different possible responses and requires only one sample. The respondent is asked to select one of the t-types of cards using a random device. Each of the t-mutually exclusive classes is described on each card, except that the order of the description is permuted from card to card. The permutation for t-cards form a latin square. The respondent reads the cards selected and report only the position of the card (i.e. $t = 1,2,\ldots$ $(t-1)$ or t) of the statement describing the class to which he/she belongs. The unrelated question design was also extended by Bourke (1974) to estimate the proportion of a population in each of t mutually exclusive classes of which $(t-1)$ are sensitive. One sample is needed if the distribution of the unrelated character is known; the design uses a deck of cards. Each of these cards contains a number of statements. The arrangement of the statements is a part of the design. Hochberg (1975) outlined an alternative scheme for estimating the t group proportions of which at the most $(t-2)$ are stigmatizing. The realizations for any sampled individuals a constitute two-stage scheme. The second stage is conditional on the random individual's response in the first stage. Drane (1976) used his "forced yes" stochastic model to estimate the proportion of more than one sensitive character. The use of supplemented block, balanced incomplete block and spring balance weighing designs were introduced by Raghavarao and Federer (1979). Their models allow the surveyor to obtain answers to several sensitive questions. Mukhopadhyay (1980), Mukherjee (1981), Tamhane (1981), Bourke (1981, 1982, 1990), Silva (1983) and Christofides (2003) have also considered the estimation of multi-attribute parameters.

Guerriero and Sandri (2007) reported that the family of models proposed by Kuk (1990) is better than the Simmons' family in terms of efficiency and privacy protection. From an empirical point of view, the study of van der Heijden *et al.* (2000) shows that Kuk's procedure seems to perform slightly better than the forced-response procedure and markedly better than face-to-face direct questioning and computer assisted self-interviewing. They also mentioned that the recommendations and successful applications of Kuk's procedure have been reported in van den Hout and van der Heijden (2002), and said that these results should be even more marked for the model proposed by Christofides (2003). In addition, they mentioned that an adequate analysis of the efficiency and the respondents' protection is always necessary when proposing new randomized response models. Thus, following Guerriero and Sandri (2007), it is worth to work further on Kuk (1990) and Franklin (1989) type models. Note that the Mangat (1994), Mangat and Singh (1990), Gjestvang and Singh (2006) and Kuk (1990) models are special cases of the Franklin (1989) model.

In the next section, thus we used the Franklin (1989) randomized response model to propose new estimators for estimating the proportions of three mutually exclusive, potentially sensitive characters which could be very useful in social, psychological, medical and environmental sciences.

2. Proposed Randomized Response Technique

In the proposed randomized response device, we say: if a person selected in the sample belongs to the first sensitive group A_1 then he/she is requested to draw a random number S_1 from a density function $f_1(s)$ and report to the interviewer; if he/she belongs to second sensitive group A_2 then he/she is requested to draw a random number S_2 from a density function $f_2(s)$ and report to the interviewer; and if he/she belongs to the third sensitive group A_3 then he/she is requested to draw a random number S_3 from a density function $f_3(s)$ and report to the interviewer. The respondent is further requested not to disclose the mode of response. Let Ω be the population under study. Obviously, $\Omega = \bigcup_{k=1}^{3} A_k$ and the groups A_k are mutually exclusive. The choice of the three densities $f_1(s)$, $f_2(s)$ and $f_3(s)$ are made such that respondents should feel safe in reporting the random number drawn by them. In other words, to keep the privacy of the respondents from all the three groups, the mean values and the variances of the three densities should not deviate too much from each other. Let π_1, π_2 and π_3 be the true proportions of persons that belong to groups A_1, A_2 and A_3 respectively such that $\pi_1 + \pi_2 + \pi_3 = 1$. Assume E denote the expected value over the proposed randomization response device, and let $\theta_1 = E(S_1)$, $\theta_2 = E(S_2)$, $\theta_3 = E(S_3)$, and $\gamma_{abc} = E(S_1 - \theta_1)^a (S_2 - \theta_2)^b (S_3 - \theta_3)^c$, where a, b and c are non-negative integers as required, are known moments of the three scrambling variables used in the proposed randomization device. Consider we selected a simple random with replacement sample (SRSWR) of n respondents. Interestingly, we show that based on only single sample information we propose three unbiased estimates of the three different parameters.

The distribution of the responses will be as follows:

$$Z_i = \begin{cases} S_1 \text{ with probability } \pi_1 \\ S_2 \text{ with probability } \pi_2 \\ S_3 \text{ with probability } \pi_3 \end{cases} \tag{2.1}$$

Obviously, we have

$$E(Z_i)=\pi_1\theta_1+\pi_2\theta_2+(1-\pi_1-\pi_2)\theta_3 \tag{2.2}$$

Following Singh and Chen (2009), we have:

$$Z_i^2=\begin{cases} S_1^2 \text{ with probability } \pi_1 \\ S_2^2 \text{ with probability } \pi_2 \\ S_3^2 \text{ with probability } \pi_3 \end{cases} \tag{2.3}$$

where $E(S_1^2)=\gamma_{200}+\theta_1^2$, $E(S_2^2)=\gamma_{020}+\theta_2^2$ and $E(S_3^2)=\gamma_{002}+\theta_3^2$.
Obviously, we have

$$E(Z_i^2)=\pi_1(\gamma_{200}+\theta_1^2)+\pi_2(\gamma_{020}+\theta_2^2)+(1-\pi_1-\pi_2)(\gamma_{002}+\theta_3^2) \tag{2.4}$$

Defining

$$\Delta=(\theta_1-\theta_3)\left\{(\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\right\}-(\theta_2-\theta_3)\left\{(\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\right\} \tag{2.5}$$

Then, we have the following theorems:

Theorem 2.1. An unbiased estimator of the population proportion π_1 is given by

$$\hat{\pi}_1=\frac{\left\{(\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}Z_i-\theta_3\right\}-(\theta_2-\theta_3)\left\{\frac{1}{n}\sum_{i=1}^{n}Z_i^2-(\gamma_{002}+\theta_3^2)\right\}}{\Delta} \tag{2.6}$$

Proof. Solving (2.2) and (2.4) for π_1 and using the method of moments, we have the theorem.

Theorem 2.2. An unbiased estimator of the population proportion π_2 is given by

$$\hat{\pi}_2=\frac{(\theta_1-\theta_3)\left\{\frac{1}{n}\sum_{i=1}^{n}Z_i^2-(\gamma_{002}+\theta_3^2)\right\}-\left\{(\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\right\}\left(\frac{1}{n}\sum_{i=1}^{n}Z_i-\theta_3\right)}{\Delta} \tag{2.7}$$

Proof. Solving (2.2) and (2.4) for π_2 and using the method of moments, we have the theorem.

Theorem 2.3. An unbiased estimator of the population proportion π_3 is given by

$$\hat{\pi}_3=1-\hat{\pi}_1-\hat{\pi}_2 \tag{2.8}$$

Proof. Obviously by taking expected values on both sides of (2.8).

Now we have the following lemmas:

Lemma 2.1. The variance of Z_i is given by

$$V(Z_i)=A_1\pi_1+A_2\pi_2+A_{11}\pi_1(1-\pi_1)+A_{22}\pi_2(1-\pi_2)+A_{12}\pi_1\pi_2+A_{00} \tag{2.9}$$

where $A_1=(\gamma_{200}-\gamma_{002})$, $A_2=(\gamma_{020}-\gamma_{002})$, $A_{00}=\gamma_{002}$, $A_{11}=(\theta_1-\theta_3)^2$, $A_{22}=(\theta_2-\theta_3)^2$ and $A_{12}=-2(\theta_1-\theta_3)(\theta_2-\theta_3)$.

Lemma 2.2. The variance of Z_i^2 is given by

$$V(Z_i^2) = B_1\pi_1 + B_2\pi_2 + B_{11}\pi_1(1-\pi_1) + B_{22}\pi_2(1-\pi_2) + B_{12}\pi_1\pi_2 + B_{00} \qquad (2.10)$$

where

$$B_1 = (\gamma_{400} - \gamma_{004}) + 4(\gamma_{300}\theta_1 - \gamma_{003}\theta_3) + 6(\gamma_{200}\theta_1^2 - \gamma_{002}\theta_3^2) + (\theta_1^4 - \theta_3^4) + (\gamma_{002} + \theta_3^2)^2 - (\gamma_{200} + \theta_1^2)^2$$
$$B_2 = (\gamma_{040} - \gamma_{004}) + 4(\gamma_{030}\theta_2 - \gamma_{003}\theta_3) + 6(\gamma_{020}\theta_2^2 - \gamma_{002}\theta_3^2) + (\theta_2^4 - \theta_3^4) + (\gamma_{002} + \theta_3^2) - (\gamma_{020} + \theta_2^2)^2$$
$$B_{11} = \left[(\gamma_{200} + \theta_1^2) - (\gamma_{002} + \theta_3^2)\right]^2, \quad B_{22} = \left[(\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right]^2$$
$$B_{12} = -2\left\{(\gamma_{200} + \theta_1^2) - (\gamma_{002} + \theta_3^2)\right\}\left\{(\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right\}$$

and $B_{00} = (\gamma_{004} + 4\gamma_{003}\theta_3 + 4\gamma_{002}\theta_3^2 - \gamma_{002}^2)$.

Lemma 2.3. The covariance between Z_i and Z_i^2 is given by

$$Cov(Z_i, Z_i^2) = C_1\pi_1 + C_2\pi_2 + C_{11}\pi_1(1-\pi_1) + C_{22}\pi_2(1-\pi_2) + C_{12}\pi_1\pi_2 + C_{00} \qquad (2.11)$$

where

$$C_1 = (\gamma_{300} - \gamma_{003}) + 3(\theta_1\gamma_{200} - \theta_3\gamma_{002}) + (\theta_1^3 - \theta_3^2) - \theta_3\{(\gamma_{200} + \theta_1^3) - (\gamma_{002} + \theta_3^2)\} - (\theta_1 - \theta_3)(\gamma_{200} + \theta_1^2)$$
$$C_2 = (\gamma_{030} - \gamma_{003}) + 3(\theta_2\gamma_{020} - \theta_3\gamma_{002}) + (\theta_2^3 - \theta_3^2) - \theta_3\{(\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\} - (\theta_2 - \theta_3)(\gamma_{020} + \theta_2^2)$$
$$C_{11} = (\theta_1 - \theta_3)\{(\gamma_{200} + \theta_1^2) - (\gamma_{002} + \theta_3^2)\}, \quad C_{22} = (\theta_2 - \theta_3)\{(\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\}$$
$$C_{12} = (\theta_2 - \theta_3)\{(\gamma_{200} + \theta_1^2) - (\gamma_{002} + \theta_3^2)\} + (\theta_1 - \theta_3)\{(\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\}$$

and $C_{00} = (\gamma_{003} + 2\theta_3\gamma_{002})$.

Now we have, without proves, the following theorems:

Theorem 2.4. The variance of the unbiased estimator $\hat{\pi}_1$ of the population proportion π_1 is given by

$$\begin{aligned}
V(\hat{\pi}_1) = \frac{1}{n\Delta^2}\Bigg[& \pi_1\left\{\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)^2 A_1 + B_1(\theta_2 - \theta_3)^2 - 2(\theta_2 - \theta_3)\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)C_1\right\} \\
& + \pi_2\left\{\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)^2 A_2 + B_2(\theta_2 - \theta_3)^2 - 2(\theta_2 - \theta_3)\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)C_2\right\} \\
& + \pi_1(1-\pi_1)\left\{\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)A_{11} + (\theta_2 - \theta_3)^2 B_{11} - 2(\theta_2 - \theta_3)\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)C_{11}\right\} \\
& + \pi_2(1-\pi_2)\left\{\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)A_{22} + (\theta_2 - \theta_3)^2 B_{22} - 2(\theta_2 - \theta_3)\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)C_{22}\right\} \\
& + \pi_1\pi_2\left\{\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)A_{12} + (\theta_2 - \theta_3)^2 B_{12} - 2(\theta_2 - \theta_3)\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)C_{12}\right\} \\
& + \left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)A_{00} + (\theta_2 - \theta_3)^2 B_{00} - 2(\theta_2 - \theta_3)\left((\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)\right)C_{00}\Bigg]
\end{aligned} \qquad (2.12)$$

Theorem 2.5.The variance of the unbiased estimator $\hat{\pi}_2$ of the population proportion π_2 is given by

$$
\begin{aligned}
V(\hat{\pi}_2) = \frac{1}{n\Delta^2}\Bigg[& \pi_1\Big\{ A_1\big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big)^2 + B_1(\theta_1-\theta_3)^2 - 2(\theta_1-\theta_3)\big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big)C_1 \Big\} \\
& + \pi_2\Big\{ \big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big)^2 A_2 + (\theta_1-\theta_3)^2 B_2 - 2(\theta_1-\theta_3)\big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big)C_2 \Big\} \\
& + \pi_1(1-\pi_1)\Big\{ \big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big)^2 A_{11} + (\theta_1-\theta_3)^2 B_{11} - 2(\theta_1-\theta_3)\big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big)C_{11} \Big\} \\
& + \pi_2(1-\pi_2)\Big\{ \big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big)^2 A_{22} + (\theta_1-\theta_3)^2 B_{22} - 2(\theta_1-\theta_3)\big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big)C_{22} \Big\} \\
& + \pi_1\pi_2\Big\{ (\theta_1-\theta_3)^2 B_{12} + \big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big)^2 A_{12} - (\theta_1-\theta_3)\big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big)C_{12} \Big\} \\
& + A_{00}\big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big)^2 + B_{00}(\theta_1-\theta_3)^2 - 2(\theta_1-\theta_3)\big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big)C_{00} \Bigg]
\end{aligned}
\tag{2.13}
$$

By using the lemmas, we have the theorem.

Theorem 2.6. The covariance between the unbiased estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ is given by

$$
\begin{aligned}
Cov(\hat{\pi}_1,\hat{\pi}_2) = \frac{1}{n\Delta^2}\Bigg[& \pi_1\Big\{ C_1\Psi - B_1(\theta_1-\theta_3)(\theta_2-\theta_3) - A_1\big((\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\big)\big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big) \Big\} \\
& + \pi_2\Big\{ C_2\Psi - B_2(\theta_1-\theta_3)(\theta_2-\theta_3) - A_2\big((\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\big)\big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big) \Big\} \\
& + \pi_1(1-\pi_1)\Big\{ C_{11}\Psi - B_{11}(\theta_1-\theta_3)(\theta_2-\theta_3) - A_{11}\big((\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\big)\big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big) \Big\} \\
& + \pi_2(1-\pi_2)\Big\{ C_{22}\Psi - B_{22}(\theta_1-\theta_3)(\theta_2-\theta_3) - A_{22}\big((\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\big)\big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big) \Big\} \\
& + C_{00}\Psi - B_{00}(\theta_1-\theta_3)(\theta_2-\theta_3) - A_{00}\Big\{(\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\Big\}\Big\{(\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\Big\} \Bigg]
\end{aligned}
\tag{2.14}
$$

where

$$\Psi = (\theta_1-\theta_3)\big((\gamma_{020}+\theta_2^2)-(\gamma_{002}+\theta_3^2)\big) + (\theta_2-\theta_3)\big((\gamma_{200}+\theta_1^2)-(\gamma_{002}+\theta_3^2)\big).$$

Theorem 2.7. The variance of the unbiased estimator $\hat{\pi}_3$ of the parameter π_3 is

$$V(\hat{\pi}_3) = V(\hat{\pi}_1) + V(\hat{\pi}_2) + 2Cov(\hat{\pi}_1, \hat{\pi}_2) \tag{2.15}$$

Proof. Obvious by the definition of variance.

In the next section, we compare the proposed estimators with a situation where the Warner (1965) estimator can be used three times instead of using the proposed randomization device.

3. Empirical Comparisons

It is natural that one could use the Warner (1965) model three times to estimate the three non-overlapping parameters π_k , $k = 1,2,3$. Each respondent selected in the sample could be provided with three randomization devices, say R_k , $k = 1,2,3$. The randomization R_k bears two types of statements, “Are you a member of group A_k ?,” and, “Are you a member of group A_k^c ?” with probabilities P_k and $(1 - P_k)$ respectively. Then, based on a sample of n respondents, if n_k reported ‘yes’ related to the k*th* group, then the unbiased estimator of π_k will be given by

$$\hat{\pi}_{k(w)} = \frac{n_k / n - (1 - P_k)}{2P_k - 1} \quad P_k \neq 0.5 \tag{3.1}$$

with variance given by

$$V(\hat{\pi}_{k(w)}) = \frac{\pi_k (1 - \pi_k)}{n} + \frac{P_k (1 - P_k)}{n(2P_k - 1)^2} . \tag{3.2}$$

The relative efficiency of the proposed estimator $\hat{\pi}_k$ with respect to the corresponding estimator $\hat{\pi}_{k(w)}$ due to Warner (1965) is given by

$$RE(k) = \frac{V(\hat{\pi}_{k(w)})}{V(\hat{\pi}_k)} \times 100\% \quad , \; k = 1,2,3 . \tag{3.3}$$

3.1 Discussion of the Results

We decided to keep $P_k = 0.7$, $k = 1,2,3$ (which is a very reasonable practical choice) in the Warner (1965) model while considering the problem of estimation of π_k with their respective estimators $\hat{\pi}_{k(w)}$ for $k = 1,2,3$. In the proposed model with three scrambling variables, we decided to make a very practical choice of the known parameters of the scrambling variables as: $\theta_1 = 57.3$, $\theta_2 = 65.2$, $\theta_3 = 60.3$, $\gamma_{200} = 6.25$, $\gamma_{020} = 16$, and $\gamma_{002} = 10$. By the three sigma empirical rule, the most of the values of the scrambling variables S_1, S_2 and S_3 could, respectively, be any real numbers in the ranges: (49.5, 64.8); (53.20, 72.30), and (50.81, 69.78). Due to an overlap between these three intervals, it will be hard to guess about the status of the respondents based on their reported responses. To study the effect of known higher order moments, such as skewness and kurtosis values, of the scrambling variables on the relative efficiencies $RE(k)$ of the proposed estimators we studied different values of $\gamma_3 = \gamma_{300} = \gamma_{030} = \gamma_{003}$ as –2, 0, 3, 5, 10 and 20; and the values of the $\gamma_4 = \gamma_{400} = \gamma_{040} = \gamma_{004}$ as 2, 3, 5 and 10. We retained minimum relative efficiency of 103% by assuming that minimum 3% gain is enough if one methodology could gain over the other without affecting the respondents' cooperation. The results so obtained are presented in Table 1 in the Appendix B. Note that while estimating rate attributes in two groups such as $\pi_1 = 0.1$, and $\pi_2 = 0.1$, then obviously $\pi_3 = 0.8$, then based on 24 observations, the average relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ remain as 308.0%, 693.1%, 492.0% with standard deviations 140.8%, 67.9% and 471.2% respectively as the values different parameters of the scrambling variables changes. The medians of the relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ remain as 249.2%, 703.6% and 285.8%. The minimum values of the relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ are 202.1%, 556.0% and 214.1% while the maximum values are 618.7%, 781.5% and 1708.7% respectively. Consider another situation when one of the attributes is rare, $\pi_1 = 0.1$, and the second attribute is moderate with $\pi_3 = 0.3$. Then, obviously $\pi_3 = 0.6$, and based on 24 observations, the average relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ remain as 333.3%, 254.1%, 144.9% with standard deviations 100.6%, 16.5% and 8.1% respectively as the values different parameters of the scrambling variables changes. The medians of the relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ remain as 293.2%, 257.4% and 143.1%. The minimum values of the relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ are 247.3%, 223.2% and 134.2% while the maximum values are 551.5%, 273.9% and 162.3% respectively. Consider another situation when all the three variables have moderate prevalence over the population as $\pi_1 = 0.3$, $\pi_2 = 0.3$ and $\pi_3 = 0.4$ then based on 14 observations, the average relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ remain as 231.1%, 242.7%, 113.9% with standard deviations 45.8%, 10.1% and 10.9% respectively as the values different parameters of the scrambling variables changes. The medians of the relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ remain as 219.6%, 243.5% and 111.3%. The minimum values of the relative efficiencies $RE(1)$, $RE(2)$ and $RE(3)$ are 187.9%, 227.5% and 103.1% while the maximum values are 301.5%, 254.7% and 131.1% respectively. In the same way, Table 1 in Appendix B could be read and interpreted. Note that in Table 1 the $RE(1)$, $RE(2)$ and $RE(3)$ for $\pi_1 = 0.1$, $\pi_2 = 0.1$ and $\pi_3 = 0.8$ are not the same as for $\pi_1 = 0.8$, $\pi_2 = 0.1$ and $\pi_3 = 0.1$ because of different choices of the mean and the variances of the scrambling variables for the three categories. The FORTRAN code used in the simulation are given in the Appendix, and all results could be reproduced, if required.

4. Generalization to the case of a Multinomial Distribution

Consider the population Ω consists of m mutually exclusive groups such that $\Omega = \bigcup_{k=1}^{m} A_k$.
Let π_k be the proportion of a sensitive attribute is the kth group. Then extending the proposed randomized response model from Section 2, that a respondent belonging to the kth group is requested to report a random number from the kth scrambling variable S_k for $k = 1,2,....,m$, then $(m-1)$ unbiased estimates of the population proportion π_k for $k = 1,2,..,(m-1)$ are given by

$$\hat{\Pi}_{(m-1)\times 1} = (A^{-1})_{(m-1)\times(m-1)} Z_{(m-1)\times 1} \tag{4.1}$$

where

$$\hat{\Pi} = \begin{bmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \\ \\ \hat{\pi}_{m-1} \end{bmatrix}, \; Z = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} Z_i - E(S_m) \\ \frac{1}{n}\sum_{i=1}^{n} Z_i^2 - E(S_m^2) \\ \\ \frac{1}{n}\sum_{i=1}^{n} Z_i^m - E(S_m^m) \end{bmatrix} \text{ and } A = \begin{bmatrix} E(S_1) - E(S_m), \ldots E(S_{m-1}) - E(S_m) \\ E(S_1^2) - E(S_m^2), \ldots. E(S_{m-1}^2) - E(S_m^2) \\ \\ E(S_1^m) - E(S_m^m), \ldots E(S_{m-1}^m) - E(S_m^m) \end{bmatrix}$$

and, the unbiased estimate of the proportion π_m is given by

$$\hat{\pi}_m = 1 - \sum_{k=1}^{m-1} \hat{\pi}_k \; . \tag{4.2}$$

Obviously, the variance of $\hat{\Pi}$ is given by

$$V(\hat{\Pi}) = (A^{-1}) V(Z) (A^{-1}) \tag{4.3}$$

and

$$V(\hat{\pi}_m) = \sum_{k=1}^{m-1} V(\hat{\pi}_k) + 2 \sum_{k<j=1}^{m-1} Cov(\hat{\pi}_k, \hat{\pi}_j) \tag{4.4}$$

where $V(Z)$ denotes the variance-covariance matrix of the scrambled responses which utilizes the higher order moments of the scrambling variables Z_i^l, $t = 1,2,3,....,m$.

5. Pseudo Bayes Estimators

Consider a trinomial distribution

$$p(\pi_1, \pi_2, \pi_3) = \frac{n!}{n_1! n_2! n_3!} \pi_1^{\alpha_1 - 1} \pi_2^{\alpha_2 - 1} \pi_3^{\alpha_3 - 1} \tag{5.1}$$

where α_i , $i = 1,2,3$ are the known priors.

Defining $D_1 = (\gamma_{200} + \theta_1^2) - (\gamma_{002} + \theta_3^2)$ and $D_2 = (\gamma_{020} + \theta_2^2) - (\gamma_{002} + \theta_3^2)$.

We consider a Pseudo Bayes' estimator of π_1 as

$$\hat{\pi}_1^b = \frac{1}{\Delta} \left[D_2 \left\{ \frac{\sum_{i=1}^{n} Z_i + \alpha_1}{n + A} - \theta_3 \right\} - (\theta_2 - \theta_3) \left\{ \frac{\sum_{i=1}^{n} Z_i^2 + \alpha_2}{n + A} - (\gamma_{002} + \theta_3^2) \right\} \right] \tag{5.2}$$

and a Pseudo Bayes' estimator of π_2 as

$$\hat{\pi}_2^b = \frac{1}{\Delta} \left[(\theta_1 - \theta_3) \left\{ \frac{\sum_{i=1}^{n} Z_i^2 + \alpha_2}{n + A} - (\gamma_{002} + \theta_3^2) \right\} - D_1 \left(\frac{\sum_{i=1}^{n} Z_i + \alpha_1}{n + A} - \theta_3 \right) \right] \tag{5.3}$$

where $A = \alpha_1 + \alpha_2 + \alpha_3$

Also we consider a Pseudo Bayes estimator of π_3 as

$$\hat{\pi}_3^b = 1 - \hat{\pi}_1^b - \hat{\pi}_2^b \tag{5.4}$$

Now we study the bias and mean squared error of the estimators $\hat{\pi}_i^b$, $i = 1,2,3$.

5.1. Bias and Variance expressions

The bias in the estimator $\hat{\pi}_i^b$ is given by

$$B(\hat{\pi}_1^b) = \frac{1}{n(1+A/n)}\left[\frac{D_2(\alpha_1 - A\theta_3) + (\theta_2 - \theta_3)\left\{A(\gamma_{002} + \theta_3^2) - \alpha_2\right\}}{\Delta} - A\pi_1\right] \tag{5.5}$$

$$B(\hat{\pi}_2^b) = \frac{1}{n(1+A/n)}\left[\frac{D_1(\theta_3 A - \alpha_1) + (\theta_1 - \theta_3)\{\alpha_2 - A(\gamma_{002} + \theta_3^2)\}}{\Delta} - A\pi_2\right] \tag{5.6}$$

and

$$B(\hat{\pi}_3^b) = -\left[B\left(\hat{\pi}_1^b\right) + B\left(\hat{\pi}_2^b\right)\right] \tag{5.7}$$

The variance of the estimator $\hat{\pi}_1^b$ is given by

$$V\left(\hat{\pi}_1^b\right) = \frac{1}{n(1+A/n)^2\Delta^2}\left[\pi_1 F_1 + \pi_2 F_2 + F_{11}\pi_1(1-\pi_1) + F_{22}\pi_2(1-\pi_2) + F_{12}\pi_1\pi_2 + F_{00}\right] \tag{5.8}$$

where

$F_1 = A_1 D_2^2 + B_1(\theta_2 - \theta_3)^2 - 2C_1 D_2(\theta_2 - \theta_3)$, $F_2 = A_2 D_2^2 + B_2(\theta_2 - \theta_3)^2 - 2C_2 D_2(\theta_2 - \theta_3)$

$F_{11} = A_{11} D_2^2 + B_{11}(\theta_2 - \theta_3)^2 - 2D_2 C_{11}(\theta_2 - \theta_3)$ $F_{22} = A_{22} D_2^2 + B_{22}(\theta_2 - \theta_3)^2 - 2D_2 C_{22}(\theta_2 - \theta_3)$

$F_{12} = A_{12} D_2^2 + B_{12}(\theta_2 - \theta_3)^2 - 2D_2 C_{12}(\theta_2 - \theta_3)$ and $F_{00} = A_{00} D_2^2 + B_{00}(\theta_2 - \theta_3)^2 - 2D_2 C_{00}(\theta_2 - \theta_3)$.

The variance of the estimator $\hat{\pi}_2^b$ is given by

$$V\left(\hat{\pi}_2^b\right) = \frac{1}{n(1+A/n)^2\Delta^2}\left[\pi_1 G_1 + \pi_2 G_2 + G_{11}\pi_1(1-\pi_1) + G_{22}\pi_2(1-\pi_2) + G_{12}\pi_1\pi_2 + G_{00}\right] \tag{5.9}$$

where

$G_1 = B_1(\theta_1 - \theta_3)^2 + A_1 D_1^2 - 2D_1 C_1(\theta_1 - \theta_3)$, $G_2 = B_2(\theta_1 - \theta_3)^2 + A_2 D_1^2 - 2D_1 C_2(\theta_1 - \theta_3)$

$G_{11} = B_{11}(\theta_1 - \theta_3)^2 + A_{11} D_1^2 - 2D_1 C_{11}(\theta_1 - \theta_3)$, $G_{22} = B_{22}(\theta_1 - \theta_3)^2 + A_{22} D_1^2 - 2D_1 C_{22}(\theta_1 - \theta_3)$

$G_{12} = B_{12}(\theta_1 - \theta_3)^2 + A_{12} D_1^2 - 2D_1 C_{12}(\theta_1 - \theta_3)$ and $G_{00} = (\theta_1 - \theta_3)^2 B_{00} + D_1^2 A_{00} - 2D_1 C_{00}(\theta_1 - \theta_3)$.

The variance of the estimator $\hat{\pi}_3^b$ is given by

$$V(\hat{\pi}_3^b) = V(\hat{\pi}_1^b) + V(\hat{\pi}_2^b) + 2Cov(\hat{\pi}_1^b, \hat{\pi}_2^b) \tag{5.10}$$

where the co-variance between the estimators $\hat{\pi}_1^b$ and $\hat{\pi}_2^b$ is given by

$$Cov\left(\hat{\pi}_1^b,\hat{\pi}_2^b\right)=\frac{1}{n(1+A/n)^2\Delta^2}\left[\pi_1 H_1+\pi_2 H_2+H_{11}\pi_1(1-\pi_1)+H_{22}\pi_2(1-\pi_2)+H_{12}\pi_1\pi_2+H_{00}\right] \tag{5.11}$$

where

$$H_1=C_1\{D_2(\theta_1-\theta_3)+D_1(\theta_2-\theta_3)\}-B_1(\theta_2-\theta_3)(\theta_1-\theta_3)-D_1D_2A_1$$
$$H_2=C_2\{D_2(\theta_1-\theta_3)+D_1(\theta_2-\theta_3)\}-B_2(\theta_2-\theta_3)(\theta_1-\theta_3)-D_1D_2A_2$$
$$H_{11}=C_{11}\{D_2(\theta_1-\theta_3)+D_1(\theta_2-\theta_3)\}-B_{11}(\theta_2-\theta_3)(\theta_1-\theta_3)-D_1D_2A_{11}$$
$$H_{22}=C_{22}\{D_2(\theta_1-\theta_3)+D_1(\theta_2-\theta_3)\}-B_{22}(\theta_2-\theta_3)(\theta_1-\theta_3)-D_1D_2A_{22}$$
$$H_{12}=C_{12}\{D_2(\theta_1-\theta_3)+D_1(\theta_2-\theta_3)\}-B_{12}(\theta_2-\theta_3)(\theta_1-\theta_3)-D_1D_2A_{12}$$

and

$$H_{00}=C_{00}\{D_2(\theta_1-\theta_3)+D_1(\theta_2-\theta_3)\}-B_{00}(\theta_2-\theta_3)(\theta_1-\theta_3)-D_1D_2A_{00}\,.$$

The mean square errors of the three estimators $\hat{\pi}_i^b$, $i=1,2,3$ are given by

$$MSE(\hat{\pi}_1^b)=V(\hat{\pi}_1^b)+\{B(\hat{\pi}_1^b)\}^2 \tag{5.12}$$

$$MSE(\hat{\pi}_2^b)=V(\hat{\pi}_2^b)+\{B(\hat{\pi}_2^b)\}^2 \tag{5.13}$$

and

$$MSE(\hat{\pi}_3^b)=V(\hat{\pi}_3^b)+\{B(\hat{\pi}_3^b)\}^2 \tag{5.14}$$

For different choices of priors, we consider the relative efficiency of the proposed Bayes estimators $\hat{\pi}_i^b$ with respect to the three unbiased estimators $\hat{\pi}_i$, $i=1,2,3$ as:

$$RE(i,b_i)=\frac{V(\hat{\pi}_i)}{MSE(\hat{\pi}_i^b)}\times 100\% \tag{5.15}$$

There could be a choice of priors, which could be searched by a computer grid method, such that the relative efficiency in (5.15) of the Pseudo Bayes estimates remains higher than 100%. We leave such an investigation, for future studies, in addition to Pseudo Empirical Bayes estimators in the next section.

6. Pseudo Empirical Bayes Estimators for Future Studies

In case α_i, $i=1,2,3$, are unknown, then we consider a Pseudo Empirical Bayes estimator of π_1 as:

$$\hat{\pi}_1^{eb}=\frac{1}{\Delta}\left[D_2\left\{\frac{\sum_{i=1}^{n}Z_i+\hat{\alpha}_1}{n+\hat{A}}-\theta_3\right\}-(\theta_2-\theta_3)\left\{\frac{\sum_{i=1}^{n}Z_i^2+\hat{\alpha}_2}{n+\hat{A}}-(\gamma_{002}+\theta_3^2)\right\}\right] \tag{6.1}$$

and a Pseudo Empirical Bayes' estimator of π_2 as

$$\hat{\pi}_2^{eb}=\frac{1}{\Delta}\left[(\theta_1-\theta_3)\left\{\frac{\sum_{i=1}^{n}Z_i^2+\hat{\alpha}_2}{n+\hat{A}}-(\gamma_{002}+\theta_3^2)\right\}-D_1\left(\frac{\sum_{i=1}^{n}Z_i+\hat{\alpha}_1}{n+\hat{A}}-\theta_3\right)\right] \tag{6.2}$$

where $\hat{A}=\hat{\alpha}_1+\hat{\alpha}_2+\hat{\alpha}_3$.

Also we consider a Pseudo Empirical Bayes estimator of π_3 as

$$\hat{\pi}_3^{eb} = 1 - \hat{\pi}_1^{eb} - \hat{\pi}_2^{eb} \tag{6.3}$$

We look forward to develop a method to estimate the priors α_i, $i = 1,2,3$, by using the scrambled responses considered in this paper.

References

Abul-Ela, A.L.A., Greenberg, B.G. and Horvitz, D.G. (1967). A multi-proportion randomized response model. *J. Amer. Statist. Assoc*., 62, 990-1008.

Blank, S.G. (2008). *Using the randomized response technique to investigate illegal fishing and contribute to abalone management in Northern California.* A 90 point thesis submitted to Victoria University of Wellington, as partial fulfillment for the degree of Master of Environmental Studies. School of Geography, Environment and Earth Sciences, Victoria University of Wellington.

Bourke, P.D. (1974). Multi-proportions randomized response using the unrelated question technique. *Report No. 74 of the Errors in Survey Research Project. Institute of Statistics, University of Stockholm (Mimeo).*

Bourke, P.D. (1981). On the analysis of some multivariate randomized response designs for categorical data. *J. Statistical Planning and Inference*, 5, 165-170.

Bourke, P.D. (1982). RR multivariate designs for categorical data. *Commun. Statist. – Theory Meth.*, A, 11, 2889-2901.

Bourke, P.D. (1990). Estimating a distribution function for each category of a sensitive variable. *Commun. Statist.-Theory Meth.*, 19(9), 3233-3241.

Bourke, P.D. and Dalenious, T. (1973). Multi-proportions randomized response using a single sample. *Report No. 68 of the Errors in Survey Research project. Institute of Statistics, University of Stockholm. (Mimeo).*

Bourke, P.D. and Dalenious, T. (1974). RR models with lying. *Technical Report –71, Institute of Statistics, University of Stockholm. (Mimeo)*

Christofides, T. C. (2003). A generalized randomized response technique. *Metrika*, 57(2), 195-200.

Drane, W. (1976). On the theory of randomized responses to two sentitive questions. *Comun. Statist.-Theory Methods,* A, 5, 565-574.

Franklin, L.A. (1989). A comparison of estimators for randomized response sampling with continuous distributions from a dichotomous population. *Commun. Statist*.- *Theory Meth*., 18 :489-505.

Gjestvang, C. R. and Singh, S. (2006). A new randomized response model. *Journal of the Royal Statistical Society , B*, 68 :523-530.

Guerriero, M. and Sandri, M.F. (2007). A note on the comparison of some randomized response procedures. *J. of Statist. Plann. and Inf.*, 137, 2184-2190.

Hochberg, Y. (1975). Two-stage randomized response scheme for estimating a multinomial. *Comuun. Statist.-Theory Methods*, 4, 1021-1032.

Kuk, A.Y.C. (1990). Asking sensitive questions indirectly. *Biomerika*, 77(2): 436-438.

Mangat, N.S. (1994). An improved randomized response strategy. *Journal of the Royal Statistical Society*, B, 56 :93-95.

Mangat, N.S. and Singh, R. (1990). An alternative randomized response procedure. *Biometrika*, 77(2) :439-442.

Mukherjee, R. (1981). Inference on confidential characters from survey data. *Calcutta Statistical Association Bull*, 30, 77-88.

Mukhopadhyay, P. (1980). On the estimation of some confidential characters from survey. *Calcutta Statistical Association Bull*., 29, 77-88.

Raghavarao, D. and Federer, W.T. (1979). Block total response as an alternative to the randomized response method. *J. R. Statist. Soc*., B, 41, 40-45.

Silva, L.C. (1983). On the generalized randomized response model with polychotomous variables. *Rev. Invest. Operac.*, 4, III, 75-100.

Singh, S. and Chen, C.C. (2009). Utilization of higher order moments of scrambling variables in randomized response sampling. *Journal of the Statistical Planning and Inference* (Article in press: JSPI 3860)

Tamhane, A.C. (1981). Randomized response techniques for multiple attributes. *J. Amer. Statist. Assoc.*, 76, 916-923.

Tracy, D.S. and Mangat, N.S. (1996). Some developments in randomized response sampling during the last decade - A follow up of review by Chaudhuri and Mukerjee. *J. Applied Statist. Sci.*, 4(2/3) ,147-158.

van der Heijden, P.G.M., van Gils, G., Bouts, J. and Hox, J.J. (2000). A comparison of randomized response, computer assisted self-interview, and face-to-face direct questioning. Eliciting sensitive information in the context of welfare and unemployment benefit. *Sociological Methods & Res.*, 28, 505-537.

van den Hout, A. and van der Heijden, P.G.M. (2002). Randomized response, statistical disclosure control and misclassification: a review. *Internat. Statist. Rev.*, 70, 269-288.

Waltz, C.F., Strickland, O.L. and Lenz, E.R. (2004). *Measurement in Nursing and Health Research*. 3rd Ed., Springer.

Warner, S.L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, *60: 63-69*

APPENDIX-A

!FORTRAN CODES USED IN THE SIMULATION STUDY

```
!      USE NUMERICAL_LIBRARIES
      IMPLICIT NONE
      INTEGER III,I,J,K,L
      DOUBLE PRECISION PI1, PI2, PI3, TH1, TH2, TH3,
     1G200, G020, G002, G300, G030, G003, G400, G040, G004,
     1DELTA,EZI,EZI2,EZI3, EZI4, VARZI,VARZI2,CZIZI2,
     1VARPI1, VPI1, RE1,VARPI2,VPI2,RE2, CPI1PI2,
     1VARPI3,VPI3,RE3,PW,API1(10),API2(10),
     1AG300(10),AG400(10)
      CHARACTER*20 OUT_FILE
      WRITE(*, '(A)') 'NAME OF THE OUTPUT FILE'
      READ(*,'(A20)') OUT_FILE
      OPEN (42, FILE=OUT_FILE, STATUS='UNKNOWN')
       DATA API1/0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9/
       DATA API2/0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9/
       DATA AG300/-2, 0, 3, 5, 10, 20/
       DATA AG400/2, 3, 5, 10/
       PW = 0.70
       TH1 = 57.3
       TH2 = 65.2
       TH3 = 60.3
       G200 = 6.25
       G020 = 16.0
       G002 = 10.0
       WRITE(42,102)TH1,TH2,TH3,G200,G020,G002,PW
102    FORMAT(2X,'TH1=',F5.1,1X,'TH2=',F5.1,1X,'TH3=',F5.1,1X,
     1 'G200=',F5.2,1X,'G020=',F5.1,1X,'G002=',F5.1,1X,'PW=',
     1 F6.3)
       DO 20 I = 1, 9
       DO 20 J = 1, 9
       PI1 = API1(I)
       PI2 = API2(J)
       PI3 = 1-PI1-PI2
       III=0
       IF (PI3.GT.0.01) THEN
       DO 10 K = 1, 6
       G300 = AG300(K)
       G030 = G300
       G003 = G300
       DO 10 L = 1, 4
       G400 =AG400(L)
       G040 = G400
       G004 = G400
       III = III+1
       WRITE(*,*)III
       DELTA = (TH1-TH3)*((G020+TH2**2)-(G002+TH3**2))
     1        -(TH2-TH3)*((G200+TH1**2)-(G002+TH3**2))
       EZI = PI1*TH1 + PI2*TH2 + PI3*TH3
       EZI2 = PI1*(G200+TH1**2)+PI2*(G020+TH2**2)+
     1 PI3*(G002+TH3**2)
       EZI3 = (G300+3*TH1*G200+TH1**3)*PI1
     1     +  (G030+3*TH2*G020+TH2**3)*PI2
     1     +  (G003+3*TH3*G002+TH3**3)*PI3
       EZI4 = (G400+4*G300*TH1+6*G200*TH1**2+TH1**4)*PI1
```

```
     1     +  (G040+4*G030*TH2+6*G020*TH2**2+TH2**4)*PI2
     1     +  (G004+4*G003*TH3+6*G002*TH3**2+TH3**4)*PI3
       VARZI  = EZI2-EZI*EZI
       VARZI2 = EZI4-EZI2*EZI2
       CZIZI2 = EZI3-EZI*EZI2
       VARPI1 = ((G020+TH2**2)-(G002+TH3**2))**2*VARZI
     1  +(TH2-TH3)**2*VARZI2
     1    - 2*((G020+TH2**2)-(G002+TH3**2))*(TH2-TH3)*CZIZI2
       VARPI1 = VARPI1/DELTA**2
       VPI1 = PI1*(1-PI1) +PW*(1-PW)/(2*PW-1)**2
       VARPI2=(TH1-TH3)**2*VARZI2
     1      +((G200+TH1**2)-(G002+TH3**2))**2*VARZI
     1      -2*(TH1-TH3)*((G200+TH1**2)-(G002+TH3**2))*CZIZI2
       VARPI2=VARPI2/DELTA**2
       VPI2 = PI2*(1-PI2) + PW*(1-PW)/(2*PW-1)**2
       CPI1PI2=(TH1-TH3)*((G020+TH2**2)-(G002+TH3**2))*CZIZI2
     1    -(TH2-TH3)*(TH1-TH3)*VARZI2
     1    -((G020+TH2**2)-(G002+TH3**2))
     1     *((G200+TH1**2)-(G002+TH3**2))*VARZI
     1    +(TH2-TH3)*((G200+TH1**2)-(G002+TH3**2))*CZIZI2
        CPI1PI2=CPI1PI2/DELTA**2
       VARPI3=VARPI1+VARPI2+2*CPI1PI2
       VPI3 = PI3*(1-PI3) + PW*(1-PW)/(2*PW-1)**2
        RE1  = VPI1*100/VARPI1
        RE2 = VPI2*100/VARPI2
        RE3 = VPI3*100/VARPI3
       IF((RE1.GE.103).AND.(RE2.GT.103).AND.(RE3.GE.103))THEN
       WRITE(42,101)PI1, PI2, PI3, G300, G030,G003, G400,G040,
     1  G004, RE1, RE2,RE3
       WRITE(*,101)PI1, PI2, PI3, G300, G030,G003, G400, G040,
     1  G004, RE1, RE2,RE3
101    FORMAT(2X,3(F7.3,1X),9(F6.1,1X),3(F9.2,1X))
       ENDIF
10     CONTINUE
       ENDIF
20    CONTINUE
      STOP
      END
```