International Journal of Statistics and Applications 2016, 6(3): 177-187
DOI: 10.5923/j.statistics.20160603.11

An Alternative Modified Item Count Technique in Sampling Survey

Farid Ibrahim

Department of Statistics, Benha University, Egypt

Abstract In this study we have proposed an alternative modification to the usual item count technique (ITC) to estimate the proportion of a sensitive characteristic in some of the fields such as health care. This technique produces two estimators to estimate the proportion. The first proposed estimator has been proven to be more efficient than the second one. Efficiency comparisons of the first proposed estimator with the estimator of Doitcour et al.'s (1991) ICT, and with the estimator of Hussain et al.'s (2012) ICT are performed. It is found that the first proposed estimator uniformly performs better than the other estimators. The optimal sample size n, in the case of minimizing the variance of the estimator, assuming that the cost of conducting the survey is fixed, will be determined.

Keywords Health surveys, Randomized response, Item count technique, Lagrange multipliers, Relative efficiency, Minimal variance

1. Introduction

In sampling surveys, it is easily conceivable that many respondents may not give truthful answers to direct questions on illegal behaviors such as habitual gambling, shoplifting, induced abortion, tax evasion, addiction to drugs, rash driving, history of past involvement in crimes, embarrassing or socially undesirable opinions and prejudices, … , etc. Measurement error in answers to previous sensitive topics may be reduced by some choices in survey design, such as open-ended questions, asking about behavior over long reference periods, tolerantly loaded introductions, and self-administration of the sensitive questions (see Tourangeau and Yan (2007)). An ingenious alternative approach to direct questioning is the randomized response technique (RRT) that was introduced by Warner (1965). In RRT the respondents employ a randomizing device to add probabilistic misclassification to their responses and conceal their true answers from the interviewer. Then several RR models have been proposed in the last decades as a valuable way for performing surveys on these sensitive topics. For details one can refer to Greenberg et al. (1969), Moors (1971), Mangat (1994), Mangat et al. (1997), Saha (2007), Singh and Tarray (2013), Barabesi et al. (2014), Adebola and Johnson (2015), and Blair et al. (2015).

Chaudhuri and Christofides (2007) gave a criticism on the randomized response technique in the sense that it demands the respondent's skill of handling the device and also asks respondents to report the information which may be useless or tricky. A clever respondent may also think that her/his reported response can be traced back to her/his actual status if she/he does not understand the mathematical logic behind the randomization device. Some of the alternatives to the randomized response technique are the Item Count Technique (ICT), the Nominative Technique and the Three Card Method. Details can be found in Droitcour et al. (1991), Droitcour and Larson (2002) and Miller (1985) respectively. These alternatives are designed because, in general, respondent evade sensitive questions especially regarding personal issues, socially deviant behaviors or illegal acts. Chaudhuri and Christofides (2007) also added that in these three alternatives to RRT respondents know that what they are revealing about themselves and that they do not need to know about any special estimation technique. Also respondents provide answers which make sense to them.

ICT has an impression, mainly, on sensitive fields such as health care. This Technique consists of selecting two independent samples (control and treatment groups). Each respondent in the control sample is presented with a list of innocuous items (control items), say g, with possible answers of "Yes" or "No" and is asked to report the total number of items that are applicable to her/him. Each respondent in the treatment sample, from the same population, is provided with the same list to which one sensitive item is added and is requested to report the total number of items that are applicable to her/him. The respondents are randomly assigned to either the control group or treatment group.

Compared ICT to the classical RRT, the ICT has the

* Corresponding author:
faridnaguib2003@hotmail.com (Farid Ibrahim)
Published online at http://journal.sapub.org/statistics

advantage of avoiding the potentially distracting act of randomization by the respondents themselves during the interview. Potential disadvantages of the ICT are that only the treatment group provides any information about the item of interest, and that the inclusion of the control items complicates the survey design and adds uncertainty to the estimation (see Kuha and Jackson (2014)).

Dalton et al. (1994) named ICT as the unmatched count technique and applied it to study the illicit behaviors of the auctioneers and compared to direct questioning they obtained higher estimates of six stigmatized items. Wimbush and Dalton (1997) applied this technique in estimating the employee theft rate in high-theft exposure business and found higher theft rates. Tsuchiya (2005) proposed two new methods, referred to as the cross-based method and the double cross-based method, by which proportions in subgroups or domains are estimated based on the data obtained via the item count technique. In order to assess the precision of the proposed methods, Tsuchiya conducted simulation experiments using data obtained from a survey of the Japanese national character. The results illustrated that the double cross-based method is much more accurate than the traditional stratified method, and is less likely to produce illogical estimates. Tsuchiya et al. (2007) conducted an experimental web survey in an attempt to compare the direct questioning technique and the ICT. Compared with the direct questioning technique, the ICT yielded higher estimates of the proportion of shoplifters by nearly 10 percentage points, whereas the difference between the estimates using these two techniques was mostly insignificant with respect to innocuous blood donation. Imai (2011) proposed new nonlinear least squares and maximum likelihood estimators for efficient multivariate regression analysis with the ICT.

Kuha and Jackson (2014) analyzed item count survey data on the illegal behavior of buying stolen goods. The analysis of an item count question was best formulated as an instance of modeling incomplete categorical data. They proposed an efficient implementation of the estimation which also provides explicit variance estimates for the parameters. Walter and Laier (2014) compared the methodological pros and cons of ICT to direct questioning (DQ). They presented findings from a face-to-face survey of 552 respondents who had all been previously convicted under criminal law prior to the survey. The results showed, first, that subjective measures of survey quality such as trust in anonymity or willingness to respond were not affected positively by ICT with the exception that interviewers feel less uncomfortable asking sensitive questions in ICT format than in DQ format. Second, all prevalence estimates of self-reported delinquent behaviors were significantly higher in ICT than in DQ format. Third, a regression model on determinants of response behavior indicated that the effect of ICT on response validity varies by gender. Overall, their results were in support of ICT.

Hussain and Shabbir (2010) and Hussain et al. (2012) proposed two modifications to the usual ICT that was proposed by Droitcour et al. (1991), and they showed that their estimators are always more efficient than the estimator of the usual ICT.

Since the two main problems of the randomized response models and their alternative techniques are the respondent's privacy and the efficiency of the estimators of these models, so in this paper we have proposed an alternative modification to the usual ICT that produces two estimators of the parameter of the item count model, and their properties are studied. The first proposed estimator, which has been proven to be more efficient than the second one, is more efficient than the estimators of the other models of ICT. Also this alternative modification provides full protection to the respondent's privacy.

So the remainder of the present research is organized as follows; Section 2 presents the usual ICT that was proposed by Droitcour et al. (1991) and Hussain et al.'s (2012) modification of the usual ICT. An alternative modification of the usual ICT, and the two estimators of the parameter of the model of ICT and their properties are presented in Section 3. The relative efficiency, of the two proposed estimators, is performed in section 4. Efficiency comparisons of the first proposed estimator, that has more efficiency compared to the other one, with the estimator of Droitcour et al.'s (1991) ICT and the estimator of Hussain et al.'s (2012) ICT are performed in section 5. The optimal sample size n, in the case of minimizing the variance of the estimator, assuming that the cost of conducting the survey is fixed, is determined in Section 6. Section 7 is devoted for conclusions and discussions.

2. The Usual ICT and Its Modification

This section presents the usual ICT which was proposed by Droitcour et al. (1991) and its modification was proposed by Hussain et al. (2012).

2.1. The Usual ICT

The usual ICT was introduced by Droitcour et al. (1991). It consists of selecting two independent samples of sizes n_1 and n_2. The i th respondent in the first sample is given a list of g innocuous items and asked to report the total number, say x_i, of items that are applicable to her/him. Similarly, the j th respondent in the second sample is provided another list of $(g+1)$ items including the sensitive item and asked to report a total number, say y_j , of the items that are applicable to her/him. The g innocuous items may or may not be the same in both the samples. Unbiased estimator of proportion of the sensitive item in the population is given by

$$\hat{\pi}_D = \bar{y} - \bar{x} \tag{2.1}$$

and its variance is given by

$$Var(\hat{\pi}_D) = \frac{\pi(1-\pi)}{n_2} + \frac{\sum_{j=1}^{g}\theta_j(1-\sum_{j=1}^{g}\theta_j)}{n_2} + \frac{\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right)+\sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k}{n_1} \tag{2.2}$$

2.2. A Modification of ICT

This modification was suggested by Hussain et al. (2012). Each respondent in a sample of size n is provided a questionnaire (list of questions) consisting of $g \geq 2$ questions. The j th question consists of queries about an unrelated item (F_j), and a sensitive characteristic S. The respondent is requested to count 1 if she/he possesses at least one of the characteristics F_j or S, otherwise, count zero, as a response to the j th question, and to report the total count based on entire questionnaire. And they assumed that z_i denote the total count of the i th respondent, and then mathematically they wrote it as

$$z_i = \sum_{j=1}^{g} \alpha_j \tag{2.3}$$

where α_j takes the values 1 and zero with probabilities $(\pi + \theta_j - \pi\theta_j)$ and $(1 - \pi - \theta_j + \pi\theta_j)$ respectively.

$$\therefore E(z_i) = \pi\left(g - \sum_{j=1}^{g} \theta_j\right) + \sum_{j=1}^{g} \theta_j \tag{2.4}$$

This suggests an unbiased estimator of π as

$$\hat{\pi}_p = \frac{\overline{z} - \sum_{j=1}^{g} \theta_j}{g - \sum_{j=1}^{g} \theta_j} \tag{2.5}$$

with variance

$$Var(\hat{\pi}_p) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)}{n\left(g - \sum_{j=1}^{g} \theta_j\right)^2}\left[\sum_{j=1}^{g} \theta_j \left(1 - \sum_{j=1}^{g} \theta_j\right) + \sum_{\substack{j,k=1 \\ j \neq k}}^{g} \theta_j \theta_k\right] \tag{2.6}$$

3. The Proposed Alternative Modification of the Usual ICT

Let a random sample of size n be selected using simple random sampling with replacement (SRSWR). The i th respondent is provided a list consists of $(g+1)$ items that including g innocuous items and one sensitive item and asked to:

- Firstly; count a number, say y_i, of the items that are applicable to her/him based on the entire list.
- Secondly; report how far away the produced number y_i is from $(g+1)$ if she/he has the sensitive item, or report the produced number y_i if she/he doesn't have it. It is to be mentioned that this idea is due to Christofides (2003).

The survey procedures are performed under the assumptions that the sensitive and innocuous items are unrelated and independent. The $(g+1)$ items are arranged randomly in the list. This technique improves the privacy protection of the respondents.

3.1. The First Proposed Estimator

Let

π be the true proportion of the population with the sensitive item.

$1 - \pi$ be the true proportion of the population without the sensitive item.

Since the j th item may be innocuous or sensitive item, and y_i be the total number produced by the i th respondent using the ICT.

$\therefore$ y_i can be written as follows

$$y_i = \sum_{j=1}^{g+1} \alpha_j \tag{3.1}$$

where

$$\alpha_j = \begin{cases} 1 & if\ the\ j\ th\ item\ is\ applicable\ to\ the\ i\ th\ respondent\ with\ pobability\ \ \pi + \theta_j \\ 0 & if\ the\ j\ th\ item\ is\ not\ applicable\ to\ the\ i\ th\ respondent\ with\ pobability\ \ 1 - \pi - \theta_j \end{cases}$$

Then, the expected value of y_i can be written as

$$E(y_i) = \sum_{j=1}^{g+1} E(\alpha_j) = \sum_{j=1}^{g+1} (\pi + \theta_j) = (g+1)\pi + \sum_{j=1}^{g} \theta_j \tag{3.2}$$

Then, from equation (3 - 2) we find

$$\hat{\pi}_1 = \frac{E(y_i) - \sum_{j=1}^{g} \theta_j}{(g+1)} \tag{3.3}$$

The expected value of $\hat{\pi}_1$ is

$$E(\hat{\pi}_1) = E\left[\frac{E(y_i) - \sum_{j=1}^{g} \theta_j}{(g+1)}\right] = \frac{E(y_i) - \sum_{j=1}^{g} \theta_j}{(g+1)} = \frac{(g+1)\pi + \sum_{j=1}^{g} \theta_j - \sum_{j=1}^{g} \theta_j}{(g+1)} = \pi$$

$\therefore$ $\hat{\pi}_1$ is an unbiased estimator of π. Also,

$$Var(y_i) = E(y_i^2) - [E(y_i)]^2$$

$$E(y_i^2) = \sum_{j=1}^{g+1} E(\alpha_j^2) + \sum_{\substack{j,k=1 \\ j\neq k}}^{g+1} E(\alpha_j \alpha_k)$$

$$= \sum_{j=1}^{g+1} (\pi + \theta_j) + \sum_{\substack{j,k=1 \\ j\neq k}}^{g+1} (\pi + \pi\theta_j + \pi\theta_k + \theta_j\theta_k)$$

$$= (g+1)^2\pi + \sum_{j=1}^{g} \theta_j + 2(g+1)\pi \sum_{j=1}^{g} \theta_j + \sum_{\substack{j,k=1 \\ j\neq k}}^{g} \theta_j\theta_k$$

$$\therefore Var(y_i) = (g+1)^2\pi + \sum_{j=1}^{g} \theta_j + 2(g+1)\pi \sum_{j=1}^{g} \theta_j + \sum_{\substack{j,k=1 \\ j\neq k}}^{g} \theta_j\theta_k - (g+1)^2\pi^2$$

$$-2(g+1)\pi \sum_{j=1}^{g} \theta_j - \left[\sum_{j=1}^{g} \theta_j\right]^2$$

$$\therefore Var(y_i) = (g+1)^2\pi(1-\pi) + \textstyle\sum_{j=1}^{g} \theta_j \left[1 - \sum_{j=1}^{g} \theta_j\right] + \sum_{\substack{j,k=1 \\ j\neq k}}^{g} \theta_j\theta_k \quad (3.4)$$

$$\therefore Var(\hat{\pi}_1) = \frac{\pi(1-\pi)}{n} + \frac{1}{n(g+1)^2}\left[\sum_{j=1}^{g} \theta_j \left(1 - \sum_{j=1}^{g} \theta_j\right) + \sum_{\substack{j,k=1 \\ j\neq k}}^{g} \theta_j\theta_k\right] \quad (3.5)$$

3.2. The Second Proposed Estimator

Let

d_i be the total number reported by the i th respondent in the sample

and

$$u_i = \begin{cases} g+1 & if\ the\ sensitive\ item\ is\ applicable\ to\ the\ i\ th\ respondent \\ 0 & if\ the\ sensitive\ item\ is\ not\ applicable\ to\ the\ i\ th\ respondent \end{cases}$$

$$\therefore Pr[u_i = g+1] = \pi \ \ and \ \ Pr[u_i = 0] = 1 - \pi$$

and

$$Pr[d_i = m] = (1-\pi)p_m + \pi p_{g+1-m} = \lambda_m \qquad m = 0, 1, 2, \dots, g \quad (3.6)$$

where

p_m is the probability that the produced total number by the i th respondent is m,

p_{g+1-m} is the probability that the produced total number by the i th respondent is $g+1-m$,

λ_m is the proportion of the respondents that report m.

$\therefore$ For the i th respondent

$d_i = |u_i - y_i|$ and similarly $y_i = |u_i - d_i|$

$$\therefore E(y_i) = \pi\, E[(g+1) - d_i] + (1-\pi)\, E[d_i]$$

$$= \pi\left[(g+1) - \bar{d}\right] + (1-\pi)\left[\bar{d}\right]$$

$$= \pi\left[(g+1) - 2\bar{d}\right] + \bar{d} \quad (3.7)$$

$$\therefore \hat{\pi}_2 = \frac{E(y_i) - \bar{d}}{[(g+1) - 2\bar{d}]} \quad (3.8)$$

The expected value of $\hat{\pi}_2$ is

$$E(\hat{\pi}_2) = E\left[\frac{E(y_i) - \bar{d}}{\left[(g+1) - 2\bar{d}\right]}\right] = \frac{E(y_i) - \bar{d}}{\left[(g+1) - 2\bar{d}\right]} = \frac{\pi\left[(g+1) - 2\bar{d}\right] + \bar{d} - \bar{d}}{\left[(g+1) - 2\bar{d}\right]} = \pi$$

$\therefore$ $\hat{\pi}_2$ is an unbiased estimator of π. Also

$$Var(y_i) = E(y_i^2) - [E(y_i)]^2$$
$$E(y_i^2) = \pi(g+1)^2 - 2\pi(g+1)\bar{d} + E[d_i^2]$$
$$[E(y_i)]^2 = \pi^2\left[(g+1) - 2\bar{d}\right]^2 + \left[\bar{d}\right]^2 + 2\pi(g+1)\bar{d} - 4\pi\left[\bar{d}\right]^2$$
$$\therefore Var(y_i) = \pi(1-\pi)\left[(g+1) - 2\bar{d}\right]^2 + Var(d_i) \quad (3.9)$$
$$\therefore Var(\hat{\pi}_2) = \frac{\pi(1-\pi)}{n} + \frac{Var(d_i)}{n[(g+1)-2\bar{d}]^2} \quad (3.10)$$

where

$$\bar{d} = \sum_{m=0}^{g} m\,\hat{\lambda}_m \quad (3.11)$$

and

$$Var(d_i) = \sum_{m=0}^{g} m^2\,\hat{\lambda}_m - \left[\sum_{m=0}^{g} m\,\hat{\lambda}_m\right]^2 \quad (3.12)$$

where $\hat{\lambda}_m$ is the proportion of the respondents that report $m\,; m = 0, 1, 2, \ldots, g$, in the sample.

From equations, of $E(y_i)$, (3 - 2) and (3 - 7) we find

$$E(y_i) = \frac{(g+1)\bar{d} + [2\bar{d} - (g+1)]\left(\sum_{j=1}^{g}\theta_j\right)}{[2\bar{d}]} \quad (3.13)$$

4. The Relative Efficiency of the Proposed Estimators of π

In this section we present efficiency comparison of the proposed estimator $\hat{\pi}_1$ with the proposed estimator $\hat{\pi}_2$. From equations (3 - 5) and (3 - 10), the relative efficiency of the two unbiased estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ is

$$REFF(\hat{\pi}_1, \hat{\pi}_2) = \frac{Var(\hat{\pi}_2)}{Var(\hat{\pi}_1)}$$
$$= \frac{\dfrac{\pi(1-\pi)\left[(g+1) - 2\bar{d}\right]^2 + Var(d_i)}{n\left[(g+1) - 2\bar{d}\right]^2}}{\dfrac{(g+1)^2\pi(1-\pi) + \left[\sum_{j=1}^{g}\theta_j\left(1 - \sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1 \\ j\neq k}}^{g}\theta_j\theta_k\right]}{n(g+1)^2}}$$
$$= \frac{(g+1)^2\pi(1-\pi)[(g+1)-2\bar{d}]^2 + (g+1)^2 Var(d_i)}{(g+1)^2\pi(1-\pi)[(g+1)-2\bar{d}]^2 + [(g+1)-2\bar{d}]^2\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1 \\ j\neq k}}^{g}\theta_j\theta_k\right]} \quad (4.1)$$

From the properties of the weighted mean the following property.

Property: Consider a set of real positive numbers $x_1, x_2, \ldots, x_n$ all are less than a positive real number *s*, and there are n positive weights w_i such that $\sum_{i=1}^{n} w_i = 1$ then

$$\left[\sum_{i=1}^{n} w_i x_i = \bar{x}_w\right] < s$$

The Prove:

$$\sum_{i=1}^{n} w_i\, x_i - s = \sum_{i=1}^{n} w_i\, x_i - s\sum_{i=1}^{n} w_i = \sum_{i=1}^{n} w_i\,[x_i - s] < 0.$$

$\therefore$ According to this property

$$(g+1) > \bar{d}$$

Which is always true for all values of g and $\bar{d}$, and

$$\left[(g+1) - 2\bar{d}\right]^2 = (g+1)^2 - 4(g+1)\bar{d} + 4\left[\bar{d}\right]^2$$
$$\therefore (g+1)\bar{d} > \left[\bar{d}\right]^2$$
$$\therefore (g+1)^2 > \left[(g+1) - 2\bar{d}\right]^2 \quad (4.2)$$

which also, is always true for all values of g and $\bar{d}$. Also, from equation (3 - 9) we find

$$Var(d_i) = Var(y_i) - \pi(1-\pi)\left[(g+1) - 2\bar{d}\right]^2 \tag{4.3}$$

and from equation (3 - 4) we find

$$\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right] = Var(y_i) - (g+1)^2\pi(1-\pi) \tag{4.4}$$

By subtracting equation (4 - 4) from equation (4 - 3)

$$Var(d_i) - \left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right] = \pi(1-\pi)\left[(g+1)^2 - \left[(g+1) - 2\bar{d}\right]^2\right]$$

According to equation (4 - 2)

$$\therefore \pi(1-\pi)\left[(g+1)^2 - \left[(g+1) - 2\bar{d}\right]^2\right] > 0$$

$$\therefore Var(d_i) > \left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right]$$

$$\therefore (g+1)^2 Var(d_i) > \left[(g+1) - 2\bar{d}\right]^2\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right]$$

$$\therefore \frac{Var(\hat{\pi}_2)}{Var(\hat{\pi}_1)} > 1$$

$\therefore \hat{\pi}_1$ has smaller variance than $\hat{\pi}_2$, i.e., $\hat{\pi}_2$ is less efficient than $\hat{\pi}_1$. We have calculated $\bar{d}$, $Var(d_i)$ and the relative efficiency of the proposed estimator $\hat{\pi}_1$ relative to the proposed estimator $\hat{\pi}_2$ for $g = 2, 3, \dots, 10$, $\pi = 0.1, 0.2, \dots, 0.5$, $n = 100$, and for $\theta_j < 0.5$ and $\theta_j \geq 0.5;\ \forall\, j$, at each value of g. The results are provided in Table (1) in Appendix. From table (1) it is indicated that;

- The proposed estimator $\hat{\pi}_1$ is more efficient than the proposed estimator $\hat{\pi}_2$.
- The sample size n does not have significant effect on the relative efficiency of the two proposed estimators.

5. Efficiency Comparisons

In this section we present efficiency comparisons of the estimator $\hat{\pi}_1$ of the proposed ICT with the estimator $\hat{\pi}_D$ of the usual ICT and with the estimator $\hat{\pi}_p$ of Hussain et al.'s ICT.

5.1. Proposed Estimator $\hat{\pi}_1$ Versus the Estimator $\hat{\pi}_D$

We compare the relative efficiency of the proposed estimator $\hat{\pi}_1$ with the estimator $\hat{\pi}_D$ in both cases of having and not having unequal $\theta_j \forall\, j$. In the case of having unequal $\theta_j \forall\, j$ the proposed estimator $\hat{\pi}_1$ would be more efficient than the estimator $\hat{\pi}_D$ if

$$Var(\hat{\pi}_D) - Var(\hat{\pi}_1) > 0$$

From equations (2 – 2) and (3 – 5), we have

$$Var(\hat{\pi}_D) - Var(\hat{\pi}_1) =$$

$$= \frac{n_1\pi(1-\pi) + n_1\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right)\right] + n_2\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right]}{n_1 n_2} -$$

$$\frac{(g+1)^2\pi(1-\pi) + \left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right]}{n(g+1)^2}$$

$$= \frac{n_1(g+1)^2\pi(1-\pi)[n-n_2] + n_1\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right)\right][n(g+1)^2-n_2] + n_2n(g+1)^2\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right)\right] + n_2\left[\sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right][n(g+1)^2-n_1]}{n_1n_2n(g+1)^2} \quad (5.1)$$

since

$n_1 \subset n$ and $n_2 \subset n$

$\therefore\ n-n_2 > 0$, $[n(g+1)^2-n_2] > 0$, and $[n(g+1)^2-n_1] > 0$ which are always true for all values of $n_1, n_2, n, and\ g$.

$\therefore The\ numenator\ of\ equation\ (5-1) > 0$

$\therefore\ Var(\hat{\pi}_D) - Var(\hat{\pi}_1) > 0$

$\therefore\ \hat{\pi}_1$ is more efficient than $\hat{\pi}_D$.

Moreover, in the case of having $\theta_j = {}^1/_g\ ; \forall j$, (which is difficult/impossible case), we find

$$\sum_{j=1}^{g}\theta_j = 1 \quad and \quad \sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k = \frac{1}{g}, \quad (5.2)$$

$$Var(\hat{\pi}_1) = \frac{\pi(1-\pi)}{n} + \frac{1}{ng(g+1)^2}, \quad (5.3)$$

and

$$Var(\hat{\pi}_D) = \frac{\pi(1-\pi)}{n_2} + \frac{1}{n_1g} \quad (5.4)$$

$$\therefore Var(\hat{\pi}_D) - Var(\hat{\pi}_1) = \frac{\pi(1-\pi)}{n_2} + \frac{1}{n_1g} - \frac{\pi(1-\pi)}{n} - \frac{1}{ng(g+1)^2}$$

$$\frac{n_1g(g+1)^2[n-n_2]\pi(1-\pi) + n_2[n(g+1)^2-n_1]}{n_1n_2ng(g+1)^2}$$

since

$n-n_2 > 0$, and $[n(g+1)^2-n_1] > 0$ which is always true for all values of $n_1, n_2, n, and\ g$. Then

$$n_1g(g+1)^2[n-n_2]\pi(1-\pi) + n_2[n(g+1)^2-n_1] > 0$$

$$\therefore Var(\hat{\pi}_D) - Var(\hat{\pi}_1) > 0$$

$\therefore\ \hat{\pi}_1$ is more efficient than $\hat{\pi}_D$.

5.2. Proposed Estimator $\hat{\pi}_1$ Versus the Estimator $\hat{\pi}_p$

We compare the relative efficiency of the proposed estimator $\hat{\pi}_1$ with the estimator $\hat{\pi}_p$ in both cases of having and not having unequal $\theta_j \forall\ j$. In the case of having unequal $\theta_j \forall\ j$ the proposed estimator $\hat{\pi}_1$ would be more efficient than the estimator $\hat{\pi}_p$ if

$$Var(\hat{\pi}_p) - Var(\hat{\pi}_1) > 0$$

From equations (2 - 6) and (3 – 5) we find,

$$Var(\hat{\pi}_p) - Var(\hat{\pi}_1) =$$

$$\frac{\left(g-\sum_{j=1}^{g}\theta_j\right)^2\pi(1-\pi) + (1-\pi)\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right]}{n\left(g-\sum_{j=1}^{g}\theta_j\right)^2} -$$

$$\frac{(g+1)^2\pi(1-\pi) + \left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right]}{n(g+1)^2}$$

$$= \frac{\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right]\left[(g+1)^2(1-\pi) - \left(g-\sum_{j=1}^{g}\theta_j\right)^2\right]}{n(g+1)^2\left(g-\sum_{j=1}^{g}\theta_j\right)^2} \quad (5.5)$$

We have calculated the relative efficiency of the proposed estimator $\hat{\pi}_1$ relative to the estimator $\hat{\pi}_p$ for $\pi = 0.1, 0.2, \dots, 0.8$, $g = 2, 3, \dots, 10$, $n = 100$, and for $\theta_j < 0.5;\ \forall j$ and $\theta_j \geq 0.5;\ \forall j$ at each value of g. The results are arranged in table (2).

Moreover, in the case of having $\theta_j = {}^1\!/_g\ ; \forall j$, (which is difficult/impossible case), we find,

$$Var(\hat{\pi}_p) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)}{ng\,(g-1)^2} \tag{5.6}$$

From equations (5 - 3) and (5 - 6) we find

$$Var(\hat{\pi}_p) - Var(\hat{\pi}_1) = \frac{(g+1)^2(1-\pi)-(g-1)^2}{ng\,(g-1)^2(g+1)^2} \tag{5.7}$$

Also, we have calculated the relative efficiency of the proposed estimator $\hat{\pi}_1$ relative to the estimator $\hat{\pi}_p$ in the case of having $\theta_j = {}^1\!/_g\ ; \forall j$ for $\pi = 0.1, 0.2, \dots, 0.5$, $g = 2, 3, \dots, 10$, $n = 100$, and for $\theta_j < 0.5;\ \forall j$ and $\theta_j \geq 0.5;\ \forall j$ at each value of g. The results are arranged in table (3).

From tables (2) and (3), it is indicated that;

- For $\pi \leq 0.5$, the proposed estimator $\hat{\pi}_1$ is more efficient than the estimator $\hat{\pi}_P$ for all values of g and $\theta_j \forall j$.
- For $\pi > 0.5$ and $\theta_j < 0.5\ \forall j$, the proposed estimator $\hat{\pi}_1$ is less efficient than the estimator $\hat{\pi}_P$ for all values of g.
- For $\pi > 0.5$ and $\theta_j \geq 0.5\ \forall j$, the proposed estimator $\hat{\pi}_1$ is more efficient than the estimator $\hat{\pi}_P$ for all values of g.
- The sample size n does not have significant effect on the relative efficiency of the proposed estimator $\hat{\pi}_1$ relative to $\hat{\pi}_p$.
- In the case of having $\theta_j = {}^1\!/_g\ ; \forall j \theta_j = {}^1\!/_g\ ; \forall j$, the relative efficiency of the two estimators $\hat{\pi}_1$ and $\hat{\pi}_p$ is approximately equivalent for all values of g and π.

6. The Optimal Sample Size n

In the sampling surveys it is necessarily to find the values of *n*, so that the variance of the estimator is minimum under fixed cost, or the cost is minimized under assumption that the variance of the estimator is predetermined (see Christofides (2005)). Consider the case of minimizing the variance of the estimator assuming that the cost of conducting the survey is fixed. Suppose that we have the following linear cost functions

$$C = c_0 + c_1 n \tag{6.1}$$

where

c_0 is the general (fixed) cost of the survey,
c_1 is the cost of interviewing an individual in the sample,
C is the total cost of the survey.

Then we want to find the optimal values of n which minimize the variance of $\hat{\pi}_1$ subject to a fixed cost. Thus, we have the following optimization problem

$$\underset{n}{Min}\left[\frac{\pi(1-\pi)}{n} + \frac{1}{n(g+1)^2}\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1 \\ j\neq k}}^{g}\theta_j\theta_k\right]\right]$$

subject to

$$C = c_0 + c_1 n \tag{6.2}$$

Using the method of Lagrange multipliers, we can equivalently have the unconstrained minimization problem

$$Min\ f(n, \beta)$$

where

$$f(n,\beta) = \frac{\pi(1-\pi)}{n} + \frac{1}{n(g+1)^2}\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1 \\ j\neq k}}^{g}\theta_j\theta_k\right] + \beta[c_0 + c_1 n - C] \quad (6-3)$$

Taking partial derivatives, we find

$$\frac{\partial f}{\partial n} = -\frac{\pi(1-\pi)}{n^2} - \frac{(g+1)^2\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right)+\sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right]}{n^2(g+1)^4} + \beta c_1 = 0 \tag{6.4}$$

$$\frac{\partial f}{\partial n} = c_0 + c_1 n - C = 0 \tag{6.5}$$

From equations (6 - 4) and (6 - 5), we find

$$\beta c_1 = \frac{(g+1)^2\pi(1-\pi)+\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right)+\sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right]}{n^2(g+1)^2} \tag{6.6}$$

$$c_1 n = C - c_0 \tag{6.7}$$

Solving the system of the equations (6 - 6) and (6 - 7), it follows that

$$n = \frac{(C-c_0)}{c_1} \tag{6.8}$$

If we substituting the value of n in equation (3 - 4) (Section 3), then we obtain the minimal value of variance $\hat{\pi}_1$ as follows;

$$\underset{m}{Var}(\hat{\pi}_1) = \frac{c_1\left[(g+1)^2\pi(1-\pi)+\left[\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right)+\sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right]\right]}{(g+1)^2(C-c_0)}$$

7. Conclusions and Discussions

The aim of this article was to provide an alternative modification of the usual ICT to estimate the proportion of a sensitive characteristic in some of the fields such as health care. The main features of this alternative modification are that, first of all it does not need to select two subsamples of sizes n_1 and n_2. Therefore, we do not require to worry about the optimal values n_1 and n_2 as is the usual ICT estimator $\hat{\pi}_D$. Secondly the first estimator of this alternative modification has been proven to be more efficient than the estimators of the other ICT techniques. Thirdly this technique provides full protection to the respondent's privacy since the reported number by the respondent may differ from the produced number and each reported number falls in the range $[0-g]$ and means that the respondent may have or may not have the sensitive item. This technique produced two estimators to estimate the proportion. We proved that the first proposed estimator to be more efficient than the second one. Also we proved that the first proposed estimator of the proposed item count technique is more efficient than the estimator of the usual ICT $\hat{\pi}_D$. Also, it has been observed that the first proposed estimator performs better than the estimator $\hat{\pi}_p$ of Hussain et al.'s (2012) ICT when $\pi \leq 0.5$ (which is the logical event). We determined the optimal sample size n, in the case of minimizing the variance of the estimator, assuming that the cost of conducting the survey is fixed. In summary, based on the findings of Sections 4 and 5, and the concluding discussion above we recommend the use of the proposed ICT with the first proposed estimator in surveys about sensitive items instead of the usual ICT and Hussain et al.'s item count technique.

ACKNOWLEDGMENTS

The author is especially appreciative for Dr. Tasos, C., Chistofides of Cyprus University, Nicosia, Cyprus. Dr. Zawar Hussain of Faculty of Sciences, King Abdul-Aziz University, Jeddah, Kingdom of Sudia Arabia. Dr. Ejaz Ali Shah and Dr. Javid Shabbir of Quaid-I-Azam University, Islamabad, Pakistan.

Also, the author is grateful to the Editor-in-Chief and to the learned referee for their valuable suggestions regarding improvement of the paper.

Appendix

Table (1). The percent relative efficiency of the two proposed estimators

g	$\sum_{j=1}^{g}\theta_j$	$\sum_{\substack{j,k=1 \\ j\neq k}}^{g}\theta_j\theta_k$	$\bar{d}$	$Var(d_i)$	π				
					0.1	0.2	0.3	0.4	0.5
2	0.5	0.12	0.9	0.69	434	318	275	256	251
2	1.5	1.12	1.2	0.56	1255	853	703	639	620
3	0.9	0.52	1.1	0.89	285	219	195	185	182
3	1.8	2.14	1.7	0.81	1750	1183	970	878	851
4	1.0	0.7	1.6	1.44	453	322	275	255	250
4	2.3	3.94	2.1	1.29	1645	1099	897	811	787
5	1.1	0.9	2.4	2.04	1346	867	701	633	613
5	3.0	7.16	2.7	2.01	4642	2988	2392	2139	2067
6	1.5	1.82	3.2	2.96	7433	4610	3637	3232	3117
6	4.2	14.6	3.4	2.34	51542	31937	25125	22278	21467
7	1.6	2.12	3.25	3.187	1393	885	713	642	622
7	5.0	21.32	3.85	3.3275	33503	20558	16122	14278	13754
8	2.1	3.4	4.0	3.7	3663	2225	1750	1555	1499
8	5.6	27.32	4.3	3.91	22449	13722	10751	9519	9169
9	2.3	4.6	4.5	5.05	4845	2959	2326	2066	1992
9	6.5	37.4	4.7	5.21	13674	8290	6482	5736	5524
10	2.7	6.44	5.1	6.09	9123	5520	4317	3821	3681
10	7.0	43.9	5.3	6.41	37986	22892	17843	15762	15172

Table (2). The percent relative efficiency of $\left[Var(\hat{\pi}_p)/Var(\hat{\pi}_1)\right]$ in the case of not having $\theta_j = 1/g\ ;\ \forall j$

g	$\sum_{j=1}^{g}\theta_j$	$\sum_{\substack{j,k=1 \\ j\neq k}}^{g}\theta_j\theta_k$	π							
			0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
2	0.5	0.12	182	145	130	121	114	109	103	96
2	1.5	1.12	1085	668	496	401	340	296	260	227
3	0.9	0.52	167	137	124	116	111	106	101	95
3	1.8	2.14	394	269	217	187	168	153	140	126
4	1.0	0.7	136	118	111	107	104	101	99	93
4	2.3	3.94	301	214	178	157	144	134	124	114
5	1.1	0.9	122	111	106	104	102	100	97	94
5	3.0	7.16	287	204	171	152	140	131	123	113
6	1.5	1.82	123	111	107	104	102	100	97	94
6	4.2	14.6	363	243	197	173	157	145	136	126
7	1.6	2.12	116	108	104	102	101	100	97	94
7	5.0	21.32	350	235	191	168	153	143	134	125
8	2.1	3.4	114	107	104	102	101	100	98	96
8	5.6	27.32	306	210	174	155	143	134	127	120
9	2.3	4.6	115	107	104	102	101	100	98	95
9	6.5	37.4	308	210	174	155	143	138	128	121
10	2.7	6.44	115	107	104	102	101	100	98	95
10	7.0	43.9	265	187	159	143	134	127	121	115

Table (3). The percent relative efficiency of $[Var(\hat{\pi}_p)/Var(\hat{\pi}_1)]$ in the case of having $\theta_j = 1/g\ ;\ \forall j$

g	π				
	0.1	0.2	0.3	0.4	0.5
2	371	260	211	183	164
3	149	125	116	111	108
4	115	107	104	103	102
5	106	103	102	101	101
6	103	101	101	100	100
7	102	101	100	100	100
8	101	100	100	100	100
9	101	100	100	100	100
10	101	100	100	100	100

REFERENCES

[1] Adebola, F., and Johnson, O., (2015): An Improved Warner's Randomized Response Model. International Journal of Statistics and Applications, Vol. 5, pp. 263-267.

[2] Barabesi, L., Dianna, G., and Perri, P., (2014): Horvitz-Thompson Estimation with Randomized Response and Non-Response Model. Assist. Statist. Appl., Vol. 9, pp. 3-10.

[3] Blair, G., Imai, K., and Zhou, Y.-Y., (2015): Design and Analysis of the Randomized Response Technique. JASA, Vol. 110, pp. 1304-1319.

[4] Chaudhuri, A., and Christofides, T., (2007): Item Count Technique in Estimating the Proportion of People with a Sensitive Feature. Journal of Statistical Planning and Inference, Vol. 137, pp. 589–593.

[5] Christofides, T., (2003). A Generalized Randomized Response Technique. *Metrika, Vol*. 57, *pp*. 195-200.

[6] Christofides, T., (2005): Randomized Response in Stratified Sampling. *Journal of Statistical Planning and Inference, Vol.* 128*, pp*. 303-310.

[7] Dalton,D., Wimbush, J., and M. Daily, C., (1994): Using the Unmatched Count Technique (UCT) to Estimate the Base Rates for Sensitive Behavior. Personnel Psychology, Vol. 47, pp. 817–828.

[8] Droitcour, J., Caspar, R., Hubbard, M., Parsley, T., Visscher, W., and Ezzati, T., (1991): The Item Count Technique as a Method of Indirect Questioning: a Review of its Development and a Case Study Application, in Measurement Errors in Surveys, Biemer, P., Groves, R., Lyberg, L., Mathiowetz, N., and Sudman, S., (editors). Wiley, New York.

[9] Droitcour, J., and Larson, E., (2002): An Innovative Technique for Asking Sensitive Questions: The Three Card Method. Sociological Methodological Bulletin, Vol. 75, pp. 5–23.

[10] Greenberg, B., Abul-Ela, A., Simmons, W., and Horvitz, D., (1969): The Unrelated Question Randomized Response Model: Theoretical Framework. JASA, Vol. 64, pp. 520-539.

[11] Hussain, Z., Ali Shah, E., and Shabir, J., (2012): An Alternative Item Count Technique in Sensitive Surveys. Revista Colombiana de Estadística, Vol. 35, pp. 39- 54.

[12] Hussain, Z., and Shabbir, J., (2010): On Item Count Technique in Survey Sampling, Journal of Informatics and Mathematical Sciences. Vol. 2, pp. 161-169.

[13] Imai, K., (2011): Multivariate Regression Analysis for the Item Count Technique. JASA, Vol. 106, pp. 407-416.

[14] Kuha, J., and Jackson, J., (2014): The Item Count Method for Sensitive Survey Questions: Modeling Criminal Behavior. Journal of the Royal Statistical Society: Series C (Applied Statistics), Vol. 63, pp. 321-341.

[15] Mangat, N., (1994): An Improved Randomized Response Strategy. Jour. Royal Statist. Soc. Ser. B, Vol. 56, pp. 93-95.

[16] Mangat, N., Singh, R., and Singh, S., (1997): Violation of Respondent's Privacy in Moors' Model- its Rectification Through a Random Group Strategy Response Model. Commun. Statist. Theory Methods, Vol. 26, pp. 243-255.

[17] Miller, J., (1985): The Nominative Technique: A New Method of Estimating Heroin Prevalence. NIDA Research Monograph, Vol. 57, pp. 104–124.

[18] Moors, J., (1971): Optimization of the Unrelated Question Randomized Response Model. JASA, Vol. 66, pp. 627-629.

[19] Tsuchiya, T., (2005): Domain Estimators for the Item Count Technique. Statistics Canada, Vol. 31, pp. 41-51.

[20] Tsuchiya, T., Hirai, Y., and Ono, S., (2007): A Study of the Properties of the Item Count Technique. Public Opinion Quarterly, Vol. 71, pp. 253–272.

[21] Tourangeau, R., and Yan, T., (2007): Sensitive Questions in Surveys. Psychological Bulletin, Vol. 133, pp. 859–883.

[22] Saha, A., (2007): A Simple Randomized Response Technique in Complex Surveys. METRON-International Journal of Statistics, Vol. LXV, pp. 55-66.

[23] Singh, H., and Tarray, T., (2013): An Alternative to Kim and Warde's Mixed Randomized Response Technique. Statistica, Vol. LXXIII, pp. 379-402.

[24] Warner, S., (1965): Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. JASA, Vol. 60, pp. 63–69.

[25] Walter, F., and Laier, B., (2014): The Effectiveness of the Item Count Technique in Eliciting Valid Answers to Sensitive Questions. An Evaluation in the Context of Self-Reported Delinquency. Survey Research Methods, Vol. 8, pp. 153-168.

[26] Wimbush, J., and Dalton, D., (1997): Base Rate for Employee Theft: Convergence of Multiple Methods. Journal of Applied Psychology, Vol. 82, pp. 756–763.