A Modified Stratified Randomized Response Techniques

F.B. Adebola , O.O. Johnson , N.A. Adegoke *

Department of Statistics, Federal University of Technology Akure, P.M.B. 704, Nigeria.

* E-mail of the corresponding author: nurudeen.adegoke@yahoo.com

Abstract
This paper presents a modification of Kim and Warde’s (2004) Stratified Warner’s Randomized Response Technique (SWRRT) and Kim and Elam’s (2007) Stratified Unrelated Question Randomized Response Technique with known distribution (SURRT) using optimal allocation. Our models further reduce the non-response bias by introducing the concept of sub-samples of non-respondent developed by Hansen and Hurwitz (1946) to the above mentioned models. In this paper we perform an empirical practice of our model and we also perform the empirical comparison of our models with both Kim and Warde (2004) and Kim and Elam (2007) models. We discovered a new note in the empirical comparative study of our models.

Keywords: Close supervision, Sensitive behavior, Non-respondent, Sub-sample, Randomized response techniques.

1. Introduction
In other to reduce biasness of respondent in surveys, most especially in surveys that the questions are so sensitive that might cause the respondent not to even answer the questions or falsely answer the questions; in this regard we appeal to Randomized Response Techniques. Randomized Response Techniques is the method of protecting the respondent’s identity so as to collect reliable information on sensitive characteristics. Sensitive characteristics involve; rape, abortion, Sexually Transmitted Infection/HIV etc. Randomized Response Techniques help to reduce non-sampling error; non-sampling error involves response bias and non-response bias Warner (1965). Response bias is when the respondent gives a false answer and non-response bias is when the respondent refuses to answer the question.

Randomized Response Technique was introduced by Warner (1965) to estimate the proportion of the people in a given population who is involve in a particular social or legal stigmatized behavior. Warner uses a randomized device with known probability to encourage the respondent for truthful answer such that the researcher does not know which of the two questions he or she answers. Greenberg et al. (1969) proposed the unrelated question RR model that is a variation of Warner’s (1965) RR model in which the respondent is presented with one sensitive question and one unrelated question (which is not related to the stigmatized behavior). Hong et al. (1994) presented a stratified randomized response technique using proportional allocation. Kim and Warde (2004) presented Stratified Warner’s Randomized Response Model (SWRRT) using an optimal allocation. Kim and Elam (2007) presented the Stratified Unrelated Question Randomized Response Model (SURRT) using an optimal allocation.

This paper develops a Modified Stratified Randomized Response Model using optimal allocation. We developed for Warner (1965) and for Greenberg et al. (1969). As stated earlier that Randomized Response Techniques helps to reduce response and non-response bias. Our model further reduces the non-response bias by introducing the concept of sub-samples of non-respondent developed by Hansen and Hurwitz (1946). A comparative study of our two models was done as well so as to determine the most efficient.

2 Randomized Response Techniques

Warner (1965) gave a genius idea by using randomized device to encourage truthful answer from the respondent with respect to a sensitive behavior. The randomizing device, such as a spinning arrow, dice or coins is used to select one of the two questions; such as,

“I have raped before” (Presented with probability P)

“I have not raped before” (Presented with probability 1-P)

The respondents have the options “Yes” or “No” presented to him or her. The interviewer does not know which question any respondent has answered but knows the probability P and 1-P with which the two statements are presented. Here, with a random sample of n respondents, the interviewer records a binomial estimate
\[ \hat{\theta} = \frac{y}{n} \]

of the proportion \( \theta \) of “Yes” answers, where \( y \) is the number of yes answers.

If the questions are answered truthfully, the relation between \( \theta_h \) and \( \pi_{Wh} \) in the population is given as:

\[ \theta_h = \pi_{Wh} p_h + (1 - p_h)(1 - \pi_{Wh}) \]  \hspace{1cm} (2.1)

Where \( \pi_{Wh} \) is the proportion of people with the stigmatized or sensitive behavior in stratum \( h \) using Warner’s techniques and \( p_h \) is the probability of selecting the sensitive question in stratum \( h \).

\[ \hat{\pi}_{Wh} = \frac{\theta_h - (1 - p_h)}{(2p_h - 1)} \]  \hspace{1cm} (2.2)

The Variance is given by

\[ V(\hat{\pi}_{Wh}) = \frac{\pi_{Wh}(1 - \pi_{Wh})}{n_h} + \frac{p_h(1 - p_h)}{n_h(2p_h - 1)^2} \]  \hspace{1cm} (2.3)

Horvitz et al. (1967) proposed the unrelated question design which is an alternative to the Warner’s method. They discussed that combining one sensitive question and one unrelated question would enhance more cooperation of the respondents.

The unrelated question design was further improved by Greenberg et al. (1969). They examined the same design where the distribution of the non-sensitive question is known in advance. The unrelated question design with a known distribution is a simplified version of the original unrelated question design. To illustrate the pattern of the design;

Question 1: Do you have more than one girlfriend?

Question 2: Did you watch the 9.00pm cartoon yesterday?

By presenting the second question in this form gives the distribution of the unrelated question.

If all respond truthfully, the population proportion of “yes” answers is given by

\[ \theta_h = p_h \pi_{Uh} + (1 - p_h) \pi_{Kh} \]  \hspace{1cm} (2.4)

Where \( \pi_{Kh} \) is the proportion in the sampled population who watch the 9:00pm cartoon yesterday in stratum \( h \) which is always assume to be known.

If \( \pi_{Kh} \) is known, the obvious (and maximum likelihood) estimate of \( \pi_{Uh} \) is

\[ \hat{\pi}_{Uh} = \frac{\hat{\theta}_h - (1 - p_h) \hat{\pi}_{Kh}}{p_h} \]  \hspace{1cm} (2.5)

Where \( \pi_{Uh} \) is the proportion of people with the stigmatized or sensitive behavior in stratum \( h \) using Unrelated design techniques and \( p_h \) is the probability of selecting the sensitive question in stratum \( h \).

The Variance is given by

\[ V(\hat{\pi}_{Uh}) = \frac{\theta_h (1 - \theta_h)}{n_h p_h^2} \]  \hspace{1cm} (2.6)

For a review, Hong et al. (1994) suggested a Stratified Randomized Response Techniques (RRT) under the proportional sampling assumption. Under Hong et al.’s (1994) proportional sampling assumption, it may be easy to derive the variance of the proposed estimator. However, it may come at a high cost in terms of time, effort, and money.
To proffer solution to the deficiency of Hong et al. (1994), Kim and Warde (2004) and Kim and Elam (2005) both presented Stratified Randomized Response Techniques (SRRT) using an optimal allocation which are more efficient than the Hong et al. (1994) Stratified Randomized Response Techniques (SRRT) using a proportional allocation. (See Kim and Warde (2004))

They derived the probability \( \theta_h \) of a "Yes" answer in stratum \( h \) for this procedure as: If all respond truthfully,

\[
\theta_h = \pi_{Wh} p_h + (1 - p_h)(1 - \pi_{Wh})
\]

for \( h = 1, 2, \ldots, L \)

Where \( \pi_{Wh} \) is the proportion of people with the sensitive trait in stratum \( h \). The maximum likelihood estimate \( \hat{\pi}_W \) of a sensitive proportion \( \pi_W \) is given by:

\[
\hat{\pi}_W = \frac{\sum_{h=1}^{L} W_h \hat{\pi}_Wh}{\sum_{h=1}^{L} W_h}
\]

(2.8)

Where \( W_h = \frac{N_h}{N} \) for \( h = 1, 2, \ldots, L \) so that \( W = \sum_{h=1}^{L} W_h = 1 \) ( \( N \) is the number of unit in the population and \( N_h \) is the total number of units in stratum \( h \)). The minimum variance of the estimator \( \hat{\pi}_W \) is given by (See Kim and Elam (2005)):

\[
V(\hat{\pi}_W) = \frac{1}{n} \left[ \frac{\sum_{h=1}^{L} W_h (\pi_{Wh} (1 - \pi_{Wh}) + p_h (1 - p_h) (2p_h - 1)^2)^2}{(\sum_{h=1}^{L} W_h) \sum_{h=1}^{L} W_h} \right]
\]

(2.9)


They derived the probability \( \theta_h \) of a "Yes" answer in stratum \( h \) for this procedure as: If all respond truthfully,

\[
\theta_h = p_h \pi_{Uh} + (1 - p_h) \pi_{Kh}
\]

(2.10)

For \( h = 1, 2, \ldots, L \)

Where \( \pi_{Kh} \) is the proportion in the sampled population who watch the 10:00pm news yesterday and it is known. Where \( \pi_{Uh} \) is the proportion of people with the sensitive trait in stratum \( h \). The maximum likelihood estimate \( \hat{\pi}_U \) of a sensitive proportion \( \pi_U \) is given by:

\[
\hat{\pi}_U = \sum_{h=1}^{L} W_h \hat{\pi}_{Uh}
\]

(2.11)

Where \( W_h = \frac{N_h}{N} \) for \( h = 1, 2, \ldots, L \) so that \( W = \sum_{h=1}^{L} W_h = 1 \) ( \( N \) is the number of unit in the population and \( N_h \) is the total number of units in stratum \( h \))

The minimum variance of the estimator \( \hat{\pi}_U \) given \( \pi_{Kh} \) is given by (See Kim and Elam (2007)):
Here, we develop a Modified Stratified Warner’s Randomized Response Technique (MSWRRT) and Modified Stratified Unrelated Question Randomized Response Technique with known distribution (MSURRT) using optimal allocation.

3. Our Model

Hansen and Hurwitz (1946) provided a technique for dealing with the problem of the unit non-response in mail surveys. The technique is as follows:

The population considered is denoted by $N$ and the population is stratified into $h$ ($h > 1$) strata. We have $N_h$ the number of units in the population in stratum $h$. Stratification is the method of using auxiliary information to increase the precision of the estimate of the population characteristics. (See Okafor 2002).

A sample of size $n$ was determined, the number of sample in each stratum was determined using the optimal allocation. Then we have the number of unit in each stratum as $n_h$ such that $\sum n_h = n$. By Simple Random Sampling without Replacement; a questionnaire is mailed to each stratum and $n'_h$ is the number of unit responded and not responded in stratum $h$ respectively. We have $n_h = n'_h + n''_h$.

3.1 For the Respondent:

The interviewer records a binomial estimate $\hat{\theta}_h = \frac{y}{n'_h}$ of the proportion $\theta$ of “Yes” answers, where $y$ is the number of yes answers.

For Warner (1965):

If the questions are answered truthfully, the relation between $\hat{\theta}_h$ and $\pi_{Wh}'$ in the population is given as:

$$\pi_{Wh}' = \pi_{Wh} p_h + (1-p_h)(1-\pi_{Wh}')$$

(3.1)

Where $\pi_{Wh}'$ is the proportion of people with the stigmatized or sensitive behavior in stratum $h$ using Warner’s techniques and $p_h$ is the probability of selecting the sensitive question in stratum $h$.

$$\hat{\pi}_{Wh}' = \frac{\hat{\theta}_h - (1-p_h)}{2p_h - 1}$$

(3.2)

The maximum likelihood estimate $\hat{\pi}_{W1}'$ of a sensitive proportion $\pi_{W1}'$ for the respondent is given by:

$$\hat{\pi}_{W1}' = \sum_{h=1}^{L} W_h \hat{\pi}_{Wh}'$$

(3.3)

The Variance for stratum $h$ is given by

$$V(\hat{\pi}_{Wh}') = \frac{\pi_{Wh}'(1-\pi_{Wh}')}{n'_h} + \frac{p_h(1-p_h)}{n'_h (2p_h - 1)^2}$$

(3.4)
Where \( W_h = \frac{N_h}{N} \) for \( h = 1, 2, \ldots, L \) so that \( W = \sum_{h=1}^{L} W_h = 1 \) (\( N \) is the number of unit in the population and \( N_h \) is the total number of units in stratum \( h \))

The minimum variance of the estimator \( \hat{\pi}'_{W1} \) for the respondent is given by:

\[
V(\hat{\pi}'_{W1}) = \frac{1}{n_1'} \left[ \sum_{h=1}^{L} W_h \left( \pi'_W (1-\pi'_W) + \frac{p_h (1-p_h)}{(2p_h-1)^2} \right) \right]^2
\]

(3.5)

Where \( n_1' = \sum_{h=1}^{L} n_{h1}' \)

For Greenberg et al. (1969):

If all respond truthfully, the population proportion of “yes” answers is given by

\[
\theta'_h = p_h \pi'_U h + (1-p_h) \pi'_K h
\]

(3.6)

Where \( \pi'_K h \) is the proportion in the sampled population who watch the 9:00pm cartoon yesterday in stratum \( h \) which is always assume to be known.

If \( \pi'_K h \) is known, the obvious (and maximum likelihood) estimate of \( \pi'_U h \) is

\[
\hat{\pi}'_{U h} = \frac{\hat{\theta}'_h (1-p_h) \hat{\pi}'_K h}{p_h}
\]

(3.7)

Where \( \pi'_U h \) is the proportion of people with the stigmatized or sensitive behavior in stratum \( h \) using Unrelated design techniques and \( p_h \) is the probability of selecting the sensitive question in stratum \( h \).

The maximum likelihood estimate \( \hat{\pi}'_{U1} \) of a sensitive proportion \( \pi'_U \) for the respondent is given by:

\[
\hat{\pi}'_{U1} = \sum_{h=1}^{L} W_h \hat{\pi}'_{U h}
\]

(3.8)

Where \( W_h = \frac{N_h}{N} \) for \( h = 1, 2, \ldots, L \) so that \( W = \sum_{h=1}^{L} W_h = 1 \) (\( N \) is the number of unit in the population and \( N_h \) is the total number of units in stratum \( h \)). The Variance for stratum \( h \) is given by

\[
V(\hat{\pi}'_{U h} \backslash \pi'_K h) = \frac{\theta'_h (1-\theta'_h)}{n_{h1}' p_h^2}
\]

(3.9)

The minimum variance of the estimator \( \hat{\pi}'_{U1} \) given \( \pi'_K \) is given by (See Kim and Elam (2007)):

\[
V(\hat{\pi}'_{U1} \backslash \pi'_K h) = \frac{1}{n_1'} \left[ \sum_{h=1}^{L} \frac{W_h}{p_h} \sqrt{\theta'_h (1-\theta'_h)} \right]^2
\]

(3.10)
Where \( n'_h = \sum_{k=1}^{L} n'_{hk1} \)

### 3.2 For the Non-Respondent:

A subsample of \( \left( m_h = \frac{n'_{h2}}{k}, k \geq 1 \right) \) units is selected from the \( n'_h \) non-respondents by Simple Random Sampling without Replacement in stratum \( h \). The desired information is now collected from these \( m_h \) units by close supervision. This time around it is assumed that all the \( m_h \) units respond.

The interviewer records a binomial estimate \( \hat{\theta}'_h = \frac{y}{m_h} \) of the proportion \( \theta \) of “Yes” answers, where \( y \) is the number of yes answers and \( m_h \) is the sub-sample of the non-respondent in stratum \( h \).

**For Warner (1965):**

If the questions are answered truthfully, the relation between \( \theta'_h \) and \( \pi'_w_h \) in the population is given as:

\[
\theta'_h = \pi'_w_h p_h + (1 - p_h)(1 - \pi'_w_h) \tag{3.11}
\]

Where \( \pi'_w_h \) is the proportion of people with the stigmatized or sensitive behavior in stratum \( h \) using Warner’s techniques and \( p_h \) is the probability of selecting the sensitive question in stratum \( h \).

\[
\hat{\pi}'_{wh} = \frac{\hat{\theta}'_h - (1 - p_h)}{(2p_h - 1)} \tag{3.12}
\]

The maximum likelihood estimate \( \hat{\pi}'_{w2} \) of a sensitive proportion \( \pi'_w2 \) for the respondent is given by:

\[
\hat{\pi}'_{w2} = \sum_{h=1}^{L} W_h \hat{\pi}'_{wh} \tag{3.13}
\]

The Variance for stratum \( h \) is given by

\[
V(\hat{\pi}'_{wh}) = \frac{\pi'_w_h(1 - \pi'_w_h)}{m_h} + \frac{p_h(1 - p_h)}{m_h(2p_h - 1)^2} \tag{3.14}
\]

Where \( W_h = \frac{N_h}{N} \) for \( h = 1,2,\ldots, L \) so that \( W = \sum_{h=1}^{L} W_h = 1 \) (\( N \) is the number of unit in the population and \( N_h \) is the total number of units in stratum \( h \)).
The minimum variance of the estimator $\hat{\pi}_{W1}'$ for the respondent is given by:

$$V(\hat{\pi}_{W1}') = \frac{1}{m} \left[ \sum_{h=1}^{L} W_h \left( \pi_{Wh}' (1 - \pi_{Wh}') + \frac{p_h (1 - p_h)}{(2p_h - 1)^2} \right) \right]$$

(3.15)

Where $m = \sum_{h=1}^{L} m_h$

For Greenberg et al. (1969):

If all respond truthfully, the population proportion of “yes” answers is given by

$$\theta'_h = p_h \pi_{Uh}' + (1 - p_h) \pi_{Kh}'$$

(3.16)

Where $\pi_{Kh}'$ is the proportion in the sampled population who watch the 9:00pm cartoon yesterday in stratum $h$ which is always assume to be known.

If $\pi_{Kh}'$ is known, the obvious (and maximum likelihood) estimate of $\pi_{Uh}'$ is

$$\hat{\pi}_{Uh}' = \frac{\theta'_h - (1 - p_h) \pi_{Kh}'}{p_h}$$

(3.17)

Where $\pi_{Uh}'$ is the proportion of people with the stigmatized or sensitive behavior in stratum $h$ using Unrelated design techniques and $p_h$ is the probability of selecting the sensitive question in stratum $h$.

The maximum likelihood estimate $\hat{\pi}_{U2}'$ of a sensitive proportion $\pi_{U2}'$ for the respondent is given by:

$$\hat{\pi}_{U2}' = \sum_{h=1}^{L} W_h \hat{\pi}_{Uh}'$$

(3.18)

Where $W_h = \frac{N_h}{N}$ for $h = 1, 2, ..., L$ so that $W = \sum_{h=1}^{L} W_h = 1$ ( $N$ is the number of unit in the population and $N_h$ is the total number of units in stratum $h$)

The Variance for stratum $h$ is given by
\[ V(\hat{\pi}_{U_2} \setminus \pi'_{Kh}) = \frac{\theta_h'(1-\theta_h')}{m_h p_h^2} \]  

(3.19)

The minimum variance of the estimator \( \hat{\pi}_{U_2} \) given \( \pi'_{Kh} \) is given by (See Kim and Elam (2007)):

\[ V(\hat{\pi}_{U_2} \setminus \pi'_{Kh}) = \frac{1}{m} \left[ \sum_{h=1}^{L_h} \frac{W_h}{p_h} \sqrt{\theta_h'(1-\theta_h')} \right]^2 \]  

(3.20)

Where \( m = \sum_{h=1}^{L_h} m_h \)

### 3.3 The True Proportion of the Sensitive Behavior

The two independent estimates of the proportion of the sensitive behavior obtained from the respondent and non-respondent strata are suitably weighted to form the estimate of the population proportion of the sensitive behavior under study (See Hansen and Hurwitz (1946)).

Now, extending Hansen and Hurwitz (1946) work to randomized response techniques,

Then we have,

The unbiased estimator for proportion of the sensitive behavior is given by:

\[ \hat{\pi}' = \frac{n_1' \pi_1' + n_2' \pi_2'}{n} = w_1' \pi_1' + w_2' \pi_2' \]  

(3.21)

\( \pi_1' \) is the estimate of the proportion of the sensitive behavior of the \( n_1' \) respondents at the first mailings.

\( \pi_2' \) is the estimate of the proportion of the sensitive behavior of the \( m \) sub-sampled non-respondents at the second call.

Where, \( w_1' = \frac{n_1'}{n} \) and \( w_2' = \frac{n_2'}{n} \)

The unbiased variance of the sensitive behavior given by:

\[ \hat{V}(\hat{\pi}') = \frac{N-n}{Nn(n-1)} \left\{ (n_1' - 1) \frac{n_1' \pi_1'(1-\pi_1')}{{n_1'}^2} + (n_2' - k) \frac{m \pi_2'(1-\pi_2')}{{m-1}} + n_1' \pi_1'^2 + n_2' \pi_2'^2 - n \pi'^2 \right\} \]

\[ + \frac{N-1}{N(n-1)} w_2'(k-1) \frac{m \pi_2'(1-\pi_2')}{m-1} \]  

(3.22)
The unbiased variance of $\hat{\pi}'$ of $\pi'$ is obtained upon replacing $s_1^2$ is replaced by $\frac{n'_1\pi'_i(1-\pi'_i)}{n'_1-1}$, $s_{2m}^2$ by

$$\frac{m\pi'_1(1-\pi'_2)}{m-1}; \quad \bar{y}_1 \text{ is replaced by } \pi'_1; \quad \bar{y}_{2m} \text{ by } \pi'_2 \text{ and } \bar{y}_r \text{ by } \pi' \text{ (See Hansen and Hurwitz (1946)) and Okafor (2002).}$$

Equation (3.21) and (3.22) can be used for both Warner (1965) and Greenberg et al. (1969).

4 Our Study

Kim and Warde (2004) and Kim and Elam (2007) have presented Stratified Randomized Response Techniques (SRRT) and has done the efficiency comparison (See Kim and Elam (2007)), but we still have to compare our model with this two models. Here, we perform an empirical practice of our model and we also perform the empirical comparison of our models with both Kim and Warde (2004) and Kim and Elam (2007) models. We apply our model to sexual behavior among students in which we administered five different sensitive questions. The questions involve:

1. Have you contracted Sexually transmitted disease before?
2. Have you aborted for a lady or abort a pregnancy before?
3. Have you raped or being raped before?

These questions were considered sensitive because they are social undesirable attributes such that people do not feel comfortable to disclose information about it. For Greenberg et al (1969), we paired these questions with an unrelated question with known distribution.

4.1 Study Design

The population considered in the survey was the undergraduate students of the Federal University of Technology, Akure, Nigeria. The population was stratified into two in which we have the Male and the Female (that is $h = 1$, 2). Stratification is the method of using auxiliary information to increase the precision of the estimate of the population characteristics. (See Okafor 2002).

Two questionnaires were design to evaluate the performance of our model; one using the Warner (1965) design and the other using the Greenberg et al (1969) design. We asked some questions before asking the sensitive questions.

The total population of students in the university ($N = 12051$, (8782 males and 3269 females)), we chose a sample size of $n=250$ for each design, optimal allocation was used to determine the number of people in each stratum, we have $n_1 = 181$ males and $n_2 = 69$ females. Out of the 250 questionnaire sent out for the Stratified Warner’s Randomized Response Techniques, only 183 (144 males and 39 females) was useable (respondent), 67 was recorded non-respondent and out of the 250 questionnaire sent out for Stratified Unrelated Question Randomized Response Techniques, only 200 (155 males and 45 females) was useable(respondent), 50 was recorded non-respondent.

We perform a sub-sample of non-respondent by choosing $k =2$, this choice of $k$ is to get a large sub-sample. As the procedure implies, the sub-sample was done with close supervision (this does not mean that we actually knew which question the respondent answered but we help to simplify the instruction so as to get more cooperation of the respondent). All the $k$ sub-sample responses were now useable.
4.2 Results

**Table 1. The Response and the Non-Response of the Survey**

<table>
<thead>
<tr>
<th></th>
<th>*SWRRT</th>
<th></th>
<th>**SURRT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MALE</td>
<td>FEMALE</td>
<td>MALE</td>
<td>FEMALE</td>
</tr>
<tr>
<td>Number originally sent</td>
<td>181</td>
<td>69</td>
<td>181</td>
<td>69</td>
</tr>
<tr>
<td>Number unreturned</td>
<td>15</td>
<td>10</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>Number returned</td>
<td>166</td>
<td>59</td>
<td>169</td>
<td>63</td>
</tr>
<tr>
<td>Invalid response</td>
<td>22</td>
<td>20</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>Useable response</td>
<td>144</td>
<td>39</td>
<td>155</td>
<td>45</td>
</tr>
<tr>
<td>Non-respondent</td>
<td>37</td>
<td>30</td>
<td>26</td>
<td>24</td>
</tr>
</tbody>
</table>

* Stratified Warner’s Randomized Response Techniques (Kim and Warde (2004))

** Stratified Unrelated Question Randomized Response Techniques (Kim and Elam (2007))

**Table 2. For Male**

<table>
<thead>
<tr>
<th>Type of Sexual behavior</th>
<th>SWRRT*</th>
<th>SURRT**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Useable response</td>
<td>Yes Response</td>
</tr>
<tr>
<td>Q1</td>
<td>144</td>
<td>45</td>
</tr>
<tr>
<td>Q2</td>
<td>144</td>
<td>56</td>
</tr>
<tr>
<td>Q3</td>
<td>144</td>
<td>44</td>
</tr>
</tbody>
</table>

**Table 3. For Female**

<table>
<thead>
<tr>
<th>Type of Sexual behavior</th>
<th>SWRRT*</th>
<th>SURRT**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Useable response</td>
<td>Yes Response</td>
</tr>
<tr>
<td>Q1</td>
<td>39</td>
<td>13</td>
</tr>
<tr>
<td>Q2</td>
<td>39</td>
<td>14</td>
</tr>
<tr>
<td>Q3</td>
<td>39</td>
<td>12</td>
</tr>
</tbody>
</table>
Table 4. Proportion of the Sensitive Behavior (Male and Female) For Respondents

<table>
<thead>
<tr>
<th>Type of Sexual behavior</th>
<th>SWRRT*</th>
<th>SURRT**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion with the sensitive behavior</td>
<td>Variance of the sensitive behavior</td>
<td>Proportion with the sensitive behavior</td>
</tr>
<tr>
<td>Q1</td>
<td>4.56%</td>
<td>0.0076</td>
</tr>
<tr>
<td>Q2</td>
<td>20.16%</td>
<td>0.0082</td>
</tr>
<tr>
<td>Q3</td>
<td>1.54%</td>
<td>0.0074</td>
</tr>
</tbody>
</table>

4.2.2 Sub-Sample of Non-Respondent

Table 5. The Table Below Shows the Result of the Survey:

<table>
<thead>
<tr>
<th></th>
<th>SWRRT*</th>
<th>SURRT**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MALE</td>
<td>%</td>
</tr>
<tr>
<td>Total non-response</td>
<td>37</td>
<td>30</td>
</tr>
<tr>
<td>Number sub-sampled</td>
<td>19</td>
<td>100%</td>
</tr>
<tr>
<td>Number returned</td>
<td>19</td>
<td>100%</td>
</tr>
<tr>
<td>Useable Response</td>
<td>19</td>
<td>100%</td>
</tr>
</tbody>
</table>

The table below presents the number of yes response and the proportion of student with the sensitive behavior alongside with the corresponding variances for subsample of non-response.

Table 6. For Male

<table>
<thead>
<tr>
<th></th>
<th>SWRRT*</th>
<th>SURRT**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Useable response</td>
<td>Yes Response</td>
</tr>
<tr>
<td>Q1</td>
<td>19</td>
<td>9</td>
</tr>
<tr>
<td>Q2</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>Q3</td>
<td>19</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 7. For Female

<table>
<thead>
<tr>
<th></th>
<th>SWRRT</th>
<th>SURRT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Useable response</td>
<td>Yes Response</td>
</tr>
<tr>
<td>Q1</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Q2</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>Q3</td>
<td>15</td>
<td>5</td>
</tr>
</tbody>
</table>
Table 8. The Total Proportion of the Sensitive Behavior for the Non-Respondents

<table>
<thead>
<tr>
<th>Type of Sexual behavior</th>
<th>Proportion with the sensitive behavior</th>
<th>Variance of the sensitive behavior</th>
<th>Proportion with the sensitive behavior</th>
<th>Variance of the sensitive behavior</th>
<th>Relative Efficiency: R.E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>13.18%</td>
<td>0.0462</td>
<td>38.57%</td>
<td>0.0188</td>
<td>246%</td>
</tr>
<tr>
<td>Q2</td>
<td>42.64%</td>
<td>0.0490</td>
<td>55.71%</td>
<td>0.0203</td>
<td>241%</td>
</tr>
<tr>
<td>Q3</td>
<td>5.91%</td>
<td>0.0445</td>
<td>50.00%</td>
<td>0.0201</td>
<td>222%</td>
</tr>
</tbody>
</table>

4.2.3 The true estimate (Combined Estimate)

Here we combine the estimates from the respondents and the non-respondents using the method developed by Hansen and Hurwitz (1946).

Table 9. The True Proportion of the Sensitive Behavior for the Respondents and Non-Respondents.

<table>
<thead>
<tr>
<th>Type of Sexual behavior</th>
<th>Proportion with the sensitive behavior</th>
<th>Variance of the sensitive behavior</th>
<th>Proportion with the sensitive behavior</th>
<th>Variance of the sensitive behavior</th>
<th>Relative Efficiency: $RE = \frac{Var(\hat{\pi}<em>{AW})}{Var(\hat{\pi}</em>{AV})}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>6.87%</td>
<td>0.0004</td>
<td>30.76%</td>
<td>0.0010</td>
<td>40%</td>
</tr>
<tr>
<td>Q2</td>
<td>26.18%</td>
<td>0.0010</td>
<td>47.30%</td>
<td>0.0012</td>
<td>83%</td>
</tr>
<tr>
<td>Q3</td>
<td>2.71%</td>
<td>0.0002</td>
<td>37.96%</td>
<td>0.0011</td>
<td>18%</td>
</tr>
</tbody>
</table>

4.3 Comparing the Efficiency of our Model with Kim and Warde (2004) and Kim and Elam (2007) Model

Table 10. The Variance of the Models and the Relative Efficiency

<table>
<thead>
<tr>
<th>QUESTIONS</th>
<th>SWRRT (2004)</th>
<th>SURRT (2007)</th>
<th>SWRRT (OUR MODEL)</th>
<th>SURRT (OUR MODEL)</th>
<th>* R.E 1 $= \frac{A}{C} \times 100$</th>
<th>** R.E 2 $= \frac{B}{D} \times 100$</th>
<th>***R.E3 $= \frac{C}{D} \times 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>0.0076</td>
<td>0.0021</td>
<td>0.0004</td>
<td>0.0010</td>
<td>1900%</td>
<td>210%</td>
<td>40%</td>
</tr>
<tr>
<td>Q2</td>
<td>0.0082</td>
<td>0.0025</td>
<td>0.0010</td>
<td>0.0012</td>
<td>820%</td>
<td>208%</td>
<td>83%</td>
</tr>
<tr>
<td>Q3</td>
<td>0.0074</td>
<td>0.0023</td>
<td>0.0002</td>
<td>0.0011</td>
<td>3700%</td>
<td>209%</td>
<td>18%</td>
</tr>
</tbody>
</table>

*R.E 1 is the relative efficiency of Kim and Warde (2004) and our model for Warner (1965).

**R.E 2 is the relative efficiency of Kim and Elam (2007) model and our model for Greenberg et al. (1969).
***R.E 3 is the relative efficiency of our model for Warner (1965) and our model for Greenberg et al. (1969).
Relative efficiency greater than 100% implies that denominator is better (or more efficient), otherwise the numerator is better (or more efficient).

From our results in the table above we can see that the Variances in column C and D (which are our models) are lesser than the variances in column A and B (which are the Kim and Warde (2004) and Kim and Elam (2007) models).

From R.E 1, our model for Warner (1965) is more efficient than the Kim and Warde (2004) considering the relative efficiency.
From R.E 2, our model for Greenberg et al (1969) is more efficient than Kim and Elam (2007) considering the relative efficiency.

We can also deduce from R.E 3 that the comparative study shows that our model for Warner (1965) is more efficient than our model for Greenberg et al. (1969).

![Graph showing relative efficiency](image)

**Fig 1. The relative efficiency of the variances.**

5 Conclusion

We have successfully presented a model for Warner (1965) using optimal allocation and for Greenberg et al. (1969) using optimal allocation. The empirical comparative study as shown that our model is more efficient in terms of variance than the Kim and Warde (2004) and Kim and Elam (2007) models. Likewise, the comparative study of our two models shows that our model for Warner (1965) is more efficient in terms of variance than our model for Greenberg et al. (1969).

As discussed earlier that Greenberg et al. (1969) encourage more cooperation from the respondent because of the unrelated paired question, which make it more efficient than Warner (1965) (See Greenberg et al. (1969)), but we have been able to cater for the deficiency of Warner (1965) by introducing subsample of non-respondent and this gave birth to a new note in the comparative study.

Conclusively, as stated earlier that Randomized Response Techniques helps to reduce response and non-response bias, in addition, our model further reduces the non-response bias.
References


The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage: http://www.iiste.org

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: http://www.iiste.org/journals/ All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digital Library, NewJour, Google Scholar