An Alternative to Warner’s Randomized Response Model

Sat Gupta
University of North Carolina at Greensboro

Javid Shabbir
Quaid-I-Azam University, jsqa@uq.edu

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm
Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation
DOI: 10.22237/jmasm/1162353900
Available at: http://digitalcommons.wayne.edu/jmasm/vol5/iss2/6
An Alternative to Warner’s Randomized Response Model

Sat Gupta                          Javid Shabbir
University of North Carolina at Greensboro     Quaid-I-Azam University

A modification to Warner’s (1965) Randomized Response Model is suggested. The suggested model is more efficient than the original model.

Key words: randomized response model, survey bias

Introduction

Warner (1965) suggested an Indirect Questioning method to circumvent social desirability response bias in surveys involving sensitive questions. If $\pi$ is the proportion of subjects in a population who have a sensitive characteristic A, then Warner’s method recommends using a randomization device, such as a deck of cards, to scramble the true response. A known proportion ($p$) of the cards in the deck have the statement “I have characteristic A” and the remaining cards in the deck have the statement “I do not have characteristic A”. A participant in the survey draws a card randomly from the deck and reports his/her agreement/disagreement with the statement on the card. Thus, a respondent who actually has the characteristic A, but draws a “I do not have characteristic A” card, will give a response “no” indicating lack of agreement with the statement on the card. Probability of a “yes” response ($p_y$) is given by

$$p_y = p\pi + (1 - p)(1 - \pi). \quad (1)$$

Equation (1) can be rewritten as

$$\pi = \frac{p_y - (1 - p)}{2p - 1}. \quad (2)$$

Equation (2) suggests estimating $\pi$ by $\tilde{\pi}_w$ where

$$\tilde{\pi}_w = \frac{n - (1 - p)}{2p - 1}, \quad p \neq .5 \quad (3)$$

where $n_1$ is proportion of “yes” responses in a simple random sample with replacement of size $n$. The fact that $n_1$ has a binomial distribution with parameters ($n, p_y$) can be used to prove that $\tilde{\pi}_w$ is a maximum likelihood estimator of $\pi$. Its variance is given by

$$V(\tilde{\pi}_w) = \frac{\pi(1 - \pi)}{n} + \frac{p(1 - p)}{n(1 - 2p)^2}. \quad (4)$$

The second term in the above expression is the penalty due to indirect responding. Note that the penalty is smallest when $p$ is closest to zero or one.

Several variations of Warner’s model have been proposed in the literature. These include models by Greenberg et. al (1969), Mangat and Siingh (1990) and Christofides (2002). Gupta and Thornton (2002) have attempted to validate some of these models with actual survey data.
Proposed Alternative Strategy

It is clear from (4) that Warner’s model works best when \( p \) is very close to zero or to one. But, both of these cases make the scrambling deck look very suspicious because almost all of the cards will be of the same type. Using two decks of the type described above is proposed, one with a low value of \( p \) (say \( p_1 \)) and the other with a high value of \( p \) (say \( p_2 \)). This will increase cooperation because the respondent is less suspicious in using decks of both kinds – one with a high value of \( p \) and one with a low value of \( p \).

A simple random sample with replacement of size \( n \) is selected and each respondent is asked to give a response using each of the two decks. Let \( (Z_{1i}, Z_{2i}), i = 1, 2, \ldots, n \) be the responses where \( Z_{ki} = 1 \) if the response using the \( k \text{th} \) deck \( (k = 1, 2) \) is “yes” and \( Z_{ki} = 0 \) if the response is “no”. Let \( n_{1i}, i = 1, 2, \ldots \) be the number of “yes” responses from the two decks. Then one can construct two estimators of the type (3). These are given by

\[
\tilde{\pi}_{w_1} = \frac{n_{1i} - (1 - p_1)}{2p_1 - 1}, \quad p_1 \neq .5, \quad (5)
\]

\[
\tilde{\pi}_{w_2} = \frac{n_{1i} - (1 - p_2)}{2p_2 - 1}, \quad p_2 \neq .5 \quad (6)
\]

It is easy to note from (4) that both of these estimators have the same variance if \( p_1 \) and \( p_2 \) are symmetric about \( .5 \).

We now propose the estimator

\[
\tilde{\pi}_p = k_1 \tilde{\pi}_{w_1} + k_2 \tilde{\pi}_{w_2}, \quad k_1 + k_2 = 1. \quad (7)
\]

Obviously \( \tilde{\pi}_p \) is unbiased because both \( \tilde{\pi}_{w_1} \) and \( \tilde{\pi}_{w_2} \) are unbiased. Also, variance of \( \tilde{\pi}_p \) is given by

\[
V(\tilde{\pi}_p) = k_1^2 V(\tilde{\pi}_{w_1}) + k_2^2 V(\tilde{\pi}_{w_2}) + 2k_1k_2 \text{Cov}(\tilde{\pi}_{w_1}, \tilde{\pi}_{w_2}). \quad (8)
\]

The following lemma is proven before exploring this variance further.

Lemma 1:

\[
\text{Cov}(\tilde{\pi}_{w_1}, \tilde{\pi}_{w_2}) = \frac{\pi(1 - \pi)}{n} \quad (9)
\]

Proof:

Note that \( n_{1i} = \sum_i Z_{1i} \).

Hence,

\[
\text{Cov}(n_{1i}, n_{12}) = \text{Cov}(\sum_i Z_{1i}, \sum_i Z_{2i}) = \sum_i \text{Cov}(Z_{1i}, Z_{2i})
\]

In the above block in the middle equation, please change the second summation to \( \sum_j Z_{2j} \) because \( Z_{1i} \) and \( Z_{2j} \) are independent for \( i \neq j \).

Hence,

\[
\text{Cov}(n_{1i}, n_{12}) = n\text{Cov}(Z_{1i}, Z_{12}) = n\{E(Z_{1i}Z_{12}) - E(Z_{1i})E(Z_{12})\}
\]

\[
= n\{p_{yi} - p_{yi}p_{y2}\}, \quad (10)
\]

where \( p_{yi} \) is the probability of a “yes” response with both decks, \( p_{yi} \) is the probability of a “yes” response with Deck 1 and \( p_{y2} \) is the probability of a “yes” response with Deck 2. The following is provided as in (1).

\[
p_{yi} = p_i \pi + (1 - p_i)(1 - \pi), i = 1, 2 \quad (11)
\]
and
\[ p_{xy} = \pi p_1 p_2 + (1 - \pi)(1 - p_1)(1 - p_2). \tag{12} \]
Substituting (11) and (12) in (10), one can easily obtain
\[ \text{Cov}(n_{i1}, n_{i2}) = n\pi(1 - \pi)(2p_1 - 1)(2p_2 - 1) \]
\[ \tag{13} \]

The lemma follows easily from (5), (6) and (13). Also, it is easy to verify that when \( p_1 + p_2 = 1, p_1 \neq p_2 \), the optimum values of \((k_1, k_2)\) in (8) are (.5, .5). This is because \( V(\hat{\pi}_{w1}) = V(\hat{\pi}_{w2}) \) if \( p_2 = 1 - p_1 \). With these choices for \((k_1, k_2)\), our proposed estimator becomes
\[ \hat{\pi}_p = \frac{\hat{\pi}_{w1} + \hat{\pi}_{w2}}{2}. \tag{14} \]

As remarked earlier, \( \hat{\pi}_p \) is unbiased because both \( \hat{\pi}_{w1} \) and \( \hat{\pi}_{w2} \) are unbiased.

Theorem 1:
When \( p_1 + p_2 = 1 \) and \( p_1 \neq p_2 \), estimator \( \hat{\pi}_p \) is more efficient than Warner’s estimators \( \hat{\pi}_{w1} \) and \( \hat{\pi}_{w2} \).

Proof:
Note that
\[ V(\hat{\pi}_p) = \frac{1}{4} V(\hat{\pi}_{w1}) + V(\hat{\pi}_{w2}) + 2 \text{Cov}(\hat{\pi}_{w1}, \hat{\pi}_{w2}) \]
\[ = \frac{1}{2} V(\hat{\pi}_{w1}) + \frac{1}{2} \frac{\pi(1 - \pi)}{n}, \]
because \( V(\hat{\pi}_{w1}) = V(\hat{\pi}_{w2}) \).

because
\[ \frac{\pi(1 - \pi)}{n} < V(\hat{\pi}_{w1}), \text{ from (4)}. \]

Numerical Examples
In this section, the efficiency of the proposed estimator is compared with Warner’s estimator for various choices of \( \pi, p_1 \) and \( p_2 (p_2 = 1 - p_1) \). Note that the proposed estimator is more efficient than Warner’s estimator, as expected, for all choices of the parameters.

Table 1: Efficiency of the proposed estimator compared to Warner’s estimator

<table>
<thead>
<tr>
<th>( \pi )</th>
<th>( p_1 = 0.1 )</th>
<th>( p_1 = 0.2 )</th>
<th>( p_1 = 0.3 )</th>
<th>( p_1 = 0.4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>143.860</td>
<td>171.171</td>
<td>187.940</td>
<td>197.087</td>
</tr>
<tr>
<td>0.2</td>
<td>130.529</td>
<td>158.140</td>
<td>180.398</td>
<td>194.937</td>
</tr>
<tr>
<td>0.3</td>
<td>125.084</td>
<td>151.414</td>
<td>175.758</td>
<td>193.458</td>
</tr>
<tr>
<td>0.4</td>
<td>122.659</td>
<td>148.077</td>
<td>173.222</td>
<td>192.593</td>
</tr>
<tr>
<td>0.5</td>
<td>121.951</td>
<td>147.059</td>
<td>172.414</td>
<td>192.308</td>
</tr>
<tr>
<td>0.6</td>
<td>122.659</td>
<td>148.077</td>
<td>173.222</td>
<td>192.593</td>
</tr>
<tr>
<td>0.7</td>
<td>125.084</td>
<td>151.414</td>
<td>175.758</td>
<td>193.458</td>
</tr>
<tr>
<td>0.8</td>
<td>130.529</td>
<td>158.140</td>
<td>180.398</td>
<td>194.937</td>
</tr>
<tr>
<td>0.9</td>
<td>143.860</td>
<td>171.171</td>
<td>187.940</td>
<td>197.087</td>
</tr>
</tbody>
</table>
Conclusion

The proposed strategy is likely to induce greater cooperation from the survey participants because it provides greater diversity in the scrambling process. Moreover, the proposed strategy is clearly more efficient than Warner’s model, particularly for higher values of $p_1$.

References


