Reinterview: A Modified Approach

Ghosh, Dhiren
Synectics for Management Decisions Inc.
1901 N. Moore St., Suite 900
Arlington, VA 22209, USA
E-mail: dhiren@smdi.com

Introduction

In any survey, simple response variance (SRV), correlated response variance (CRV) and response bias (RB) are important ingredients of measurement error. In the US Census Bureau model[1] for measurement errors, simple response variance is included in the estimate for sampling variance, but correlated response variance and response bias obviously are not included in the estimate for sampling variance. In the Census Bureau model correlated response variance arises due to the fact that an interviewer collecting data from a cluster of households which tends to record some values of some variables which are correlated. The reinterview technique is a powerful technique for estimating response variance and response bias. However there is a different kind of correlated response variance which arises because some respondents tend to remember the response given at the initial interview. The usual analysis of data obtained in the initial interview and reinterview for dichotomous data for simple random sampling is as follows:

		Interview 1	Interview 0	
Reinterview	1	a	b	a + b
	0	c	d	c + d
		a + c	b + d	

n = a + b + c + d.

Gross Difference Rate (GDR) = $\frac{b+c}{n}$. It can be shown E(GDR) = 2*SRV. A natural estimator of the proportion of random error that is associated with measurement error is given by the Index of Inconsistency (IOI) = $\frac{SRV}{SRV+SV} = \frac{GDR}{2s^2}$. Where s^2 is the estimate of the sampling variance from the sample. Thus the IOI takes the simple form $\frac{b+c}{2np(1-p)}$ where p = $\frac{a+c}{n}$. The chief reason for doing a reconciliation procedure after the interview is to estimate size of the response bias. An unbiased estimate of the response bias is given by Net Difference Rate (NDR) which can be shown to be $\frac{|c-b|}{n}$.

In the NDR calculation equal number of error of opposite direction offset each other and the remaining non offsetting part of the error is counted. In the GDR calculation there is no opportunity of the offsetting and thus the every error is counted.

In the calculation of SRV under the above model there is however the inclusion of the kind of correlated response variance mentioned above which arises because some respondents tend to remember the initial response. This kind of correlation response variance (correlated due to memory) as CRV^M.

We introduce a device in the reinterview technique for separating the simple response variance. The respondents are asked whether they remembered the initial response and on that basis respondents are divided into two groups. For the group that does not remember the initial response the usual estimate of response variable as above is an estimate of the true simple response variance. For the other group the variance between the responses constitute the correlated response variance CRV^M.

This method applies both for quantitative and qualitative data. A future paper would include the analysis for quantitative data.

The Mathematical Model

$$y_{it} = \mu_i + \varepsilon_{it}$$

$$E(\varepsilon_{it}|i) = 0$$

$$Var(\varepsilon_{it}|i) = \sigma_i^2$$

$$E\left(\sigma_i^2\right) = \sigma^2$$

$$Cov(\varepsilon_{it}\,\varepsilon_{i't'}) = 0 \quad for\ i \neq i', \qquad t = t'$$

$$Cov(\varepsilon_{it}\,\varepsilon_{i't'}) = 0 \quad for\ t \neq t'$$

$$Var(\bar{y}_i) = Var(\bar{\mu}) + \frac{\sigma^2}{n}$$

$$= SV + SRV$$

$$Where \qquad \bar{y}_i = \frac{1}{n}\sum y_{it} \ ; \ \bar{\mu} = \frac{1}{n}\sum \mu_i$$

$$GDR = \frac{b+c}{n} \qquad\qquad (binary\ case)$$

$$GDR = \frac{1}{n}\sum (y_{i1} - y_{i2})^2 \qquad (numeric\ data)$$

$$SRV = \frac{GDR}{2}$$

$$TRV = \frac{1}{n}\frac{\sum(y_{i1} - y_{i2})^2}{2} \qquad (numerical\ data)$$

$$TRV = \frac{1}{2n}\,(b + c) \qquad (binary\ data)$$

$$SRV = \frac{1}{2n}\sum(y_{i1} - y_{i2})^2 \qquad i = 1, \dots, n^-$$

$$TRV - SRV = Correlated\ Response\ Variance\ (CRV)$$

$$= \frac{1}{2n}\,(b + c) - \frac{1}{2n^-}\,(b^- + c^-)$$

$$= \frac{1}{2}\left[\frac{n^-(b + c) - n\,(b^- + c^-)}{nn^-}\right]$$

Where

$$a^- + a^+ = a$$
$$b^- + b^+ = b$$
$$c^- + c^+ = c$$
$$d^- + d^+ = d$$
$$n^- + n^+ = n$$

The Tables

Table 1

	Original Interview		
Reinterview	1	0	Total
1	a	b	a + b
0	c	d	c + d
Total	a + c	b + d	n = a + b + c + d

Table 2

Respondents who remember

	Original Interview		
Reinterview	**1**	**0**	**Total**
1	a^+	b^+	a^+ + b+
0	c+	d^+	$c^+ + d^+$
Total	$a^+ + c^+$	$b^+ + d^+$	$n^+ = a^+ + b^+ + c^+ + d^+$

Respondents who do not remember

	Original Interview		
Reinterview	**1**	**0**	**Total**
1	a^-	b^-	$a^- + b^-$
0	c^-	d^-	$c^- + d^-$
Total	$a^- + c^-$	$b^- + d^-$	$n^- = a^- + b^- + c^- + d^-$

REFERENCES

[1]Hansen, M., Hurwitz, W. and Pritzker, L. 1964. "The Estimation and Interpretation of Gross Differences and the Simple Response Variance", in C. Rao (ed.), *Contributions to Statistics.* Calcutta: Statistical Publishing Society.