Interviewer effect

Ruitao Zhang

In this paper, we will explore survey measurement errors which are mainly due to interviewers. By our method, we will be able to clearly solve some problems which were not very clearly stated in previous work. We will use the methods of permutation model, expanded permutation model and potential observable population to solve the problem.

First, we can clearly distinguish three different sources of randomness due to sampling, interviewer allocation, and interviewing. Second, we will be able to clearly describe and tell the difference among several ways of interviewer assignment. As a result, we can distinguish whether the interviewer effect is associated with labeled subjects or sample indices. Third, we will be able to estimate domain means and PSU means in different sampling design and interviewer assignment.

Interviewer assignment

The context we consider is for a finite population with units labeled by $s = 1, \ldots, N$ and finite interviewers labeled by $k = 1, \ldots, K$. The population is located in a finite number of geographic areas labeled by $a = 1, \ldots, A$. Suppose the data are collected by interviewers. A common survey question is: how to use the interviewers? i.e., how to draw the interviewer from interviewer group and assign them to the various individuals? Hansen and his college (1951) using mathematical language to describe a way to assign the interviewers:

(a) $n$ of the $N$ individuals in the population are selected at random without restriction.
(b) $k_a$ interviewers are selected at random without restriction from the $d^a$ interviewer group to interview those sample individuals who are available for interview by this interviewer group. Let $K = \sum_{a=1}^{d} k_a$ be the total number of interviewers selected.

(c) The same number, $\bar{n}$ of individuals is assigned to each of the $k_a$ interviewers. The $\bar{n}$ individuals assigned to any interviewer are a random subsample of all the sample individuals available for interview by this interviewer group.

Let $K = A$ and $k_a = 1$, Sarndal and his college (1992) summarize the assignment of interviewers into three ways: 1.) Deterministic assignment interviewers. 2.) Random assignment interviewers. 3. Interpenetrating subsample.

**Deterministic assignment interviewers (DAI)**

For each population subject, it is known beforehand who will interview him.

- there is a fixed set of interviewers labeled by $k = 1, \ldots, K$.
- prior to the survey, there is a fixed partitioning of the population into $K$ groups of subjects, so that each interviewer $k$ is linked to a unique group.
- the preassigned interviewer carries out all interviews with sampled elements from his own group.

**Example1.**

Suppose a finite population defined by a listing of 6 subjects, indexed by $s = 1, \ldots, 6$ and $a = 1, \ldots, 3$ interviewers. The population is located in $h = 1, \ldots, 3$ areas. Each area has two subjects (see Table1a). We use $y_{sa}$ to represent the response for subject $s$ by interviewer $a$ where $s = 1, \ldots, 6$ and $a = 1, 2, 3$. Interviewer 1, 2 and 3 are assigned to area 1, 2 and 3 respectively.
Table 1a. Potential observable responses for deterministic assignment

<table>
<thead>
<tr>
<th>Areas (h)</th>
<th>Subjects (s)</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>y_{11}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>y_{21}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>y_{32}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>y_{42}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td>y_{53}</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td>y_{63}</td>
</tr>
</tbody>
</table>

We use this concept of potential observable outcomes in formulating the interviewer assignment design. The average for interviewer \( a \) is defined by 

\[
\mu_{a} = \frac{1}{N_{a}} \sum_{s=2a-1}^{2a+2} y_{sa}. 
\]

In general, the average for interviewer \( a \) can be written as 

\[
\mu_{a} = \frac{1}{N_{a}} \sum_{s=1}^{N_{a}} y_{sa} \text{ where, } N_{a} \text{ is the total number of the subjects interviewed by interviewer } a. \]

The overall mean is 

\[
\mu = \frac{1}{N} \sum_{a=1}^{3} \sum_{s=2a-1}^{2a+2} y_{sa}. \]

In general, the overall average can be written as 

\[
\mu = \frac{1}{N} \sum_{s=1}^{N} y_{sa} \text{ where, } N \text{ is the total number of the subjects in the study.} \]

If \( a = 1 \) represents control, and \( a = 2 \) represents a treatment, the average causal effect of the treatment is given by 

\[
\mu_{2} - \mu_{1}. \]