Survey on Sexual Harassment by a family member, relative or close acquaintance

(by using Randomized Response Technique)

By,

SHRAMANA BHATTACHARYA
Roll No.- 427

Supervisor's Name :
Dr. Surupa Chakraborty

Declaration : ***I affirm that I have identified all my sources and that no part of my dissertation paper uses unacknowledged materials.***

INTRODUCTION

Survey research is useful in studying a variety of problems. The topic of this survey **is "*SEXUAL HARASSMENT BY A FAMILY MEMBER, RELATIVE OR ANY CLOSE ACQUAINTANCE*"** . My aim is to obtain an estimate of the population proportion of individuals who have been victims of such an incident, thereby getting an idea of the extent of such crimes still prevalent in our society.

Unwelcome sexual advances, requests for sexual favours, and other verbal or physical conduct of a sexual nature constitutes *sexual harassment*. When this sexual harassment is inflicted upon by someone close to the victim, it becomes a more grievous incident. Many such cases are hushed up inside the family; nonetheless the victim suffers from trauma.

Statistics provided by *RAINN* of America states that 79% of survivors who were victimized by a family member, close friend or acquaintance experience professional or emotional issues, including moderate to severe distress, or increased problems at work or school. In a *study of elderly female sexual abuse victims*, 81% of abuse was perpetrated by the victim's primary caregiver. 78% was perpetrated by family members, of whom 39% were sons. In *India*, almost all of these people who have been sexually harassed; are raped by their neighbours, parents, a close family member or other relatives. *Rapes in 15 out of India's 35 states and Union Territories are only committed by people known to the survivor*.

This survey has been performed by considering the undergraduate students of *Statistics Department of St. Xavier's College, Kolkata* who are aged between 18-21 years; as the population. It was difficult for me to carry out the survey on a larger population and hence I have performed it on a limited scenario. *Direct Survey* method has been used to get an unbiased estimate of the population proportion of victims. For various reasons individuals in a sample survey may prefer not to confide to the interviewer the correct answers to direct questions on such a sensitive topic. Hence, another appropriate method, that is, *Randomized Response Technique (RRT)* has been used to get an unbiased estimate of the population proportion.

Later, another survey has been carried out on a topic which is *not sensitive*; by both direct method and RRT. The topic of the second survey is **"*PREFERENCE FOR READING PAPER BOOKS OVER E-BOOKS*."** This survey has been done to assess the efficiency of the method of randomized response technique , so as to be certain about the findings of the first survey.

METHODOLOGY

A. DIRECT SURVEY METHOD

A random sample of size n is obtained from a population of size N possessing a particular character, say A. The individuals present in the sample are interrogated directly where the interviewer gets to know the respondent's identity. Let π be the proportion of individuals in the population who possess A.

$\hat{\pi}_D$ = **r/n** where r is the number of persons in the sample who possess A and $\hat{\pi}_D$ is an unbiased estimator of π.

B. RANDOMISED RESPONSE TECHNIQUE (RRT)

Randomized response is a research method used in structured survey interview which was first proposed by ***S. L. Warner*** in 1965 and later modified by ***B. G. Greenberg*** in 1969. It allows respondents to respond to sensitive issues while maintaining confidentiality. Examples are questions on intentional tax evasion, consumption of illegal drugs and so on. Respondents might feel embarrassed to answer truthfully to sensitive questions if asked directly in a direct survey, thereby leading to the technique of randomized response.

Two important randomized response techniques when the population is dichotomous are:-

✶ *DICHOTOMOUS POPULATION*

- *Warner's Model*
- *Unrelated Question Method*

I will only highlight the *Warner's Model* under a dichotomous population set-up and base my analysis of the survey data on this method.

WARNER's MODEL

Let a population possess a sensitive character A. We are interested to estimate the proportion of persons who possess the character A in the population. The population is thus dichotomous; some possessing the character A and the others possessing the complementary character Ã (i.e. they do not possess the character A). Because of the sensitive and often stigmatic nature of A, direct questions would result in biased estimates of the population proportion π as most of the respondents would give untruthful or evasive answers. To avoid this, Warner suggested the following randomized device.

A simple random sample of size n is drawn with replacement from a population of N individuals. Each respondent is provided with a randomization device using which he chooses one of the two questions:—

(a) *Do you belong to A*?
(b) *Do you belong to Ã*? ,
-- with probability P and (1-P), respectively, without the knowledge of the interviewer and answers "yes" or "no" according to his status with respect to the chosen question.

The value of P is known to the survey sampler.

The randomization device may be a spinner moving on a circular disc 100P% of the area of which denotes selection of question (a) and the remaining 100(1-P)% selection of question (b). The respondent will spin the pointer unnoticed by the interviewer. Many such randomization devices can be thought of, like a pack of cards containing two types of cards, denoting question (a) and (b) in the proportions P and (1-P) respectively. Privacy of the respondent is maintained because the interviewer gets only the "yes" or "no" answers without knowing the question to which the answer belongs.

Assuming that the respondent answers truthfully through the randomized response device, the probability of a "yes" answer is--

$$\lambda = \pi P + (1-\pi)(1-P) = (1-P) + (2P-1)\pi$$

Let r be the number of "yes" answers in the sample. An unbiased estimator of λ is-- $\hat{\lambda} = r/n$

Hence, an unbiased estimator of π is $\hat{\pi}_w = \frac{P-1}{2P-1} + \frac{r}{(2P-1)n}$, taking $P \neq \frac{1}{2}$

When P = 1, a direct survey response occurs and $\hat{\pi}_D = r/n$ which is the usual unbiased estimate of π.

$$V(\hat{\pi}_D) = \frac{\pi(1-\pi)}{n}$$

As, r follows Binomial(n,λ);

$$V(\hat{\pi}_w) = \frac{\lambda(1-\lambda)}{n(2P-1)^2} = \frac{\pi(1-\pi)}{n} + \frac{1}{n}\left(\frac{1}{16(P-0.5)^2} - \frac{1}{4}\right) = \frac{\pi(1-\pi)}{n} + \frac{P(1-P)}{n(2P-1)^2}$$

Since, $E\left(\frac{\hat{\lambda}(1-\hat{\lambda})}{n-1}\right) = \frac{\lambda(1-\lambda)}{n}$; an unbiased estimator of $V(\hat{\pi}_w)$ is

$$\widehat{V(\hat{\pi}_w)} = \frac{\hat{\lambda}(1-\hat{\lambda})}{(n-1)(2P-1)^2} = \frac{\widehat{\pi_w}(1-\widehat{\pi})}{n-1} + \frac{1}{n-1}\left(\frac{1}{16(P-0.5)^2} - \frac{1}{4}\right)$$

For direct survey, an unbiased estimator for $V(\hat{\pi}_D)$ is $\widehat{V(\hat{\pi}_D)} = \frac{\hat{\pi}_D(1-\hat{\pi}_D)}{n-1}$

$V(\hat{\pi}_w)$ has two parts and one part is same as $V(\hat{\pi}_D)$**.** The other part is due to the randomization introduced in the responses.

Clearly $V(\hat{\pi}_w) > V(\hat{\pi}_D)$ unless P = 1 or 0.

When P=1 or 0, $V(\hat{\pi}_w)$ is minimum.

When P is close to 1 or 0, $\hat{\pi}_w$ will have high efficiency.

For value of P close to 0.5, $V(\hat{\pi}_w)$ will be high.

If P=1, the first question is chosen with high frequency and the respondent will hesitate to answer truthfully because they would understand that a "yes" will almost always disclose their identity.

If P is close to 0.5, the privacy of the respondent is protected but now variance of estimate increases.

Thus the efficiency of the estimator and maintenance of privacy are conflicting matters.

From practical considerations, a value of P between 0.60 to 0.75 seems to be recommendable.

DATA ANALYSIS

- ***Data Description:***

I have considered a simple random sample with replacement of size n=100 out of N=184 students in the department of Statistics of St. Xavier's College; Kolkata. The sample consists of undergraduate students of first, second and third years. The data collected here is primary.

- ***Analysis:***

➢ *Direct Survey Method:* To perform the **direct survey**, every student in the sample has filled up questionnaires containing relevant questions regarding the topic. I have obtained an estimate of population proportion of victims which is $\hat{\pi}_D = \frac{10}{100} = 0.1$, where r=10 and n=100.

➢ *Randomized Response Technique:*

Warner's Model: I have considered a pack of 40 cards as my randomization device which contained 26 black and 14 red cards. The respondents used this randomization device to choose one of the two questions:

i. *Have you ever been sexually harassed by any of your family members, relatives or close acquaintances?*

ii. *Say 'yes' only if you support the following statement:*
"I have never faced any sexual harassment from any of my family members, relatives or close acquaintances."

If the respondent got a black card then he/she was supposed to answer the first question or else the second one should have been answered.

Here, P=26/40=0.65

n=100 and r is found out to be 44.

Hence, $\hat{\lambda}_w$=0.44

$\hat{\pi}_w$= 0.3, $\widehat{V(\hat{\pi}_w)}$ = 0.0276543 and $\hat{\pi}_D = 0.1$, $\widehat{V(\hat{\pi}_D)} = 0.0009091$

For the second survey……

Survey: Preference for reading paper books over e-books

With the advent of internets, several books are available online at zero or low cost and can be read anytime and anywhere. This survey has been performed to get an idea whether paper books are still in the preference of people over e-books.

A simple random sample of size n=100 have been taken out of N=184 students and the same procedure for direct survey and RRT have been followed here to obtain the following results:

- $\hat{\lambda}_d = \frac{49}{100} = 0.49 \approx 0.5 = \hat{\pi}_D$
- $\hat{\lambda}_w = \frac{51}{100} = 0.51$
- $\hat{\pi}_w = 0.5333 \approx 0.5$
- $\widehat{V(\hat{\pi}w)} = 0.0280471$
- $\widehat{V(\hat{\pi}_D)} = 0.0025242$

CONCLUSION

- The *estimate of the population proportion* of individuals who have been sexually harassed by a family member, relative or any close acquaintance obtained by Direct Survey and RRT, **has increased from 0.1 to 0.3.** The **variance has also increased from 0.0009091 to 0.0276543** which is quite evident from the formula used**.**

  Now, we can consider the following points:

  - The topic being sensitive, some respondents who have provided incorrect answers in direct survey have provided truthful answers in case of RRT which has lead to the increase in proportion estimate.
  - There might be presence of errors due to non-responses in survey and also due to reporting.
  - There is a small increase in proportion estimate. This can suggest that direct survey has obtained quite a number of truthful answers from respondents. This is quite satisfactory in this topic because it signifies that many victims are no longer trying to hide themselves and they want to speak about the wrong that had happened to them.

- In order to confirm the efficiency of Warner's Model and support the method of RRT thoroughly; a second survey on a lighter topic has been performed. The estimate of population proportion of individuals who prefer reading paper books to e-books is 0.5; that is half of the population still prefers reading paper books while the other half prefers e-books.

  - Here, there is no significant increase in the proportions obtained by direct survey and RRT. In fact, the two values are approximately same ($\hat{\pi}_D = 0.49 \approx \hat{\pi}_W$=0.5333). This topic being not at all sensitive, respondents have provided true answers when asked directly. Hence the proportion estimate in case of RRT and direct survey are approximately equal. The second survey helps us to confirm that the method of RRT is appropriate.
  - Evidently, the variance has increased from 0.0025242 to 0.0280471.

This survey has provided an estimate of sexually harassed students of Statistics Department of St. Xavier's College, Kolkata; by their own family members, relatives or close acquaintances. It has been obtained by the method of RRT, that 30% of the students have been victims of such an incident in the department. From this survey we can get an idea of the alarming situation regarding sexual harassment by someone known to the victim, still prevailing in our country where the respondents of this survey comprise of a very small part of the entire country's population.

BIBLIOGRAPHY

I have taken help from the following books and websites to complete my dissertation:-

- ***Theory And Methods of Survey Sampling***
  By,
  Parimal Mukhopadhyay

- ***Sampling Techniques***
  By,
  William G. Cochran
  Publisher: John Wiley and Sons

- ***https://en.wikipedia.org/wiki/Randomized_response***

- ***https://www.soundvision.com/article/15-tips-for-victims-on-how-to-deal-with-sexual-assault-abuse-and-harassment-in-the-west***

- ***http://www.aware.org.sg/2015/03/survey-1-in-3-young-people-have-faced-sexual-violence-few-seek-or-receive-help/***

- ***https://www.popline.org/node/345751***