

A System for Sensitive Characteristics on Internet Survey

Ki-Seong Nam¹⁾, Gi-Sung Lee²⁾, Geum-Min Song³⁾, Joo-Hwan Kim⁴⁾, Chang-Kyoon Son⁵⁾

¹⁾ Changwon University, Department of Education (kisnam@dreamx.net)

²⁾ Woosuk University, Division of Computer and Information Science (gisung@woosuk.ac.kr)

³⁾ Dongbusan College, Division of Digital Management (metakr@data2kdd.net)

⁴⁾ Dongguk University, Department of Science and Information Science, jhk@mail.dongguk.ac.kr)

⁵⁾ Dongshin University, Department of Computer Science (ckson85@blue.dongshinu.ac.kr)

Abstract

In this paper, we suggest and implement an internet survey system cooperating with the randomized response technique (RRT). RRT is an indirect question method to get truthful information from the respondent who is conceived to have sensitive characteristics without revealing his/her true status. We embody the system(method) based on database system, so it could be commonly used to all kinds of data obtained through internet surveys. This system also can be used on spot-survey independently.

We implement five RRTs : related question technique of Warner(1965), polychotomous question technique of Abul-Ela et al.(1967), unrelated question technique of Greenberg et al.(1969), forced answer technique of Loynes(1976) in qualitative data, and unrelated question technique of Greenberg et al.(1971) in quantitative data.

1. Introduction

Even if the on-line survey has a disadvantage such as lack of representativeness of survey sample, this survey methodology has developed rapidly as pointed by Kim et al.(1999), because this have several advantages relative to the off-line survey.

Recently, most business group carries out the on-line survey to get the management advice from customer, especially on-line members or employees. But it is difficult for an interviewer to take sensitive information from the on-line survey, because the respondent refuses to reveal his or her privacy. If the respondent is asked a sensitive question in on-line survey like the off-line, then they refuse the truthful answer because he or she is afraid of revealing his or her privacy or injustice treatment after response.

In a business survey to get the sensitive information such as agreement or disagreement for enterprise decision making or the amount of misappropriation of public goods, we consider that the indirect response method, randomized response technique (RRT) embodied in the on-line survey methodology.

Our survey system is developed to use on-line survey system, and at the same time on the independent spot-survey. Also our survey system has the following features; first of all, it restricts the answer at several times for the same ID or IP in the login site or non-login site. Secondly, it allows to use many people such as in computer laboratory.

This paper is composed of four chapters. The necessity of randomized response technique in on-line survey system for sensitive character is devoted in section 1. The randomized response technique is devoted in section 2. The implemented RRT system for on-line survey is devoted in section 3. Finally, the conclusion and further study are devoted in section 4.

2. Randomized Response Technique

In the survey having sociological sensitive characteristics, the respondent make refusal or untruthful answer to the survey question. The RRT that is answered indirectly by insuring true status or privacy of respondent is suggested by Warner(1965) to derive the information for sensitive characteristics. He would get the qualitative information for the sensitive characteristics, using the randomization device composed of the sensitive question and exclusive question.

Abul-Ela et al.(1967) proposed the polychotomous question technique which was extended the related question

technique for dichotomous population suggested by Warner. Greenberg et al.(1969) proposed the unrelated question technique, which was used unrelated question replaced by an exclusive question and Greenberg et al.(1971) extended it to the quantitative question technique. Also Loynes(1976) suggested the forced answer technique which was asked “yes” replaced by an exclusive question in Warner. Fox and Tracy(1986) and Chaudhuri and Mukerjee(1988) are summarized and arranged the RRT. Recently, it is interested in practical usage of RRT, especially it try to find the applications to the several science as sociology, management science and medical science. In Korea, Ryu et al.(1993) published the RRT textbook, and they suggested the practical usage of RRT by comparing and analyzing the case of this method.

In general, the RRT can be classified by the data and method used as in Table 1.

Table 1. Classification of RRT

Data	Classification		Remark
Qualitative data	Related question technique		Warner(1965)
	Unrelated question technique	Known characteristic	Greenberg et al.(1969)
		Unknown characteristic	
	Forced answer technique		Loynes(1976)
	Polytochomous question technique		Abul-Ela et al.(1967)
Quantitative data	Unrelated question technique	Known mean	Greenberg et al.(1971)
		Unknown mean	

To utilize the RRT and get more accurate information for the sensitive characteristics, it is desirable for the practical survey to imply from the following step.

First, investigate the survey topic is sensitive or not. Because the RRT is efficient to the survey with sensitive characteristics, first above all it should be investigated the sensitivity of the question from the pilot survey. The sensitivity of the question varies according to the region, gender, and age of survey unit.

Second, select the suitable RRT. One of problems in applying of RRT is the lack of understanding of the respondents for the method. Therefore, we should select RRT easier understanding and guarantee the anonymity of them.

Third, select a survey methodology. He or she must select a survey methodology such as face-to-face survey, mail survey, telephone survey, and internet survey by considering survey time and cost. But, the RRT and randomization device may be different with the selected survey methodology, it should be considered with these points.

Forth, select the randomization device. The randomization device should be selected with familiarity and ease by the respondent. In addition to, it would be sufficient to guarantee anonymity of respondent.

Fifth, describe of RRT and random device.

3. System implementation

3.1. Environment and System flow

In implementation of survey system, the computer languages used are the ColdFusion, Java and HTML, and the operating environment is developed on Linux. In addition, the database is MySQL of Microsoft.

The RRT survey system is composed of the administrator (interviewer) mode and the respondent (interviewee) mode. In the administrator mode, it operates as the following steps.

Step 1> Input survey topic.

Step 2> Construct survey items (including the survey period and target etc.).

Step 3> Select RRT, randomization device and probability for each item.

Step 4> Make the URL.

Step 5> Complete survey questionnaire and implement of survey.

The respondent mode is structured to answer for actual respondent. When the population ratio of unrelated characteristics is unknown or the population mean of unrelated variable is unknown, the respondent would be

participated in the answer separated sample 1 and sample 2 in accessing order.

The response result is stored in the database, the estimate of the population ratio(or mean) of this result only shows to the respondent, and all estimates including variance estimate of the population ratio(or mean) show to the administrator.

This survey system is based on the database in all processing from input to output of data. In this reason the response result is flexible to join with the other one in the repeated measurement for same respondent. Because of using the query in data processing, the implementation speed is faster than the file system. In addition, it is able to analyze the trend in longitudinal survey, because the system uses the database in data processing.

The database table for this survey system follows as Table 2.

Table 2. Data Base Table

Table: rrt_resp_item (Information)			
Logical Name	Physical Name	Data Type	Remark
Item Code	rr_i_Serial	Integer	PK, Index
Questionnaire code	rr_i_rsiSerial	Integer	FK, Index
Item List Code	rr_i_rilSerial	Float	FK, Index
Sensitive or Not	rr_i_Sense1_Nosense2	Integer	
Data Type	rr_i_ItemType	Date	
Item Contents	rr_i_Content	longtext	
Width	rr_i_Write_Length	Integer	
Total Respondents	rr_i_Resp_Count	Integer	
Yes Respondents	rr_i_Yes_Count	Integer	
Table: rrt_survey_info (Basic Information)			
Table: rrt_resp_list(Order Item)			
Table: rrt_resp_case (General item)			
Table: rrt_resp_data(Result)			

3.2 A example of survey system Implementation

This survey system is consists of the administrator mode and the respondent mode. We first explain the administrator mode. If you access to this survey site and log in, then you can see the window for survey items. On this site, if you click the “Add Questionnaire” button, then it appears to the next window as Fig.1. You must input survey topic, survey period and introduction of survey in this window. And then you click “Adding” button placed in bottom of this window, it appears to the survey information window as Fig.2. You confirm and input information in this window, and if you click “Adding Item” button placed in bottom of this window, it appears to the “Item Registration Wizard” window as Fig.3. You select one of RRT’s and input detail questionnaire information such as selection probability on this window, it completes the survey questionnaire. If you want to link this survey questionnaire on your site, then you have only insert the URL in Fig. 2 on it. From these survey questionnaire creation steps, the administrator is ready to an internet survey for a sensitive characteristics

RRT Home >>> Questionnaire Making	
Title of Questionnaire	<input type="text" value="Survey"/>
Restriction Respondents	<input checked="" type="radio"/> Not Restriction <input type="radio"/> Same User ID <input type="radio"/> Same IP Address
Survey Results	<input type="radio"/> Secret (Not Open to Public) <input type="radio"/> Public <input checked="" type="radio"/> Public After Ending
Term of Survey	<input type="text" value="2002/05/13 01:20:19"/> ~ <input type="text" value="2002/06/13 01:20:19"/> (Ex: 2001/05/19 13:30:00)
Introduction to Survey	<div>Proceeding</div> <div></div>
<input type="button" value="Adding Questionnaire"/>	

Fig. 1 Add questionnaire

RRT Home >>> Information Control of Questionnaire			
Survey Results	Survey Results	Total Respondents : 0 Person	Pre View
Questionnaire Link	Response the Survey		
Title of Questionnaire	RRT Survey		
Restriction Respondents	<input checked="" type="radio"/> Not Restriction <input type="radio"/> Members Only		
Survey Results	<input type="radio"/> Secret (Not Open to Public) <input type="radio"/> Public <input checked="" type="radio"/> Public After Ending		
Starting Survey	<input checked="" type="radio"/> Arrangement <input type="radio"/> Processing <input type="radio"/> Ending		
Number of Items per Page	0 Items (0 : All Items in One Page)		
Term of Survey	2002/05/12 12:00:00 ~ 2002/06/12 12:00:00 (Ex : 2003/05/19 13:30:00)		
Introduction to Survey	Proceeding <div style="border: 1px solid black; height: 100px;"></div>		

Fig. 2 Survey information

RRT Home >>> Information Control of Questionnaire >>> Wizard of Adding Item : Step 1	
Type of RRT	2. Unrelated Question Technique
Sensitive Question 1	0. General Question 1. Related Question Technique 2. Unrelated Question Technique 3. Two Sample Unrelated Question Technique 4. Forced Answer Technique 5. Unrelated Question Technique (Quantity Data) 6. Two Sample Unrelated Question Technique (Quantity Data) 7. Polychotomous Question Technique
Ex) [Sensitive Question 1] Ex) [Unrelated Question 2]	Have you never driven a motor vehicle while intoxicated? <div style="border: 1px solid black; height: 40px;"></div>
Unrelated Question 2	
Selection Ratio	■ The Selection Probability of Sensitive Question 1 : 0.3 (P ≠ 0.5) ■ The Population Proportion of Unrelated Group : 0.3
<div style="text-align: center;"> <input type="button" value="Adding of Item"/> <input type="button" value="Close Window"/> </div>	

Fig. 3 Item registration wizard

This survey system would select one of various randomization devices by the respondent, the one of them shows in Fig. 4. If the administrator inputs a selection probability of Question 1, it will appear head or tail, left or right of coin, according to the probability.



Fig. 4 The coin for randomization device

When the respondent accesses the survey site, it appears to the window for respondent as Fig. 5. At this time the respondent will be answer truthful to the question, because the administrator is unknown whether the respondent answer to question 1 or question 2.

Left >>>> You answer Question 1.

Right >>>> You answer Question 2.

Question 1.
Have you ever driven a motor vehicle while intoxicated?

☐ 1) Yes ☐ 2) No

Question 2.
Flip a coin, Did you get a head?

Fig. 5 Window for Respondent

3.3 Response results

The response results on this survey system appear in two different windows, the one only shows the estimate of the population ratio for a sensitive character as Fig.6, the other is showed all kinds of estimate of the population ratio for a sensitive characteristics as Fig.7. In this window, we can see the number of respondents, the population ratio estimate, and its variance estimate.

Question 3. Have you ever driven a motor vehicle while intoxicated?

The estimate of Population Ratio		
Yes	<div style="width: 16.67%; background-color: blue;"></div>	16.67 %
No	<div style="width: 83.33%; background-color: green;"></div>	83.33 %

Fig. 6 Window for Respondent

Question 1	Have you ever driven a motor vehicle while intoxicated?
Question 2	Flip a coin, Did you get a head?

Unrelated Question Technique : GreenBerg(1969) Technique		
Item Conditions	Population Proportion of Unrelated Group	0,50000
	Selection Prob. of Sensitive Question 1 (P)	0,30000

	Total Respondents	Yes Respondents	Estimate of Population Ratio	Estimate of Variance
Total	43	18	0,22868	0,06438
Male	23	11	0,42754	0,12602
Female	20	7	0,06140	0,14363

Fig. 7 Window for Administrator

4. Conclusion and further study

In this paper, we suggest and implement an internet survey system cooperating with the randomized response technique (RRT). RRT is an indirect question method to get truthful information from the respondent who is conceived to have sensitive characteristics without revealing his/her true status. Also we embody the system(method) based on database, so it could be commonly used to all kinds of data obtained through internet surveys.

We would expect that the survey system can be used to reduce untruthful responses for private or disadvantages at an enterprise or an assistance instrument for decision making in management.

Our survey system only uses the randomization device for a sensitive question cooperating with the other survey system, so that we can compare with other questions. Also, this survey system can be used to the independent spot-survey.

Finally, the further studies it will be possible to imply the various sampling designs for drawing respondents, and it will be used in practical survey for employees or customers having sensitive characteristics in the business surveys.

Reference

- [1] Abul-El, Abdel-Latif A., Greenberg, Bernard G., Horvitz, Paniel G.; A Multiproportions Randomized Response Model, Journal of the American Statistical Association, Vol. 62, pp 990-1008.
- [2] Chaudhuri, A. and Mukerjee, R.; Randomized Response : Theory and Techniques, New York, Marcel Dekker, Inc., 1988.
- [3] Coomber, R.; Using the Internet for Survey Research. Sociological Research Online, 2(2).
<http://www.socresonline.org.uk/socresonline/2/2/2.html>, 1997.
- [4] Fox, J. A. and Tracy, P. E.; Randomized Response : A Method for Sensitive Survey, Sage Publications, 1986.
- [5] Greenberg, B. G., Abul-El, Abdel-Latif A., Simmons, W. R., and Horvitz, D. G.; The Unrelated Question Randomized Response Model : Theoretical Framework, Journal of the American Statistical Association, Vol. 64, pp 520-539, 1969.
- [6] Greenberg, B. G., Kubler, R. R., Abernathy, J. R., and Horvitz, D. G.; Applications of the RR Technique in Obtaining Quantitative Data, Journal of the American Statistical Association, Vol. 66, pp 243-250, 1971.
- [7] Kim, J. K., Kim, H. J., Nam, K. S., Park, H. C., Lee, S. C. and Jeong, J. H.; Analysis of Social Survey, Changwon Univ. Press, 1999.
- [8] Ryu, J. B., Hong, K. H., and Lee, G. S.; Randomized Response Model, Freedom Academy, 1993.
- [9] Ryu, J. B., Lee, G. O., and Lee, G. S.; A Practical Plan of Randomized Response Technique, The Korean Journal of Applied Statistics, Vol. 8(1), pp 9-26, 1995.
- [10] Schwarz, C. J. StatVillage; An On-line, WWW-Accessible, Hypothetical City Based on Real Data for Use in an Introductory Class in Survey Sampling, Journal of Statistics Education, Vol. 5(2), <http://www.amstat.org/publications/jse>, 1997.
- [11] Warner, S. L.; Randomized Response; A Survey Technique for Eliminating Evasive Answer Bias, Journal of the American Statistical Association, Vol. 60, pp 63-69, 1965.