On the Interviewing Bias

By Hirojiro AOYAMA

(Received January 25, 1954)

1. Introduction

In performing surveys we adopt often the interviewing method. In such cases, however, we suffer more or less from interviewing bias. F. Mosteller [1] investigated the interviewing bias under some conditions. He assumed that the increment of favourable answer for the issue by favourable interviewers and the decrement of that by unfavourable interviewers were constant. But this assumption was not proved to be reliable. Recently M. H. Hansen and others [2] got some results about the response error. They appreciated the response variance by means of a certain mathematical model in which the random samples in every class of the population concerned are assigned to the interviewers selected at random from the corresponding interviewer class. They, however, had not the estimate about the response bias, but only that about the response variance.

In this paper we shall give a method of estimation of the interviewing bias under the practical conditions which we see often to hold in our country.

2. Mathematical model

Let $i\,(1 \leqq i \leqq R)$ denote the class of interviewers, $j\,(1 \leqq j \leqq L_i)$ the interviewer in class i, and $k\,(1 \leqq k \leqq n_{ij})$ the sample which was assigned to interviewer j of class i.

Under certain conditions in which the influence of the communication between interviewers and samples plays an essential role (see M. H. Hansen and others [2]), we can get the true value x_{ijk} of sample k who is interviewed by interviewer j of class i. But the interviewer can get only the sample value y_{ijk} which has been transformed through interview from the sample's true value x_{ijk}, that is, he can get only

$$(1) \qquad y_{ijk} = T(x_{ijk})$$

Here we do not consider the fluctuation of interviewee's answer, for we can treat this error by the variance independently of our bias.

If we can represent this transformation by formulas in probability such as

$$(2) \qquad \begin{aligned} P_r\{T_{ij}(x_{ijk}) = x_{ijk}\} &= p_{ij} \\ P_r\{T_{ij}(x_{ijk}) = Z_{ij}\} &= 1 - p_{ij} = q_{ij} \end{aligned}$$

where Z_{ij} are the interviewers' own characteristic values, we have easily some appreciation about the interviewing bias. In the ordinary survey of our country we employ students as interviewers. Usually they bring a few false reports which come from sample's absence, their own fatigue and so on. In such cases, however, they note down the answers by their own opinions as those of the interviewee. According to our experimental research about the interviewers, which was held in practical interviews, they happen to note down wrong for the question instead of noting the previously defined right answer. But, anyway, we may think that this takes place with small probability, and by means of an experimental research we can estimate the probability p_{ij} in (2).

If these estimates are obtained, we can get the sample mean

$$(3) \qquad \bar{y}=\frac{1}{n}\sum_{i=1}^{R}\sum_{j=1}^{L_i}\sum_{k=1}^{n_{ij}} y_{ijk}$$

the expected value

$$(4) \qquad E(\bar{y})=\frac{1}{n}\sum_{i=1}^{R}\sum_{j=1}^{L_i} n_{ij}(p_{ij}\bar{X}+q_{ij}Z_{ij})$$

where $\bar{X}$ is the population mean and Z_{ij} is interviewer j's characteristic value in class i, and the variance

$$(5) \qquad V(\bar{y})=\frac{1}{n^2}\sum_{i=1}^{R}\Big[\sum_{j=1}^{L_i} n_{ij}\{p_{ij}(\sigma^2+\bar{X}^2)+q_{ij}Z_{ij}^2\}+\sum_{j=1}^{L_i} n_{ij}(n_{ij}-1)$$
$$\times(p_{ij}^2\bar{X}+2p_{ij}q_{ij}\bar{X}\bar{Z}_{ij}+q_{ij}^2Z_{ij}^2)+\sum_{j\neq j'}^{L_i}\sum n_{ij}(p_{ij}\bar{X}+q_{ij}Z_{ij})$$
$$\times n_{ij'}(p_{ij'}\bar{X}+q_{ij'}Z_{ij'})-\{\sum_{j=1}^{L_i} n_{ij}(p_{ij}\bar{X}+q_{ij}Z_{ij})\}^2\Big]$$

where σ^2 is the population variance.

Further, if we can put $L_i=L$, $n_{ij}=\dfrac{n}{RL}=\bar{n}$, $p_{ij}=p_i$, $Z_{ij}=\bar{Z}_i$, we have

$$(6) \qquad E(\bar{y})=\frac{\bar{X}}{R}\sum_{i=1}^{R}p_i+\frac{1}{R}\sum_{i=1}^{R}q_i\bar{Z}_i$$

$$(7) \qquad \text{bias}=\bar{X}-E(\bar{y})=\frac{1}{R}\sum_{i=1}^{R}q_i(\bar{X}-\bar{Z}_i)$$

$$(8) \qquad V(\bar{y})=\frac{1}{Rn}\left\{\sigma^2\sum_{i=1}^{R}p_i+\sum_{i=1}^{R}p_iq_i(\bar{X}-\bar{Z}_i)^2\right\}$$

that is, if we use the same number of interviewers in each class and samples of the same size, and can assume that the transformation probabilities and the interviewers' opinions are all the same in each class, respectively, we get very simple results as above.

3. Example

In the pre-survey of the mathematico-statistical research of the national character of Japan [3] which was performed by our Institute, we could consider that the transformation probability is 0.95, speaking on the average. And the percentages of supporting political parties of the people in Tokyo-to were as the following tables, where α, β, γ and δ indicate experienced students, inexperienced students, experienced men and inexperienced men, respectively. In these tables it may be seen that there is some bias introduced by interviewers of different classes, but there are no significant differences among the supporting percentages.

Table 1.

Party / Inter-viewer class	Liberal Party	Pro-gressive Party	Social Party, right wing	Social Party, left wing	Only Social Party	Com-munist	Mis-cella-neous	Non sup-porting	D.K.	Refuse	Total
α	30.8	2.2	24.3	8.1	4.3	—	1.1	16.8	10.8	1.6	100.0
β	33.0	5.9	12.5	12.5	10.0	1.8	—	16.6	6.5	1.2	100.0
γ	35.9	5.0	19.9	6.1	3.9	0.5	—	14.4	12.1	2.2	100.0
δ	35.4	3.2	24.5	5.2	3.2	1.3	1.3	11.6	11.6	2.6	100.0
Total	33.8	4.1	20.2	8.0	5.3	0.9	0.6	14.9	10.3	1.9	100.0

Table 2.

	Liberal—Social	Conservative—Progressive *
α	−5.9	−3.7
β	−2.0	2.1
γ	6.0	10.5
δ	2.5	2.5

*Conservative = Liberal + Progressive Party
Progressive = Social Party + Communist

In our survey we put

$$R=4, \quad n_{ij}=\bar{n}=10, \quad L_i=L=22, \quad n=4\times 22\times 10=880$$

and took interpenetrating samples of 22 spots in Tokyo-to. When we assume here

$\bar{X}=P=$ percentage supporting of Liberal Party $=0.34$

$$\bar{Z}_\alpha=\bar{Z}_\beta=0 \qquad \bar{Z}_\gamma=\bar{Z}_\delta=1$$

we have

$$E(\bar{y})=0.348$$

$$\text{bias}=0.008$$

$$V(\bar{y})=0.0002571$$

$$\text{M.S.E.}\,(\bar{y})=0.0003211$$

and the interviewing bias is so small that we can neglect it in our problem.

INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] F. Mosteller: Correction for Interviewer Bias, Appendix II of *Gauging Public Opinion* by H. Cantril and others, 1947.

[2] M. H. Hansen, W. N. Hurwitz, E. S. Marks and W. P. Mauldin: Response Error in Surveys, *Journal of American Statistical Association*, vol. 46, 1951.

[3] Members of the Institute of Statistical Mathematics: Report of the Survey of the National Character of Japan, the *comming issue of The Proceedings of the Inst. of Statistical Mathematics*, Tokyo.