

## Chapter 2

# Specification of Qualitative and Quantitative Parameters Demanding Estimation

**Abstract** In this chapter the basic and rudimentary aspects of sample surveys for finite populations are presented in a compact way. The concepts of population, sample, sampling design, survey data, estimating finite population parameters of interest and consequent errors and their control will be explained and detailed illustrations will be provided. The theory to address general issues will be explained first. Then the need for modification to cover the case of sensitive issues and how to do that will be explained. It will be clearly shown how in a general situation one may handle indirectly procured observations to estimate parameters of interest and also derive estimated measures of accuracy. Sophisticated theoretical details will be presented only in brief. Finally in this chapter we put emphasis on the fact that any probability sampling design may be employed for the purpose of estimating parameters related to stigmatizing characteristics.

### 2.1 Introduction

A labeled finite population of identifiable individuals, each bearing real numbered values including zero and one is supposed to be surveyed through sampling and ascertaining sample-wise values to estimate certain parameters. Specifically, we take up the case of individual human beings, some of whom bear sensitive, rather stigmatizing features. A way out is needed to successfully gather individual values for respective people sampled so as to estimate proportions of a community bearing such sensitive features or the total value for a stigmatizing characteristic borne collectively by the members of a community.

As the title of this monograph announces it is avowedly one on survey sampling. So, we are under an obligation to tell our readers certain basic and rudimentary aspects of sample surveys.

By  $U = (1, \dots, i, \dots, N)$  we shall mean a survey population. It refers to a known finite number  $N$  of individuals labeled for identification uniquely by  $i$  which stands for the label 1 through  $N$ , each denoting just one of these  $N$  units. Each

unit  $i$  bears a value  $y_i$  for a real variable  $y$ ,  $i = 1, \dots, N$ . Thus the vector  $\underline{Y} = (y_1, \dots, y_i, \dots, y_N)$  is defined on  $U$ . Then

$$Y = \sum_{i=1}^N y_i, \quad \bar{Y} = \frac{Y}{N}, \quad S^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2$$

are some of the parameters of common interest related to  $\underline{Y}$ . The quantities  $Y$  and  $\bar{Y}$  are usually required to be estimated, called, respectively, the total and mean of  $y$  related to  $U$ . The quantity  $S^2$  is the population variance. For estimation, a sample  $s$  of elements of  $U$  is required to be chosen with a pre-assigned probability  $p(s)$  according to a suitable probability design or design in brief for simplicity, say, such that  $0 \leq p(s) \leq 1$  for every sample  $s$  that may be selected from  $U$ . The number of units in  $s$ , say,  $n$  is called the size of  $s$ , just as  $N$  is the size of the population  $U$ . On choosing a sample, the values of  $y$  for the units in  $s$ , that is,  $y_i$  for  $i \in s$  are to be ascertained by dint of an actual sample survey. On gathering the survey data

$$d = (s, y_i | i \in s),$$

an estimator for  $Y$  may be employed as  $t = t(d)$ , usually with the properties

$$E_p(t) = \sum_s p(s)t(d) = Y$$

and

$$M_p(t) = E_p(t - Y)^2 = \sum_s p(s)(t(d) - Y)^2,$$

being suitably small. The quantity  $M_p(t)$  is called the mean square error, and  $E_p$  denotes the expectation operator with respect to the sampling design  $p$ . Such a  $t$  is called an estimator for  $Y$  with its error  $(t - Y)$  such that  $M_p(t)$  is suitably under control. If  $K$  is a pre-assigned positive number, noting that we may write

$$M_p(t) = \sum_1 p(s)(t(d) - Y)^2 + \sum_2 p(s)(t(d) - Y)^2 \geq K^2 P[s : |t(d) - Y| \geq K],$$

where  $\sum_1$  is the sum over those samples for which  $|t(d) - Y| \geq K$  and  $\sum_2$  that over the complementary set of possible samples, it follows that

$$P[t(d) - K < Y < t(d) + K] \geq 1 - \frac{M_p(t)}{K^2}.$$

The quantity

$$B_p(t) = E_p(t - Y)$$

is called the bias of  $t$  in estimating  $Y$  and

$$\sigma_p(t) = +\sqrt{\sigma_p^2(t)} = +\sqrt{V_p(t)}$$

is the standard error of  $t$ , where

$$V_p(t) = E_p(t - E_p(t))^2,$$

is the variance of  $t$ . Now

$$M_p(t) = V_p(t) + B_p^2(t).$$

Consequently, by taking  $K = \lambda\sigma_p(t)$ , with  $\lambda > 0$ , it follows that

$$P[t(d) - \lambda\sigma_p(t) < Y < t(d) + \lambda\sigma_p(t)] \geq \left(1 - \frac{1}{\lambda^2}\right) - \frac{1}{\lambda^2} \left(\frac{|B_p(t)|}{\sigma_p(t)}\right)^2.$$

If  $t$  is chosen to be unbiased for  $Y$ , then

$$P[Y \in (t(d) - \lambda\sigma_p(t), t(d) + \lambda\sigma_p(t))] \geq \left(1 - \frac{1}{\lambda^2}\right).$$

The interval

$$t(d) \pm \lambda\sigma_p(t)$$

is called a confidence interval (CI) for  $Y$  with a confidence coefficient (CC) of at least

$$\left(1 - \frac{1}{\lambda^2}\right).$$

So, it is desirable to employ an unbiased estimator  $t$  for  $Y$  with  $B_p(t) = 0$  and a small variance because in that case one may derive a confidence interval based on  $t(d)$  with a small width equal to  $2\lambda\sigma_p(t)$ . For example, a choice of  $\lambda = 3$  will yield

$$t(d) \pm 3\sigma_p(t)$$

as a confidence interval with a confidence coefficient of at least as high as 8/9. [Neyman \(1934\)](#) gave us this gift for survey sampling extending Chebyshev's inequality in probability theory.

Unfortunately however, as [Basu \(1971\)](#) has shown, unless a census, i.e., a complete enumeration of a survey population is undertaken, no sampling design admits an unbiased estimator for a finite population total ensuring a minimum value for its variance, uniformly for every  $\underline{Y} = (y_1, \dots, y_i, \dots, y_N)$  assigning any real

number to each coordinate of  $\underline{Y}$ . Following [Godambe \(1955\)](#), let us restrict to the use of an estimator  $t = t(d)$  of the form

$$t = t_b = \sum_{i \in s} y_i b_{si}$$

such that  $b_{si}$  are numbers free of  $\underline{Y}$  subject to the restriction

$$\sum_{s \ni i} p(s) b_{si} = 1, \quad \forall i \in U, \quad (2.1)$$

where  $\sum_{s \ni i}$  means summation over all samples  $s$  which contain the unit labeled  $i$ . The condition given by (2.1) is necessary as well as sufficient to render  $t_b$  unbiased for  $Y$ . Unfortunately again, [Godambe \(1955\)](#) has shown that in the class of all such estimators for  $Y$  of the form  $t_b$  as above, called the class of Homogeneous Linear Unbiased Estimators (HLUE), no one exists with a Uniformly Minimum Variance (UMV), so long as it is based on a “general” class of designs  $p$ . [Godambe \(1955\)](#) did not specify what he meant by his “general” class. But [Hege \(1965\)](#) and [Hanurav \(1966\)](#) independently showed that a general class of designs called Unclustered Class of Designs (UCD) exists for which the above negative result does not hold. A design  $p$  belonging to a UCD is a design such that for any two samples  $s_1$  and  $s_2$  with  $p(s_1) > 0$  and  $p(s_2) > 0$  either

1.  $s_1 \cap s_2$  is empty, i.e.,  $s_2$  does not intersect with  $s_1$

or

2.  $s_1 \sim s_2$ , i.e., every unit of  $s_1$  is in  $s_2$  and vice versa.

Thus, [Godambe's \(1955\)](#) above celebrated nonexistence result is valid only for non uni-cluster-designs (NUCD).

[Hege \(1965\)](#), [Hanurav \(1966\)](#), and quite elegantly [Lanke \(1975\)](#) have shown that a UCD admits a UMV estimator for  $Y$  in the HLUE class. Importantly, this UMV estimator in the class HLUE for  $Y$  is of the form

$$t_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i},$$

where  $\pi_i = \sum_{s \ni i} p(s)$ . The estimator  $t_{HT}$  is given by [Horvitz and Thompson \(1952\)](#). The quantity  $\pi_i$  is called the “inclusion-probability” of a unit  $i$  in  $U$  being included in a sample chosen according to a design  $p$ . It may be noted that there is no problem in taking  $\pi_i$  into the denominator in  $t_{HT}$  because a well-known fact in “Survey Sampling Theory” is that a “Necessary and Sufficient Condition” that an unbiased estimator for  $Y$  may exist is that

$$\pi_i > 0, \quad \forall i \in U.$$

See [Chaudhuri \(2010\)](#) for details.

Most of the estimators for  $Y$  used in practice are in the HLUE class. If we add the terms  $b_s$  to  $t_b$  such that  $b_s$  is free of  $\underline{Y}$  and  $E_p(b_s) = 0$ , then we get

$$t_L = b_s + \sum_{i \in s} y_i b_{si},$$

a nonhomogeneous Linear Unbiased Estimator for  $Y$ , the LUE class. No estimator for  $Y$  outside this LUE class is ever put into practice except that the unbiasedness conditions

$$E_p(b_s) = 0 \text{ and } \sum_{s \ni i} p(s) b_{si} = 1 \quad \forall i \in U$$

are often relaxed. The estimators  $t_L$ ,  $t_b$  for  $Y$  are admitted provided the quantities  $B_p(t_L)$  and  $B_p(t_b)$  are numerically small enough, at least for large  $n$ . The HLUEs for  $Y$  and in particular the [Horvitz and Thompson's \(1952\)](#) estimator, say, are the most popular. Added to them, [Hajek's \(1971\)](#) ratio estimator

$$t_H = \frac{X \sum_{i \in s} (y_i / \pi_i)}{\sum_{i \in s} (x_i / \pi_i)},$$

where  $x_i$  are known (positive) values of a variable  $x$ , highly and positively correlated with  $y$  and where  $X = \sum_{i=1}^N x_i$ , mostly exhaust the popular estimators for  $Y$ . This ratio estimator with its denominator even as a random variable as  $s$  is so, is yet in Godambe's class of linear homogeneous estimators for  $Y$  because this  $t_H$  is linear in the  $y_i$ 's and there is no term in it free of  $\underline{Y}$  and cannot be nonhomogeneous.

We consider it important to observe:

$$V_p(t_b) = E_p(t_b - Y)^2 = \sum y_i^2 C_i + \sum_i \sum_{j, j \neq i} y_i y_j C_{ij},$$

where

$$C_i = \sum_{s \ni i} p(s) b_{si}^2 - 1$$

and

$$C_{ij} = \sum_{s \ni i, j} p(s) b_{si} b_{sj} - 1.$$

The quantity

$$v_p(t_b) = \sum_{i \in s} y_i^2 C_{si} + \sum_{i \in s} \sum_{j \in s, j \neq i} y_i y_j C_{sij}$$

is an unbiased estimator for  $V_p(t_b)$  on choosing  $C_{si}$  and  $C_{sij}$ 's as free of  $\underline{Y}$  subject to the conditions

$$\sum_s p(s) C_{si} = C_i, \quad \forall i \in U$$

and

$$\sum_s p(s) C_{sij} = C_{ij}, \quad \forall i, j \in U, i \neq j.$$

More generally, on choosing

$$\underline{W} = (w_1, \dots, w_i, \dots, w_N), \quad w_i \neq 0 \quad \forall i \in U$$

as a vector of pre-assigned constants, it follows that

$$V_p(t_b) = - \sum_i \sum_{j, j > i} d_{ij} w_i w_j \left( \frac{y_i}{w_i} - \frac{y_j}{w_j} \right)^2 + \sum_i \frac{y_i^2}{w_i} \alpha_i,$$

where

$$d_{ij} = E_p(b_{si} I_{si} - 1)(b_{sj} I_{sj} - 1),$$

$$I_{si} = \begin{cases} 1 & \text{if } s \ni i \\ 0 & \text{otherwise} \end{cases}$$

and

$$\alpha_i = \sum_j d_{ij} w_j.$$

Writing

$$\pi_{ij} = \sum_{s \ni i, j} p(s)$$

as the probability of  $i$  and  $j$  ( $i \neq j$ ) both included in a sample chosen according to a design  $p$  and in addition supposing  $\pi_{ij} > 0 \quad \forall i \neq j$  and of course  $\pi_i > 0 \quad \forall i \in U$  to get  $t_b$  unbiased for  $Y$ , one may get an unbiased estimator for  $V_p(t_b)$  as

$$v_p(t_b) = - \sum_{i \in s} \sum_{j \in s, j > i} d_{sij} w_i w_j \left( \frac{y_i}{w_i} - \frac{y_j}{w_j} \right)^2 + \sum_{i \in s} \frac{y_i^2}{w_i} \frac{\alpha_i}{\pi_i},$$

on taking the  $d_{sij}$ 's as free of  $\underline{Y}$  subject to

$$\sum_s p(s) d_{sij} = d_{ij}, \quad \forall i \neq j,$$

e.g., as  $d_{sij} = (d_{ij}/\pi_{ij}) \forall i \neq j$  in  $s$  and in  $U$ . In particular,

$$V_p(t_{HT}) = \sum_{i=1}^N y_i^2 \frac{1 - \pi_i}{\pi_i} + \sum_i \sum_{j, j \neq i} y_i y_j \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}$$

and

$$v_p(t_{HT}) = \sum_{i \in s} y_i^2 \frac{1 - \pi_i}{\pi_i} \frac{I_{si}}{\pi_i} + \sum_{i \in s} \sum_{j \in s, j \neq i} y_i y_j \left( \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \right) \frac{I_{sij}}{\pi_{ij}}.$$

Also, alternatively,

$$V_p(t_{HT}) = \sum_i \sum_{j, j > i} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 + \sum \frac{y_i^2}{w_i} \beta_i,$$

taking  $w_i = \pi_i$  and writing

$$\beta_i = 1 + \frac{1}{\pi_i} \sum_{j, j \neq i} \pi_{ij} - \sum_{i=1}^N \pi_i$$

and an unbiased estimator for  $V_p(t_{HT})$  is

$$v_p(t_{HT}) = \sum_{i \in s} \sum_{j \in s, j > i} \left( \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \right) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 + \sum \frac{y_i^2}{\pi_i} \beta_i \frac{I_{si}}{\pi_i}.$$

It should be noted that if every sample  $s$  with  $p(s) > 0$  has a constant number of distinct units in it, then  $\beta_i = 0$ . Consequently  $V_p(t_{HT})$  and  $v_p(t_{HT})$  take the familiar forms due to [Yates and Grundy \(1953\)](#) namely,

$$V_{YG} = \sum_i \sum_{j, j > i} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2$$

and

$$v_{YG} = \sum_{i \in s} \sum_{j \in s, j > i} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \frac{1}{\pi_{ij}},$$

respectively.

*Remark 2.1.* As in Chaudhuri's (2011) textbook, the spirit in which this monograph has been written demands a reader/researcher to feel comfortable with randomized response data (and data produced by other indirect questioning techniques to be illustrated in this monograph), being procured from sampled persons in specifically prescribed ways no matter how selected, so that inferences may be drawn and suitably assessed on choosing the samples in desirable ways, with equal or unequal selection probabilities, with or without replacement. In a vast majority of the space covered in publications on data gathered by indirect questioning techniques, samples are chosen by Simple Random Sampling with Replacement (SRSWR). In such a case, the simple arithmetic mean of suitably transformed individual observations is employed to unbiasedly estimate  $\bar{Y}$  and variances and variance estimators are derived along the standard SRSWR procedure. In very few cases, when  $y$  is just a binary variable taking on the values 0 and 1 representing an innocuous and stigmatizing feature respectively borne by a person, samples are not chosen by simple random sampling with replacement.

Chaudhuri's (2011) textbook is claimed to offer a quick appreciation of our approach set forth here.

## 2.2 Estimating Parameters

We shall consider only two possibilities:

- (I) Every  $y_i$  is either 1, implying a person labeled  $i$  bears a stigmatizing characteristic or attribute, say,  $A$ , or 0 implying the  $i$ -th person's attribute being  $A^c$ , i.e., the complement of  $A$ , meaning that the person does not have the stigmatizing characteristic; our problem is to estimate  $\bar{Y}$  to be denoted by  $\theta$ , the proportion in a community  $U$  bearing  $A$ .
- (II) Every  $y_i$  is a real number and our aim is to estimate the population total  $Y$ .

From our treatment in Sect. 2.1 an outline of a possible procedure readily follows. As a measure of accuracy in estimation we recommend evaluating the coefficient of variation (CV) as

$$CV = 100 \frac{+\sqrt{v_p(t)}}{t}$$

for a choice of  $t$  as illustrated.

Since  $t_H$  is not unbiased for  $Y$  we need to consider only its Mean Square Error (MSE)

$$M(t_H) = E_p \left[ X \frac{t_{HT}(y)}{t_{HT}(x)} - Y \right]^2$$



approximately taken as, on assuming  $n$  quite large,

$$M(t_H) = X^2 V_p(t_{HT}) \mid_{\underline{Y}=\underline{Y}-R\underline{X}},$$

where  $R = (Y/X)$  and recognizing  $V_p(t_{HT})$  as a formula for  $V_p(\sum_{i \in s} (y_i/\pi_i))$  evaluated with  $y_i$  replaced by  $d_i = y_i - Rx_i$ ,  $i \in U$ . By  $t(y)$ ,  $t(x)$  we mean an estimator for  $Y$  using  $y_i$ ,  $i \in s$  and the same one with  $x_i$ ,  $i \in s$ . Then a plausible estimator for  $M(t_H)$  is taken as

$$m(t_H) = [t_{HT}(x)]^2 v_p(t_{HT}) \mid_{\underline{Y}=\underline{Y}-\hat{R}\underline{X}},$$

writing  $\hat{R} = t_{HT}(y)/t_{HT}(x)$  and in the formula for  $v_p(t_{HT})$  taking  $y_i - \hat{R}x_i$  throughout for  $i \in s$  in lieu of  $y_i$ ,  $i \in s$ . Chaudhuri (2010) may be consulted for clarification. Needless to mention the coefficient of variation for  $t_H$  will be taken as

$$CV(t_H) = 100 \frac{+\sqrt{m(t_H)}}{t_H}$$

as an approximation, assuming  $n$  large enough.

All these are applicable only provided  $y_i$  is available for  $i$  in  $s$ . But in the situations of our interest, they are not. An investigator, out of sheer delicacy hesitates to ask a respondent to disclose his/her value on a sensitive variable. Even if he/she dares, a respondent is likely to refuse or hide the truth giving an untruthful or misleading response.

In the subsequent chapters we shall narrate diverse procedures to tackle this situation. The gist is that we may point out at this stage that by a suitable “random” mechanism we claim we may gather observations  $r_i$ , “independently” across  $i$  in  $s$ , such that

$$E_R(r_i) = y_i, \forall i \in U,$$

where  $E_R$  denotes a generic expectation operator with respect to such a random mechanism employed.

In situation (I) when  $y_i$  takes on either the value zero or one only,

$$\begin{aligned} V_R(r_i) &= E_R(r_i - E_R(r_i))^2 \\ &= E_R(r_i^2) - y_i^2 \\ &= E_R(r_i^2) - y_i \\ &= E_R[r_i(r_i - 1)] \\ &= V_i, \end{aligned}$$

say, writing  $V_R$  as the generic variance operator for the variance calculation with respect to the random mechanism mentioned above. Then  $v_i = r_i(r_i - 1)$  provides an unbiased estimator for  $V_i$ .

In situation (II) we shall be able to show for our illustrated applications of the “random” mechanisms that for real  $y_i$ , an unbiased estimator  $r_i$  will be derivable independently across  $i$  in  $s$  so that  $E_R(r_i) = y_i, \forall i \in U$ ; moreover, generically,  $V_R(r_i)$  will be shown to be of the form

$$V_i = V_R(r_i) = \alpha_i y_i^2 + \beta_i y_i + \Psi_i$$

with  $\alpha_i, \beta_i, \Psi_i$  as known constants and consequently,

$$v_i = \frac{1}{1 + \alpha_i} (\alpha_i r_i^2 + \beta_i r_i + \Psi_i)$$

is an unbiased estimator for  $V_R(r_i)$ . Then, it will follow that given  $t_b = t_b(d)$  with  $E_p(t) = Y$ , so that  $\sum_{s \ni i} p(s) b_{si} = 1$ ,

$$V_p(t_b) = \sum y_i^2 C_i + \sum_i \sum_{j, j \neq i} y_i y_j C_{ij}$$

and

$$v_p(t_b) = \sum_{i \in s} y_i^2 C_{si} + \sum_{i \in s} \sum_{j \in s, j \neq i} y_i y_j C_{sij},$$

one might employ, writing  $\underline{R} = (r_1, \dots, r_i, \dots, r_N)$

$$e_b = t_b \mid_{\underline{Y}=\underline{R}} = \sum r_i b_{si},$$

$$E(e_b) = E_p E_R(e_b) = E_R E_p(e_b) = Y,$$

$$V(e_b) = E_p V_R(e_b) + V_p E_R(e_b) = E_R V_p(e_b) + V_R E_p(e_b).$$

Then

$$v_1(e_b) = v_p(e_b) \mid_{\underline{Y}=\underline{R}} + \sum_{i \in s} v_i b_{si}$$

and

$$v_2(e_b) = v_p(e_b) \mid_{\underline{Y}=\underline{R}} + \sum_{i \in s} v_i (b_{si}^2 - C_{si})$$

will both satisfy

$$E(v_1(e_b)) = V(e_b) = E(v_2(e_b)),$$

where

$$E = E_p E_R = E_R E_p$$

and

$$V = E_p V_R + V_p E_R = E_R V_p + V_R E_p.$$

Thus, it is clear how in a general situation one may handle indirectly procured observations to estimate  $\theta$  and  $Y$  and also derive estimated measures of accuracy in estimation, in terms of coefficients of variation (CV). As a rule of thumb, we could consider an estimate as excellent if for an estimated CV, it is true that  $CV \leq 10\%$ , as satisfactory if  $10\% < CV \leq 20\%$ , as acceptable if  $20\% < CV \leq 30\%$  and as unacceptable if  $CV > 30\%$ .

## 2.3 How to Sample?

Any probability sampling scheme depending upon available resources may be employed. Chaudhuri's (2010) may be utilized as a companion to take care of this for simplicity. Chaudhuri's (2011) monograph may be utilized in gathering guidelines in their utilization in estimation.

## 2.4 How to Gather Sensitive Data?

We may recognize that facing a given situation, an investigator may consider an issue to be sensitive enough so that an indirect questioning technique may seem to be necessary. For this, requisite procedures are narrated in subsequent chapters. If the investigator deems it doubtful if he/she should go for a direct or indirect questioning, then also appropriate procedures may be followed as given in later chapters. Corresponding estimation procedures are also set forth in the right places in detail.

## References

- Basu, D. (1971). An essay on the logical foundations of survey sampling, Part 1. In V.P. Godambe, D.A. Sprott (Eds.), *Foundations of statistical inference* (pp. 203–242). Toronto: Holt, Rinehart & Winston.
- Chaudhuri, A. (2010). *Essentials of survey sampling*. New Delhi: Prentice Hall of India.
- Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys*. Boca Raton: Chapman & Hall, CRC Press, Taylor & Francis Group.

- Godambe, V.P. (1955). A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society: Series B*, 17, 269–278.
- Hajek, J. (1971). Comment on a paper by Basu. In V.P. Godambe, D.A. Sprott (Eds.), *Foundations of statistical inference* (pp. 203–242). Toronto: Holt, Rinehart and Winston.
- Hanurav, T.V. (1966). Some aspects of unified sampling theory. *Sankhya, Series A*, 28, 175–204.
- Hege, V.S. (1965). Sampling designs which admit uniformly minimum unbiased estimators. *Calcutta Statistical Association Bulletin*, 14, 160–162.
- Horvitz, D.G., & Thompson, D.J. (1952). A generalization of sampling without replacement from finite universe. *Journal of American Statistical Association*, 47, 663–685.
- Lanke, J. (1975). *Some contributions to the theory of survey sampling*. Ph. D. Thesis, University of Lund, Lund, Sweden.
- Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society: Series B*, 97, 199–214.
- Yates, F., & Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society: Series B*, 15, 253–261.

Indirect Questioning in Sample Surveys

Chaudhuri, A.; Christofides, T.

2013, XIII, 177 p. 1 illus., Hardcover

ISBN: 978-3-642-36275-0