Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/120220

Bayesian Estimation for Parameters of Truncated Data Based on a Two-Phase Sampling Plan

Sohair F. Higazi [1,*], *Sarah F. Aboud* [2], *and Khaled Bedair* [1]

[1]Department of Statistics, Mathematics, and Insurance, Faculty of Commerce, Tanta University, Egypt
[2]Department of Statistics, Mathematics, and Insurance, the Higher Institute of Computers, Information and Management Technology, Tanta, Egypt

Received: 22 Dec. 2022, Revised: 22 Jan. 2023, Accepted: 19 Feb. 2023
Published online: 1 May 2023

Abstract: This paper provides insights to researchers when conducting a survey and the response rate is very low, and it is of interest to increase response rate and to estimate the optimal population mean. The paper considers a population that consists of two strata, respondents, and non-respondents. The sampling design is to select a random sample of size n, then when only r observations respond to the survey, select an optimal random sample of size m that minimizes the expected loss, from the remaining (n-r) observations. A truncated Bayesian approach sampling plan is considered [1], such that the posterior distribution of the first stage is treated as a prior to the second stage, and an over-all mean is estimated. The paper illustrates Ericson approach to two random data sets with two sets of priors where the estimated overall mean is obtained for each stage and the expected loss is computed for the two prior sets. It is concluded that priors on means affect the optimal estimate for the mean; under the selected two priors, the final covariance matrix is approximately the same, and the losses are approximately equal when the r responses are more than 20%.

Keywords: Gamma Prior Distribution, Beta Normal Density, Optimal Second Stage Sample Size, Quadratic Loss function, First stage Variance-Covariance Matrix, Second stage Variance-Covariance Matrix.

1 Introduction

When conducting a survey, researchers decide on the sample size, using one of sample size determination formulas; and then send the survey either by mail, email or online to respondents. When the response rate is less than 100%, then Nonresponse error exists, that may lead to biased estimates [2]. Web-bases surveys are relatively easy and costless, but they suffer from low response rate that those conducted in the traditional way [3]. The size of non-response error is a function of the proportion of nonrespondents [4-6], and causes higher variability for the estimates, as the sample size you end up is smaller than the size you originally had determined; and thus the need for a second stage sample.

One of the approaches to solve the problem of non-response rate is the follow-up approach that involves contacting a random sample of non-respondents and having them complete the survey [7, 8]; or withdraw a second random sample of only nonrespondents. Used estimation methods for this scenario include ignorable maximum likelihood (IML) (e.g., [9]), Bayesian approach (i.e.,[10, 11]), and multiple imputation (e.g., [12]).

The objective of this paper is to estimate the mean of variable using responses from both samples: on how to determine the optimal second sample size that minimizes the losses and on how to combine all responses to reach the optimal estimate for the population mean. Bayesian algorithms that consider prior information for both samples [13]; prior information could be some results from pilot studies or previous research findings are recommended. The Bayesian approach is an extremely powerful approach for estimating or modeling uncertainty of a random variable especially when data is limited, and the

[*]Corresponding author e-mail: sohair_Higazi@commerce.tanta.edu.eg

© 2023 NSP
Natural Sciences Publishing Cor.

researcher is worried about overfitting [14] Bayesian selection models which utilize the additional data from phase II and develop an efficient Markov chain Monte Carlo algorithm for the posterior computation were introduced (i.e., [15, 16]).

This paper is an application of the Bayesian approach developed by Ericson [1] for the estimation of population mean from a truncated sequential two- stage sampling plan. Thus, the purpose of this paper is to apply Ericson`s approach using some prior information and show how is to estimate the population mean from a first stage sampling plan with r [r=0(10)150] responses, how to assign the second sample size from the remaining (n-r) observations, and how to incorporate prior information from both samples to reach an optimal estimate for the population mean. Section 2 , gives the sampling situation; Section 3 gives the theoretical frame work for the prior and the posterior distributions in phase I and II; section 4 gives the optimal second stage sample size and the optimal estimate for the mean ; and Section 5 gives discussions on computations results discussion; conclusions and recommendations are given in Section 6.

2 Sampling Situation

It is assumed that the population consists of two strata; in a sense that the first stratum has subjects, who will respond on the first contact, it has an unknown mean μ_1 and known variance σ_1^2; the second stratum is composed of subjects who did not response on the first trial with unknown mean μ_2 and known variance σ_2^2. The researcher selected a random sample of size n. The sampling model involves a truncated sequential sampling plan [1, 13] in a sense that the first sample has obtained r responses, with responses $x_1, x_2, \ldots x_r$ and conditional on the sample results, an optimal fixed size second sample of size m from the remaining (n-r) respondents is chosen. Thus, the choice of the optimal m will depend on the random outcome of the first sample; and m is a random quantity which is determined only when the outcome of the first sample is known [1].

Based on the researcher prior information which could be based on pilot studies that yield estimate for μ_1 , μ_2 , π , where π is an expected response rate. Applying Bayes approach, the joint distribution of μ_1 , μ_2 , π is obtained. The first posterior distribution is used to determine the optimal sampling of nonrespondents. The second sample is withdrawn yielding more information about μ_1 and μ_2 ;and the first posterior distribution is used as a prior distribution and with combination with the second sample data; a final distribution of μ_1 , μ_2 , π is obtained. A detailed discussion on the topic is given in [1, 17-19].

The over-all population mean to be estimates as:

$$\mu = \pi\mu_1 + (1-\pi)\mu_2 \tag{1}$$

Where π is an unknown proportion of respondents in the first stage, (1- π) is the unknown proportion of nonrespondents, and there is a joint prior distribution for (μ_1 , μ_2 , π).With π being independent of (μ_1, μ_2), and if π is as given in Equation (1),then using Lemma (1) in [1]:

$$\left.\begin{aligned} E(\mu) &= p\bar{\mu}_1 + (1-\pi)\bar{\mu}_2 \\ Var(\mu) &= h\left[\left(\bar{\mu}_1 - \bar{\mu}_2\right)^2 + v_{11} + v_{22} - 2v_{12}\right] + p^2 v_{11}\,(1-p)^2 v_{22} + 2p(1-p)v_{12} \\ p &= \frac{r}{n} \quad \text{and} \quad h = \frac{r(n-1)}{n^2\,(n+1)} \end{aligned}\right\} \tag{2}$$

A squared error loss function is considered, and Bayes terminal rule is the mean of the posterior distribution [1, 13].The optimal second phase sample is the number that minimizes the expected quadratic risk function, which takes the form:

$$R(n,r,m,y,\hat{u},\mu) = K((\mu-\hat{u})^2 + cn + c_1 r + c_2 m. \tag{3}$$

Where, $R(.)$is the loss plus cost; K is constant that gives the trade off between errors of estimation and sampling costs; and c, c_1 and c_2 are non-negative values represent basic sampling cost (survey design cost,…etc.), cost of one unit of the first phase, and cost of one unit of the first phase.

In this paper, the superscripts (0, 1, 2) to represent priors (0) for prior; (1) for posterior of stage 1; and (2) for posterior of stage 2.

© 2023 NSP
Natural Sciences Publishing Cor.

3 Theoretical Framework

In this section we give the prior and the posterior distributions for first and second stage samples.

3.1 The Prior Distribution

The prior distribution of π is given by a beta prior: Beta $(r^0, (n^0 - r^0)$, thus it takes the form [18]:

$$f_{\beta(\pi\,|n^0, r^0)} = \frac{\Gamma(n^0)}{\Gamma r^0 \Gamma(n^0 - r^0)} \pi^{n^0-1} (1-\pi)^{n^0-r^0-1} ,$$

where π is the unknown proportion of respondents. The joint prior on μ_1, μ_2 *and* π is a Beta-normal density with known parameters n^0 and r^0 [20], and

$$f_{(\mu_1, \mu_2, \pi)} = f^{(0)}{}_{\beta(\pi\,|n^0, r^0)} \times f^{(0)}{}_{N(\mu\,|\,\mu_1^{(0)}, \mu_2^{(0)}, v^{(0)})} \tag{4}$$

and $f_{N(\mu\,|\,\mu_1^{(0)}, \mu_2^{(0)}, v^{(0)})}$ is a bivariate normal density with:

$$\underline{\mu}^{(0)} = \begin{bmatrix} \mu_1^{(0)} \\ \mu_2^{(0)} \end{bmatrix} \quad \text{and} \quad V^{(0)} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}. \tag{5}$$

Conditioning on a given value of π [Equation (4)], the overall mean has a univariate normal prior density with mean:

$$g(\pi) = \pi\mu_1^0 + (1-\pi)\mu_2^0 = \pi(\mu_1^0 - \mu_2^0) + \mu_2^0$$

And variance

$$\sigma^2(\pi) = h^{-1}(\pi) = \pi^2 v_{11}^0 + (1-\pi)^2 v_{22}^0 + 2\pi(1-\pi)v_{12}^0$$

And the prior distribution of μ

$$\propto \int_0^1 \exp\left[-\frac{1}{2}\, h(\pi)((\mu - g(\pi)^2\right] \pi^{r^0-1}(1-\pi)^{n^0-r^0-1} \tag{6}$$

Thus and from (6), the prior distribution on r (number of respondents from first stage sampling, follow a binomial random variable ($Bin(r, \pi)$) and the responses $X`s$ follow a normal distribution $N(\mu, \frac{1}{h})$; thus, the mean of the $\overline{X}`s$ $is \sim N(\mu, \frac{1}{h_1 r})$, (When there are more than two strata, the prior on π follows a Dirichlet distribution [17].

3.2 Stage 1: Posterior

The likelihood function of π for the first stage sample is [1]:

$$L\left(\mu_1, \pi\, \middle|\, n, r, \overline{x}\right) \propto \pi^r (1-\pi)^{n-r} \exp\left[-\frac{1}{2} r h_1 (\mu_1 - \overline{X}_r)^2\right]$$

The joint posterior distribution for the first stage sample is:

$$f^{(1)}{}_{(\mu_1, \mu_2, \pi)} = f^{(0)}{}_{\beta(\pi\,|n^{(1)}, r^{(1)})} \times L\left(\mu_1, \pi\, \middle|\, n, r, \overline{x}\right). \tag{7}$$

Moreover, the distribution of (7) is a Beta-Normal density with parameters:

$$r^{(1)} = r^{(0)} + r \quad and \quad n^{(1)} = n^0 + n. \tag{8}$$

The first-stage estimators for μ_1 *and* μ_2 [18] are

© 2023 NSP
Natural Sciences Publishing Cor.

$$\left.\begin{aligned}\hat{\mu}_1{}^{(1)} &= \mu^0{}_1 + \frac{r\, v_{11}^0\, h_1(\bar{x} - \mu^0{}_1)}{d}\\ \hat{\mu}_2{}^{(1)} &= \mu^0{}_2 + \frac{r\, v_{22}^0\, h_2(\bar{y} - \mu^0{}_2)}{d}\end{aligned}\right\}, \tag{9}$$

where

$$d = 1 + rh_i\, v_{ii}^0 \qquad i = 1,2,$$

and the posterior Variance-Covariance matrix for stage 1 is:

$$V^{(1)} = \frac{1}{d}\begin{bmatrix} v_{11}^{(0)} & v_{12}^{(0)} \\ v_{21}^{(0)} & d\, v_{22}^{(0)} - rh_1\, {v_{12}^{(0)}}^2 \end{bmatrix}. \tag{10}$$

This yields the first-stage posterior correlation coefficient μ_1 , μ_2 as:

$$\rho^{(1)} = \frac{v_{12}^{(1)}}{\sqrt{v_{11}^{(1)}\, v_{22}^{(1)}}} = \frac{v_{12}^{(0)}}{\sqrt{v_{11}^{(0)}\, v_{22}^{(0)}}}\sqrt{\frac{v_{22}^{(0)}}{v_{22}^{(0)} + rh_1(\, v_{11}^{(0)}\, v_{22}^{(0)} - v_{12}^{(0)})^2}}. \tag{11}$$

3.3 Stage 2: Prior

For the second sample, with m observations, yielding a mean $\bar{y}$, the likelihood function [1] becomes:

$$L\left(\mu_2, \pi\ \middle|m, \bar{y}\right) \propto \quad \exp\left[-\frac{1}{2} mh_2\, (\bar{y} - \mu_2)^2,\right.$$

and the likelihood from both samples is:

$$\propto \pi^r (1-\pi)^{n-r} \exp\left[-\frac{1}{2}\, rh_1(\bar{x} - \mu_1)^2 - \frac{1}{2}\, mh_2\, (\bar{y} - \mu_2)^2.\right.$$

3.4 Stage 2: Posterior

The second stage posterior distribution on $(\mu_1 , \mu_2, \pi\,)$ denoted $f^{(2)}{}_{(\mu_1,\mu_2\,,\pi\,|n,r,\bar{x},m,\bar{y})}$ is proportional to:

$$\propto \quad f^{(1)}{}_{(\mu_1,\mu_2,\pi)} \times L\left(\mu_2, \pi\ \middle|m, \bar{y}\right).$$

Then, $f^{(2)}{}_{(\mu_1,\mu_2\,,\pi\,|n,r,\bar{x},m,\bar{y})}$ follows a bivariate normal density with parameters : $\mu_1^{(2)}$, $\mu_2^{(2)}$ and variance covariance matrix $V^{(2)}$, where

$$\left.\begin{aligned}\mu_1{}^{(2)} &= \frac{1}{e}\left[\left(1 + mh_2\, v_{22}^{(0)}\right)\left(\mu_1^{(0)} + rh_1\bar{x}\ \, v_{22}^{(0)}\right) + \ mh_2\, v_{12}^{(0)}\left(\bar{y} - \mu_2^{(0)} - rh_1\bar{x}\ \, v_{12}^{(0)}\right)\right]\\ \mu_2{}^{(2)} &= \frac{1}{e}\left[\left(1 + rh_1\, v_{11}^{(0)}\right)\left(\mu_2^{(0)} + mh_2\bar{y}\ \, v_{22}^{(0)}\right) + rh_1\ \ v_{12}^{(0)}\left(\bar{x} - \mu_1^{(0)} - mh_2\bar{y}\ \, v_{12}^{(0)}\right)\right]\\ e &= \left(1 + \ mh_2\, v_{22}^{(0)}\right)\left(1 + rh_1\, v_{11}^{(0)}\right) - \left(rmh_1h_2\, (v_{12}^{(0)})^2\right)\end{aligned}\right\}. \tag{12}$$

The posterior Variance-Covariance matrix for stage 2 is:

$$V^{(2)} = \frac{1}{e}\begin{bmatrix}\left(\, v_{11}^{(0)}(1 + mh_2\, (v_{22}^{(0)} - \ (v_{12}^{(0)})^2\right) & v_{12}^{(0)} \\ v_{21}^{(0)} & \left(\, v_{22}^{(0)}(1 + rh_1(\, v_{11}^{(0)} - \ \ (v_{21}^{(0)})^2\right)\end{bmatrix}. \tag{13}$$

3.5 Optimal Estimation of μ

The estimator $\hat{\mu}$ is chosen to minimize the expected risk [equation 2], thus:

© 2023 NSP
Natural Sciences Publishing Cor.

$$\hat{\mu} = min_{\mu} \left(k\left(\mu - \hat{\hat{\mu}}\right)^2\right) + cn + c_1 r + c_2 m$$

Where expectation is taken with respect to the second stage posterior distribution ofμ. Using Equation (2), Bayes terminal rule is found to be:

$$\hat{\hat{\mu}} = \frac{r + r^{(0)}}{n + n^{(0)}} \mu_1^{(2)} + \frac{n + n^{(0)} - r + r^{(0)}}{n + n^{(0)}} \mu_2^{(2)} \qquad (14)$$

3.6 Expected Loss

The risk function for the posterior distribution is as given in (3).

$$L(n, r, \overline{X}_r m, \overline{Y}_r) = k\, V_\mu^{(2)} + cn + c_1 r + c_2 m$$

Where $V_\mu^{(2)}$ is the final posterior variance of μ given in equation (13) and substituting from (2) we get the loss function in the second stage as follows:

$$\left.\begin{array}{c} L\left(n, r, \overline{X}_r, m, \overline{Y}_m\right) = KV_\mu{}^{(2)} + c + c_1 r + c_2 m \\ V_\mu^{(2)} = \{h[\left(\overline{\mu}_1{}^{(2)} - \overline{\mu}_2{}^{(2)}\right)^2 + \nu_{11}{}^{(2)} + \nu_{22}{}^{(2)} - 2\nu_{12}{}^{(2)}] + p^2 \nu_{11}{}^{(2)} (1 - p)^{2)} \nu_{22} + 2p(1-p)\nu_{12} \\ p = \dfrac{r^{(0)} + r}{n + n^{(0)}} \\ h = \dfrac{(r + r^{(0)})(n + n^{(0)} - r^{(0)} - r)}{(n + n^{(0)})^2 (n^{(0)} + n + 1)} \\ \nu_{ii} \text{ are elements of } V^{(2)} \end{array}\right\} \qquad (15)$$

4 Optimal Second stage sample size

The first stage produces $(n, r, \overline{x})$; the optimal number of the remaining $(n - r)$ nonrespondents to be sampled at stage 2, is the number that minimizing the risk function. The expected loss is given by

$$L(n, r, \overline{x}, m) \equiv E_{\overline{y}_m}\left(n, r, \overline{x}, m, \overline{y}_m\right).$$

The optimal m is the value that minimizes the expected loss [Eq. 3]. The loss function depends on $(n, r, \overline{x}, m, \overline{y})$, and the risk depends on the loss function and the cost; so, both the risk and the loss functions depend on y which depends on m (optimal size to be samples in stage 2). Theorem (1) in [1] gives the optimal proportion of the $(n - r)$ non-respondents to be sampled at second stage as follows:

$$\phi = \begin{bmatrix} \dfrac{h_2 W - 1}{\upsilon_{22}^{(1)} (n - r) h_2} & n > r \\ 0 & n = r \end{bmatrix}. \qquad (16)$$

Where,

$$W = \left(\frac{K}{c_2 h_2}\right)^{1/2} [P^{(1)} \upsilon_{12}^{(1)} + (1 - P^{(1)}) \upsilon_{22}^{(1)}, \qquad (17)$$

where, c_2 is the cost of sampling one observation in stage 2, and $h_2 = \frac{1}{\sigma_2^2}$.

The optimal second stage sample size is:

$$m = \begin{cases} 0 & \phi = 0 \\ \phi(n - r) & 0 < \phi < 1. \\ n - r & \phi > 1 \end{cases} \qquad (18)$$

© 2023 NSP
Natural Sciences Publishing Cor.

5 Analytic Results

As [1] has mentioned "While this class of priors and posterior distributions seems analytically in-tractable in several respects, lack of simple expressions need not present any practical difficulty given today's computing hardware". A customized computational algorithm was developed to apply the proposed approach. The following prior information are used:

$$\mu^{(0)} = \begin{bmatrix} 40 \\ 45 \end{bmatrix}, \quad V^{(0)} = \begin{bmatrix} 12 & 4 \\ 4 & 9 \end{bmatrix}, \quad n = 150 \quad n^0 = 10 \quad r^0 = 2$$

$$\sigma_1^2 = \sigma_2^2 = 16 \quad c = 1500 \quad c_1 = 2 \quad c_2 = 5$$

r random numbers are generated $(0 \le X_i \le 100)$ to represent the responses of the first sample : $x_1, x_2, \ldots, x_r$, and $0 \le r \le n$; i.e., from 0% to 100% responses. Sample means $\overline{X}_r$ are computed from the generated data. Table (1) gives the classical estimation (one stage sampling) for the population point estimates for proportion and mean for $r = 0(10)150$. Proportion of respondents P_r are estimated in this initial stage as $\frac{r}{n}$. The mean and variance of $\overline{X}_r$ and proportions of respondents are computed, and 90% confidence intervals are constructed (not considering any priors).

Table 1: Classical Estimation of Proportions and Means for responses r=0(10)150

r	Proportions				Means			
	P_r	SE	LCL	UCL	$\overline{X}_r$	SE	LCL	UCL
0	0.00	0.00	0.00	0.00	0.00	0	0.000	0.00
10	0.07	0.02	0.03	0.11	50.46	1.265	47.976	52.93
20	0.13	0.03	0.08	0.19	48.94	0.894	47.187	50.69
30	0.20	0.03	0.14	0.26	56.94	0.730	55.509	58.37
40	0.27	0.04	0.20	0.34	48.39	0.632	47.150	49.63
50	0.33	0.04	0.26	0.41	53.75	0.566	52.641	54.86
60	0.40	0.04	0.32	0.48	50.25	0.516	49.238	51.26
70	0.47	0.04	0.39	0.55	52.64	0.478	51.703	53.58
80	0.53	0.04	0.45	0.61	47.85	0.447	46.973	48.73
90	0.60	0.04	0.52	0.68	53.40	0.422	52.574	54.23
100	0.67	0.04	0.59	0.74	50.82	0.400	50.036	51.60
110	0.73	0.04	0.66	0.80	52.65	0.381	51.902	53.40
120	0.80	0.03	0.74	0.86	53.39	0.365	52.674	54.11
130	0.87	0.03	0.81	0.92	52.76	0.351	52.072	53.45
140	0.93	0.02	0.89	0.97	53.11	0.338	52.447	53.77
150	1.00	0.00	1.00	1.00	50.20	0.327	49.560	50.84

5.1 Estimate for population mean for stage one

Using the obtained means $(\overline{x}_r)$ as given in Table (1); and applying equations (8)-(11), Table (2) is obtained for r= 0(10)150. Table (2) shows that the estimated population mean is less than sample means $\overline{x}_r$, except at r=0 and r=n; the correlation coefficient $\rho^{(1)}$ at r=0 is the same as $\rho^{(0)}$, computed using the prior covariance matrix; the posterior estimated population proportion of respondents for π and also thus $\hat{\sigma}_\pi^2$ depend mainly on the prior information n^0, r^0 and n; and they are not affected by changes of other constants. Also, the larger the number of respondents (r) the smaller posterior correlation between ($\hat{\mu}_1$ and $\hat{\mu}_2$); and there is a variance-covariance matrix for each r, except at r=0, where its variance covariance matrix $=V^{(0)}$, for example, when r=60

$$p = \frac{6+20}{150+10} = .3875\,; \quad \mu^{(1)} = 48.19 \quad and \quad V_{r=60}^{(1)} = \begin{bmatrix} .26 & .09 \\ .09 & 7.64 \end{bmatrix} \quad \rho = .0614.$$

© 2023 NSP
Natural Sciences Publishing Cor.

Table 2: Variance-Covariance Matrices and Estimates of $\hat{\mu}^{(1)}$ for r= 0(10)150

r	p	$\rho^{(1)}$	$v_{11}^{(1)}$	$v_{12}^{(1)}$	$v_{22}^{(1)}$	$\overline{X}_r$	$\mu_1^{(1)}$	$\mu_2^{(1)}$	$\hat{\mu}^{(1)}$
0	0.0125	0.3849	12.00	4.00	9.00	0.00	40.000	45.00	44.94
10	0.0750	0.1416	1.41	0.47	7.78	50.46	49.594	46.67	46.89
20	0.1375	0.1037	0.75	0.25	7.70	48.94	48.717	46.28	46.62
30	0.2000	0.0857	0.51	0.17	7.67	56.94	56.868	48.96	50.54
40	0.2625	0.0747	0.39	0.13	7.66	48.39	48.444	46.14	46.74
50	0.3250	0.0671	0.31	0.10	7.65	53.75	53.929	47.95	49.90
60	0.3875	0.0614	0.26	0.09	7.64	50.25	50.428	46.78	48.19
70	0.4500	0.0569	0.22	0.07	7.64	52.64	52.900	47.60	49.98
80	0.5125	0.0533	0.20	0.07	7.64	47.85	48.030	45.97	47.03
90	0.5750	0.0503	0.18	0.06	7.63	53.40	53.733	47.87	51.24
100	0.6375	0.0478	0.16	0.05	7.63	50.82	51.105	46.99	49.61
110	0.7000	0.0456	0.14	0.05	7.63	52.65	52.998	47.62	51.38
120	0.7625	0.0437	0.13	0.04	7.63	53.39	53.773	47.88	52.37
130	0.8250	0.0420	0.12	0.04	7.63	52.76	53.136	47.66	52.18
140	0.8875	0.0405	0.11	0.04	7.63	53.11	53.506	47.78	52.86
150	0.9500	0.0391	0.11	0.04	7.63	50.20	50.515	46.79	50.33

To determine the size of the optimal second stage sample size, m, Equations (16) - (18) are applied. The value of m depends on: a) the value of k in the risk function (Equation (3), b) the variance of the population of non-respondents, c) the percentage of respondents in stage (1) $p^{(1)}$, and d) the initial variance covariance matrix $V^{(0)}$. Table (3) give the effect of k on ϕ, ρ and m for $r = 0(10)150$.

Examination of Table (3) reveals the following:

1. The second stage sample size, m, is very sensitive to the quantity $(kh_2/ c_2$; the smaller this quantity the smaller ϕ; this quantity affects W (equation (17)); which in turns affects ϕ. Thus, the larger the variance of non-respondents, the smaller ϕ (given c_2 is constant); and the higher c_2 the smaller ϕ and the smaller m (given kh_2 is constant).
2. The larger r the smaller the m
3. ϕ Is sensitive to, r,as r increases the percentage assigned for non-respondents increase, for fixed W.
4. As r increases, the value of W decreases (all other elements in equation (17) are held constants)
5. $When$ $r = n$ ($m = 0$); no second stage sampling is needed , then Equations in (12) depend on priors of means and the prior covariance matrix and the variance of the population of respondents, and equations (12) are reduced to:

$$\mu_1{}^{(2)} = \frac{1}{e}\left[\mu_1^0 + rh_1\overline{x}\ \ \nu_{12}^0\right]$$

$$\mu_2{}^{(2)} = \frac{1}{e}\left[\mu_2^0 + (1 + \ rh_1\ \ \nu_{11}^0) + rh_1\,\nu_{12}^0(\overline{x} - \mu_1^0\right]$$

and,

$$e = \ (1 + h_1\ \ \nu_{11}^0).$$

© 2023 NSP
Natural Sciences Publishing Cor.

Table 3: The effect of K on W, ϕ, ρ and m for $r = 0(10)150$.

r	$P^{(1)}$	$\rho^{(1)}$	k=50			K=500			K=1000		
			W	ϕ	m	W	ϕ	m	W	ϕ	m
0	0.0125	0.3849	112.58	0.07154	11	356	0.2519	38	503.46	0.3611	54
10	0.0750	0.1416	106.25	0.07163	10	336	0.2540	36	475.18	0.3644	51
20	0.1375	0.1037	99.93	0.07173	9	316	0.2564	33	446.89	0.3683	48
30	0.2000	0.0857	93.60	0.07186	9	296	0.2593	31	418.61	0.3728	45
40	0.2625	0.0747	87.28	0.07200	8	276	0.2626	29	390.32	0.3781	42
50	0.3250	0.0671	80.95	0.07217	7	256	0.2667	27	362.04	0.3845	38
60	0.3875	0.0614	74.63	0.07238	7	236	0.2716	24	333.75	0.3923	35
70	0.4500	0.0569	68.31	0.07265	6	216	0.2778	22	305.47	0.4020	32
80	0.5125	0.0533	61.98	0.07299	5	196	0.2857	20	277.19	0.4146	29
90	0.5750	0.0503	55.66	0.07344	4	176	0.2963	18	248.90	0.4313	26
100	0.6375	0.0478	49.33	0.07407	4	156	0.3111	16	220.62	0.4547	23
110	0.7000	0.0456	43.01	0.07502	3	136	0.3333	13	192.33	0.4898	20
120	0.7625	0.0437	36.68	0.07660	2	116	0.3704	11	164.05	0.5483	16
130	0.8250	0.0420	30.36	0.07977	2	96	0.4444	9	135.76	0.6654	13
140	0.8875	0.0405	24.03	0.08926	1	76	0.6667	7	107.48	1.0164	10
150	0.9500	0.0391	17.71	0.00000	0	56	0.0000	0	79.20	0.0000	0

5.2 Estimate for population mean for stage two

When $m > 0$, m random numbers are generated to represent the responses of the second stage sampling: $y_1, y_2, \dots, y_m$, ; the second sample mean $\overline{y}$ is computed from the generated data; then applying the Equations in (12) to estimate μ using the prior information, and first posterior estimate for the mean and $y_1, y_2, \dots, y_m$ second stage responses.

Table (4), gives the posterior Variance-Covariance matrix and the optimal estimates of population mean at phase 2; and since the two means are independent, the estimated variance at second stage, computed using Equation (14), would be as follows:

$$V(\hat{\hat{\mu}}) = \left(\frac{r+r^{(0)}}{n+n^{(0)}}\right)^2 v_{11}^{(2)} + \left(\frac{n+n^{(0)}-r+r^{(0)}}{n+n^{(0)}}\right)^2 v_{22}^{(2)}.$$

Also, the standard error of estimate computed as $\sqrt{\frac{Var(\hat{\mu})^{(2)}}{r+m}}$ and 95% confidence interval for the optimal population mean, for r=0(10)150 is shown in Table (4).

In Table (4); comparison of estimated means $\overline{X}_r$, $\hat{\mu}^{(1)}$ and $\hat{\mu}^{(2)}$, it is found that values of $\overline{X}_r$ (except for r=0) are higher estimates for the population mean, followed by $\hat{\mu}^{(1)}$ and $\hat{\mu}^{(2)}$ is the smallest. Confidence intervals produced are tighter and limits are less than those in Table (1). For r=40, m= 32, $\overline{X} = 48.39, \overline{Y} = 49.07$, incorporating prior information reveals that the estimated population mean for the first sample $\mu_1^{(2)}$= 42.84 and for the second sample $\mu_2^{(2)}$ =43.47 ; the estimated over-all population mean from the two samples= 37.87 with variance =.37. The correlation coefficient between $(\mu_1^{(2)}, \mu_2^{(2)})$= .02 which is very weak association between the two means, this is due to the very small covariance between the two variables.

© 2023 NSP
Natural Sciences Publishing Cor.

Table 4: Second Stage sampling results for r=0 (10)150.

r	m	$\overline{X}_r$	$\overline{Y}$	$v_{11}^{(2)}$	$v_{12}^{(2)}$	$v_{22}^{(2)}$	$\mu_1^{(2)}$	$\mu_2^{(2)}$	$\hat{\mu}^{(1)}$	$\hat{\mu}^{(2)}$	$Var(\hat{\mu}^{(2)})$	Se	LCL	UCL
0	38	0.00	51.77	0.52	0.17	0.39	61.50	51.26	44.94	44.98	0.40	0.10	44.78	45.18
10	36	50.46	52.70	1.53	0.03	0.47	45.37	46.27	46.89	40.42	0.43	0.10	40.22	40.61
20	33	48.94	53.85	0.82	0.02	0.51	43.67	47.26	46.62	40.86	0.41	0.09	40.69	41.03
30	31	56.94	53.75	0.56	0.01	0.54	50.15	47.38	50.54	42.01	0.39	0.08	41.85	42.17
40	29	48.39	49.07	0.42	0.01	0.58	42.84	43.47	46.74	37.87	0.37	0.07	37.73	38.01
50	27	53.75	53.29	0.34	0.01	0.63	47.44	47.07	49.90	41.31	0.34	0.07	41.17	41.44
60	24	50.25	51.04	0.29	0.01	0.68	44.43	45.25	48.19	39.28	0.32	0.06	39.16	39.40
70	22	52.64	54.35	0.25	0.01	0.74	46.55	48.04	49.98	41.36	0.30	0.06	41.25	41.47
80	20	47.85	44.14	0.22	0.01	0.82	42.38	39.93	47.03	36.19	0.27	0.05	36.09	36.30
90	18	53.40	50.24	0.19	0.01	0.91	47.31	45.05	51.24	40.72	0.25	0.05	40.62	40.81
100	16	50.82	64.82	0.17	0.01	1.03	45.23	56.36	49.61	42.22	0.23	0.04	42.13	42.31
110	13	52.65	53.07	0.16	0.01	1.18	46.89	47.58	51.38	41.15	0.20	0.04	41.07	41.23
120	11	53.39	58.83	0.15	0.01	1.37	47.75	52.11	52.37	42.27	0.18	0.04	42.20	42.34
130	9	52.76	43.91	0.14	0.01	1.66	47.38	41.71	52.18	41.18	0.16	0.03	41.11	41.24
140	7	53.11	64.50	0.13	0.01	2.08	48.16	56.40	52.86	42.04	0.14	0.03	41.98	42.10
150	0	50.20	0.00	0.14	0.04	9.00	51.60	56.82	50.33	44.76	0.17	0.03	44.69	44.83

Table 5: Estimation of m, and Second Stage sampling results for r=0(10)150 and k=500.

r	W	φ	m	$\overline{X}$	$\overline{Y}$	$v_{12}^{(2)}$	$v_{11}^{(2)}$	$v_{22}^{(2)}$	$\mu_1^{(2)}$	$\mu_2^{(2)}$	$\hat{\mu}^{(2)}$	$Var(\hat{\mu}^{(2)})$	LCL	UCL
0	442.18	0.28	42	0.00	46.36	0.23	13.61	0.45	58.49	46.98	41.25	0.47	41.04	41.45
10	417.03	0.28	40	50.46	47.26	0.03	1.74	0.48	51.54	47.82	42.12	0.44	41.94	42.30
20	391.87	0.29	37	48.94	43.32	0.02	0.93	0.51	49.50	44.12	39.34	0.42	39.17	39.51
30	366.72	0.29	35	56.94	56.42	0.01	0.64	0.54	57.24	56.59	49.65	0.39	49.49	49.80
40	341.56	0.29	32	48.39	55.21	0.01	0.48	0.58	48.90	55.30	46.70	0.37	46.56	46.85
50	316.40	0.30	30	53.75	53.55	0.01	0.39	0.62	53.98	53.89	47.18	0.35	47.05	47.31
60	291.25	0.30	27	50.25	49.61	0.01	0.33	0.68	50.49	50.20	44.04	0.32	43.92	44.16
70	266.09	0.31	25	52.64	52.46	0.01	0.28	0.74	52.82	52.92	46.26	0.30	46.15	46.37
80	240.94	0.32	22	47.85	51.17	0.01	0.25	0.82	48.08	51.66	43.37	0.28	43.27	43.47
90	215.78	0.33	20	53.40	46.10	0.01	0.22	0.91	53.48	47.38	44.97	0.26	44.87	45.06
100	190.62	0.34	17	50.82	46.50	0.01	0.20	1.03	50.94	47.80	43.82	0.23	43.74	43.91
110	165.47	0.36	15	52.65	49.31	0.01	0.18	1.18	52.75	50.49	45.76	0.21	45.68	45.84
120	140.31	0.40	12	53.39	46.85	0.01	0.16	1.39	53.46	48.67	46.24	0.19	46.17	46.31
130	115.16	0.48	10	52.76	49.00	0.01	0.15	1.68	52.84	50.74	46.13	0.17	46.06	46.20
140	90.00	0.70	7	53.11	37.50	0.01	0.14	2.14	53.14	42.65	46.63	0.15	46.57	46.69
150	64.85	0.00	0	50.20	0.00	0.04	0.13	8.45	50.32	56.96	43.53	0.17	43.47	43.60

© 2023 NSP
Natural Sciences Publishing Cor.

5.3 Changing Priors: Priors (2)

To study the effect of priors on the final estimation process, new hypothetical priors (priors 2) are selected, as:

$$V = \begin{bmatrix} 16 & 5 \\ 5 & 10 \end{bmatrix}, \quad \sigma_1^2 = \sigma_2^2 = 20 \quad \mu_1^{(0)} = 65 \ , \mu_2^{(0)} = 60,$$

$$n = 150, \quad r^0 = 2 \quad n^0 = 10$$

Table (5) was constructed using the priors above, and K=500 for the estimation of m, equations (16-18); the second phase samples results are as shown in Table (5).

Examination of Table (5) shows that for r=40, m= 32,$\overline{X} = 48.39, \overline{Y} = 55.21$, incorporating prior information reveals that the estimated population mean for the first sample $\mu_1^{(2)}$= 48.90 and for the second sample $\mu_2^{(2)}$ =55.30 which is close to ,$\overline{Y}$; the estimated over-all population mean from the two samples= 46.70 with variance =.37. The initial correlation coefficient between ($\mu_1^{(0)}, \mu_2^{(0)}$)= .3952 while in phase 2, the correlation coefficient between ($\mu_1^{(2)}, \mu_2^{(2)}$=.0189 [Equation (11)]; so the correlation got weaker in phase 2; Table (5) also shows a very small standard error and thus the confidence limits are very narrow, due to the $V^{(0)}$ matrix elements.

Compares estimates obtained using prior (1) [Table (4)] and Prior (2) [Table (5)] at r=40, m= 29 , $\overline{X} = 48.39, \overline{Y} = 55.21$, incorporating prior information reveals that priors (1) yields lower estimates for population 1 and population 2 , and the optimal estimated mean. The correlation coefficient between ($\mu_{1,r=40}^{(2)}, \mu_{2,r=40}^{(2)}$)= .0189, which is approximately the same under both priors, and thus, the same coefficient is obtained for prior (1).

Table 6: the Expected Risk for {r and m), K=500.

r	Losses for Prior (1)				Expected Loss for Prior (2)			
	m	*p*	*h*	Risk	*m*	*p*	*h*	Risk
0	38	0.013	0.000	2167	42	0.013	0.000	8747
10	36	0.075	0.000	2875	40	0.075	0.000	3012
20	33	0.138	0.001	2554	37	0.138	0.001	2638
30	31	0.200	0.001	2443	35	0.200	0.001	2496
40	29	0.263	0.001	2392	32	0.263	0.001	2464
50	27	0.325	0.001	2372	30	0.325	0.001	2411
60	24	0.388	0.002	2369	27	0.388	0.002	2403
70	22	0.450	0.002	2380	25	0.450	0.002	2408
80	20	0.513	0.002	2403	22	0.513	0.002	2437
90	18	0.575	0.002	2434	20	0.575	0.002	2488
100	16	0.638	0.002	2580	17	0.638	0.002	2509
110	13	0.700	0.002	2540	15	0.700	0.002	2569
120	11	0.763	0.001	2645	12	0.763	0.001	2673
130	9	0.825	0.001	2784	10	0.825	0.001	2797
140	7	0.888	0.001	3004	7	0.888	0.001	3066
150	0	0.950	0.001	6409	0	0.950	0.001	6137

5.4 Expected Risk of phase two priors

Applying Equation (15), and variance-covariance estimates from Tables (5), and for K=500, C=1500, $c_1 = 3$, $c_2 = 5$. The expected losses are shown for each combination of [11] as shown in Table (6)}, where p and h as given in equations (15).

Comparing the final estimated mean (phase 2), using the two different priors for the mean of the two populations [Table (7)], reveals that, although the prior means differ in magnitude for the second prior, the final estimated mean for the second prior tends to be closer to in magnitude to the final estimated mean for the first prior .

© 2023 NSP
Natural Sciences Publishing Cor.

Table (7) shows that the $|\hat{\mu}_1^{(2)} - \hat{\mu}_2^{(2)}| >= 5$ and $\hat{\mu}_1^{(2)}$ *and* $\hat{\mu}_2^{(2)}$ are the optimal means for prior (1) and (2) respectively. Table (8) compares the effect of the different prior Covariance matrix on the second stage covariance matrix. Examination of the table shows that the values of $\upsilon_{12}^{(2)}$ are very close when ≥ 30 *for* $\forall\, m$; and thus, the prior covariance matrix has no effect on the posterior covariance matrix, and the correlation coefficient between ($\mu_1^{(2)}, \mu_2^{(2)}$) reflects very weak association.

Table 7: Effect of priors on the second stage estimated optimal mean, for r=0(10)150.

r	Classical Mean $\hat{\mu} = \bar{X}$	$\mu_1^{(0)} = 40$, $\mu_2^{(0)} = 45$		$\mu_1^{(0)} = 65$, $\mu_2^{(0)} = 60$	
		$\bar{Y}$	$\hat{\mu}_{11}^{(2)}$	$\bar{Y}$	$\hat{\mu}_{12}^{(2)}$
0	0.00	51.77	44.98	46.36	41.25
10	50.46	52.70	40.42	47.26	42.12
20	48.94	53.85	40.86	43.32	39.34
30	56.94	53.75	42.01	56.42	49.65
40	48.39	49.07	37.87	55.21	46.70
50	53.75	53.29	41.31	53.55	47.18
60	50.25	51.04	39.28	49.61	44.04
70	52.64	54.35	41.36	52.46	46.26
80	47.85	44.14	36.19	51.17	43.37
90	53.40	50.24	40.72	46.10	44.97
100	50.82	64.82	42.22	46.50	43.82
110	52.65	53.07	41.15	49.31	45.76
120	53.39	58.83	42.27	46.85	46.24
130	52.76	43.91	41.18	49.00	46.13
140	53.11	64.50	42.04	37.50	46.63
150	50.20	0.00	44.76	0.00	43.53

Table 8: The effect of different prior Covariance Matrix, for r=0(10)150 and accompanying m.

r	V= $\begin{bmatrix} 12 & 4 \\ 4 & 9 \end{bmatrix}$, $\sigma_1^2 = \sigma_2^2 = 16$				V= $\begin{bmatrix} 16 & 5 \\ 5 & 10 \end{bmatrix}$, $\sigma_1^2 = \sigma_2^2 = 20$			
	m	$\upsilon_{12}^{(2)}$	$\upsilon_{11}^{(2)}$	$\upsilon_{22}^{(2)}$	m	$\upsilon_{12}^{(2)}$	$\upsilon_{11}^{(2)}$	$\upsilon_{22}^{(2)}$
0	38	0.17	0.52	0.39	42	0.23	13.61	0.45
10	36	0.03	1.53	0.47	40	0.03	1.74	0.48
20	33	0.02	0.82	0.51	37	0.02	0.93	0.51
30	31	0.01	0.56	0.54	35	0.01	0.64	0.54
40	29	0.01	0.42	0.58	32	0.01	0.48	0.58
50	27	0.01	0.34	0.63	30	0.01	0.39	0.62
60	24	0.01	0.29	0.68	27	0.01	0.33	0.68
70	22	0.01	0.25	0.74	25	0.01	0.28	0.74
80	20	0.01	0.22	0.82	22	0.01	0.25	0.82
90	18	0.01	0.19	0.91	20	0.01	0.22	0.91
100	16	0.01	0.17	1.03	17	0.01	0.20	1.03
110	13	0.01	0.16	1.18	15	0.01	0.18	1.18
120	11	0.01	0.15	1.37	12	0.01	0.16	1.39
130	9	0.01	0.14	1.66	10	0.01	0.15	1.68
140	7	0.01	0.13	2.08	7	0.01	0.14	2.14

© 2023 NSP
Natural Sciences Publishing Cor.

6 Conclusions and Recommendations

6.1 Conclusions

This paper presents computations of Ericson`s formulas to estimate the mean of a two-stage sampling plan, using hypothetical priors. The following conclusions and recommendations are reached as follows:

1. The estimated optimal mean in stage 1 and stage 2 is less than the observed mean for ∀ values of r except for r=0 and r=n
2. The estimated optimal means in phases 1 and 2 are affected by the prior means, the optimal means get higher as prior means get higher.
3. The percentage of observations to be selected in phase 2 (m), increases as the (n-r) decreases and ϕ is sensitive to, r,as r increases the percentage assigned for non-respondents increase, for fixed W.
4. The larger r the smaller the m
5. K in the risk function affects the m, s K increases m increases, m also increases.
6. The variance of the optimal mean decreases as r increases and the confides limits get tighter.
7. The expected risk is the highest at r=0 and r=n, and it is an increasing function of r.
8. The covariance matrix is not sensitive to the prior covariance matrix.
9. The correlation coefficient between the two mean gets weaker as the number of responses get larger.
10. $When\, r = n$ ($m = 0$); no second stage sampling is needed.

6.2 Recommendations

It is recommended to investigate such approach when the population consists of three strata with unknown variances with large sample sizes, and with small sample sizes. In addition, it is recommended to construct tables for standard priors that researchers can be referred to for the second sample size given the number of responses in phase 1.

References

[1] Ericson, W.A., Optimal sample design with nonresponse. *Journal of the American Statistical Association*. **62**, 63-78 (1967).
[2] Rogelberg, S.G., et al., Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology*. **88**, 1104 (2003).
[3] Simsek, Z. and J.F. Veiga, A primer on internet organizational surveys. *Organizational research methods*. **4**, 218-235 (2001).
[4] Lavrakas, P.J., *Encyclopedia of survey research methods*: Sage publications, (2008).
[5] Reio Jr, T.G., Survey Nonresponse Bias in Social Science Research. *New Horizons in Adult Education and Human Resource Development*. **21**, 48-51 (2007).
[6] Gray, P.E., R.E. Machol, and R.E. Machol, Recent developments in information and decision processes. (1962).
[7] Rogelberg, S.G. and A. Luong, Nonresponse to mailed surveys: A review and guide. *Current directions in psychological science*. **7**, 60-65 (1998).
[8] Stoop, I.A., *The hunt for the last respondent: Nonresponse in sample surveys*: Sociaal en Cultureel Planbu, (2005).
[9] Dempster, A.P., N.M. Laird, and D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: series B*. **39**, 1-22 (1977).
[10] Sugden, R. and T. Smith, Ignorable and informative designs in survey sampling inference. *Biometrika*. **71**, 495-506 (1984).
[11] Bolstad, W.M. and J.M. Curran, *Introduction to Bayesian statistics*: John Wiley & Sons, (2016).
[12] Rubin, D.B., *Multiple imputation for nonresponse in surveys*: John Wiley & Sons, (2004).

© 2023 NSP
Natural Sciences Publishing Cor.

[13] Ferguson, T., Mathematical Statistics: A Decision Theoretic Approach. *Academic Press, New York*. (1967).

[14] Holland, P.W., *Introduction to Bayesian inference and decision*: Taylor & Francis, (1973).

[15] Zhang, N., H. Chen, and M.R. Elliott, Nonrespondent subsample multiple imputation in two-phase sampling for nonresponse. *Journal of Official Statistics*. **32**, 769-785 (2016).

[16] Zhang, Y., H. Chen, and N. Zhang, Bayesian inference for nonresponse two-phase sampling. *Statistica Sinica*. **28**, 2167-2187 (2018).

[17] Draper, N.R. and I. Guttman, Some Bayesian stratified two-phase sampling results. *Biometrika*. **55**, 131-139 (1968).

[18] Raiffa, H. and R. Schlaifer, Applied statistical decision theory. (2000).

[19] Schlaifer, R.O., *Probability and statistics for business decisions: an introduction to managerial economics under uncertainty*, (1959).

[20] Novick, M.R. and P.H. Jackson, *Statistical methods for educational and psychological research*: McGraw-Hill, (1974).

© 2023 NSP
Natural Sciences Publishing Cor.