ON USE OF PRELIMINARY TESTS OF SIGNIFICANCE IN REPEATED SURVEYS

By

B.V.S. Sisodia,

IVRI, Izatnagar

(Received: November 1979)

Summary

A preliminary test estimator for the population mean on the current occasion in case of sampling over two occasions is built up which depends on the outcome of the preliminary test. Both the cases are considered when variance-covariance of the variables on both the occasion is known and unknown. In both the cases, the preliminary test estimators are found to be better than usual estimators for large value of ρ, depending upon the proper choice of α and q.

Introduction

In repeated surveys the application of successive sampling technique: with partial replacement of sampling unit on the subsequent occasious, has many advantages such as reduction in cost, time, labour; precision, comparability etc. For estimating the population mean of a character on the second occasion in repeated surveys, the information available on the previous occasion may be used to get an improved estimator of the population mean through application of double sampling techniques.

Han (1973) developed an estimation procedure for double sampling based on a prior information on the axiliary character. He assumes that, if the population mean of the auxiliary character, *i.e.*, μ_x is approximately known (say, μ_o), then preliminary sample may be used to test the hypothesis; $\mu_x=\mu_o$. He proposed an estimator based on this preliminary test. In case of samaling on two successive occasions, there may be situations when, in addition to the information collected on the first occasion, an experimenter has got approximate information about the population mean. This information may be available on the basis of some other auxiliary character or it may be some intelligent guess obtained from the experience gained in due course of repeated surveys. On the basis of experience, the experimenter believes that the population mean on the first occasion is very close to that of its approximate priori knowledge. However, his belief is only subjective and this may be tested on the basis of sample available on the first occasion. This provides a preliminary test to the belief and if it is found correct with reasonably a good confidence, this may be used in the estimation.

In the present paper an estimation procedure which involves a preliminary test of significance in the case of sampling over two occasions is proposed. Its bias and mean-square-error (MSE) are derived. The relative efficiency (RE) of the proposed estimator as compared to the usual estimator is discussed. Both the cases are considered when variance-covariance matrix of the variables on the two occasions are known and unknown.

When variance-covariance matrix is known

Let (X, Y) have bivariate normal distribution with mean (μ_x, μ_y) and variance-covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\,\sigma_x\,\sigma_y \\ \rho\,\sigma_x\,\sigma_y & \sigma_y^2 \end{bmatrix}$$

The bivariate normal distribution is assumed for simplicity but it would be useful to investigate the behaviour of this preliminary test scheme when data are, not selected from a bivariate normal population, but are in fact, from some other population, such as bivariate gama. The bias and MSE as reported here would not be correct for non-normal data, but they may provide reasonably approximation to the bias and MSE if the procedure is robust. In fact in case of large scale surveys where the sample size is sufficiently large, it is not essential to assume the normality of original bivariate population.

The sampling procedure and minimum variance linear unbiased estimation (MVLUE)

Suppose that X and Y, the two random variables, denote the same character on the first and second occasions respectively. Consider a random sample $(x_1, x_2, .., x_n)$ drawn from the population on the first occasion. A fraction np ($np=m$, say) of the units drawn on the first occasion is retained on the second occasion and is supplemented by nq ($nq=u$, say) units drawn afresh from the population on the second occasion such that sample size remains constant over occcasions, *i.e*, $p+q=1$, Let $\bar{x}$ and $\bar{y}$ be the sample means on the first and second occasions respectively, $\bar{x}'$ and $\bar{y}'$ be the means based on the units common to both occasions and, $\bar{x}''$ and $\bar{y}''$ be the means based on the units uncommon to both occasions.

To begin the investigation of the estimation procedure, the elements of Σ (*i.e.*, ρ, σ_x^2 and σ_y^2) are assumed to be known. Supposing that $\sigma_x^2 = \sigma_y^2 = \sigma^2$, it can be further assumed, without loss of generality that $\sigma^2=1$. Thus, the usual estimator of the population mean μ_y on the second occasion is given by

$$\hat{Y}_2=c\,[\bar{y}'+\rho\,(\bar{x}-\bar{x}')]+(1-c)\,\bar{y}'' \qquad \ldots(2.1.1)$$

where c is an arbitrary constant. An optimum value of c obtained by minimising the variance of $\hat{Y}_2$ is given by

$$c_o=\frac{p}{1-q^2\rho^2} \qquad \ldots(2.1.2)$$

Replacing c by c_o in (2.1.1), it becomes MVLUE and therefore, the minimum variance of $\hat{Y}_2$ is obtained as follows:

$$V(\hat{\hat{Y}}_2)=\frac{\sigma^2}{n}\,\frac{1-q\,\rho^2}{1-q^2\,\rho^2} \qquad \ldots(2.1.3)$$

Test statistic and preliminary test estimator (PTE)

As it is already assumed that $\sigma_x^2 = \sigma_y^2 = \sigma^2$ and it is equal to unity, then to test the hypothesis H_o : $\mu_x=\mu_o$ against an alternative hypothesis H_1 : $\mu_x\neq\mu_o$, the well known normal test statistic is given by

$$Z=\sqrt{n}\,(\bar{x}-\mu_o) \qquad \ldots(2.2.1)$$

Letting $\mu_0=0$, without loss of generality, the hypothesis becomes as under

$$H_0 : \mu_x=0$$

Versus $\quad H_1 : \mu_x\neq 0 \qquad \ldots(2.2.2)$

and the test statistic (2.2.1) reduces to

$$Z=\sqrt{n}\,\bar{x} \qquad \ldots(2.2.3)$$

Now, an estimator of μ_y after preliminary test of significance is defined as below :

$$\hat{\mu}_{PTY}=\begin{cases} T_1=c\,(\bar{y}'-\rho\,\bar{x}')+(1-c)\,\bar{y}'' & \text{if } H_0 \text{ is accepted} \\ T_2=c\,[\bar{y}'+\rho\,(\bar{x}-\bar{x}')]+(1-c)\,\bar{y}'' & \text{if } H_1 \text{ is accepted.} \end{cases} \qquad \ldots(2.2.4)$$

Where $T_1 = T_2 - c\rho\bar{x}$ and T_2 is unbiased estimator of μ_y. Thus, B_1 (*i.e.*, bias of $\hat{\mu}_{PTY}) = -c\rho E(\bar{x}/H_0$ Accepted) P_r (H_0 Accepted). It is clear that if P_r (H_0 Accepted)$=1$, then $\hat{\mu}_{PTY} = T_1$ and $B_1 =$ Bias $(T_1) = -c\rho\mu_x$. Also if P_r (H_0 Accepted)$=0$, $B_1 = 0$. These above results always hold true even without deriving the actual expression on the assumption of normality.

The arbitrary constant c is taken same in both the situation just for simplicity. The $\hat{\mu}_{PTY}$ is called as preliminary test estimator (*PTE*) as it arises due to preliminary test of significance.

Evaluation of Bias, MSE and RE of $\hat{\mu}_{PTY}$

To evaluate the bias and *MSE* of $\hat{\mu}_{PTY}$, we require the joint distribution of $\bar{x}$, $\bar{x}'$ and $\bar{y}'$. It can be easily verified that the joint distribution of these is nothing but a trivariate normal distribution with mean (μ_x, μ_x, μ_y) and variance covariance matrix

$$\Sigma' = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \frac{1}{n}\rho \\ \frac{1}{n} & \frac{1}{m} & \frac{1}{m}\rho \\ \frac{1}{n}\rho & \frac{1}{m}\rho & \frac{1}{m} \end{bmatrix} \qquad \ldots(2.3.1)$$

Generally, m and u are not sufficiently large. However, if m and u are sufficiently large, one could assume the trivariate normal approximation for $(\bar{x}, \bar{x}', \bar{y}')$ even if the original bivariate population is not normal.

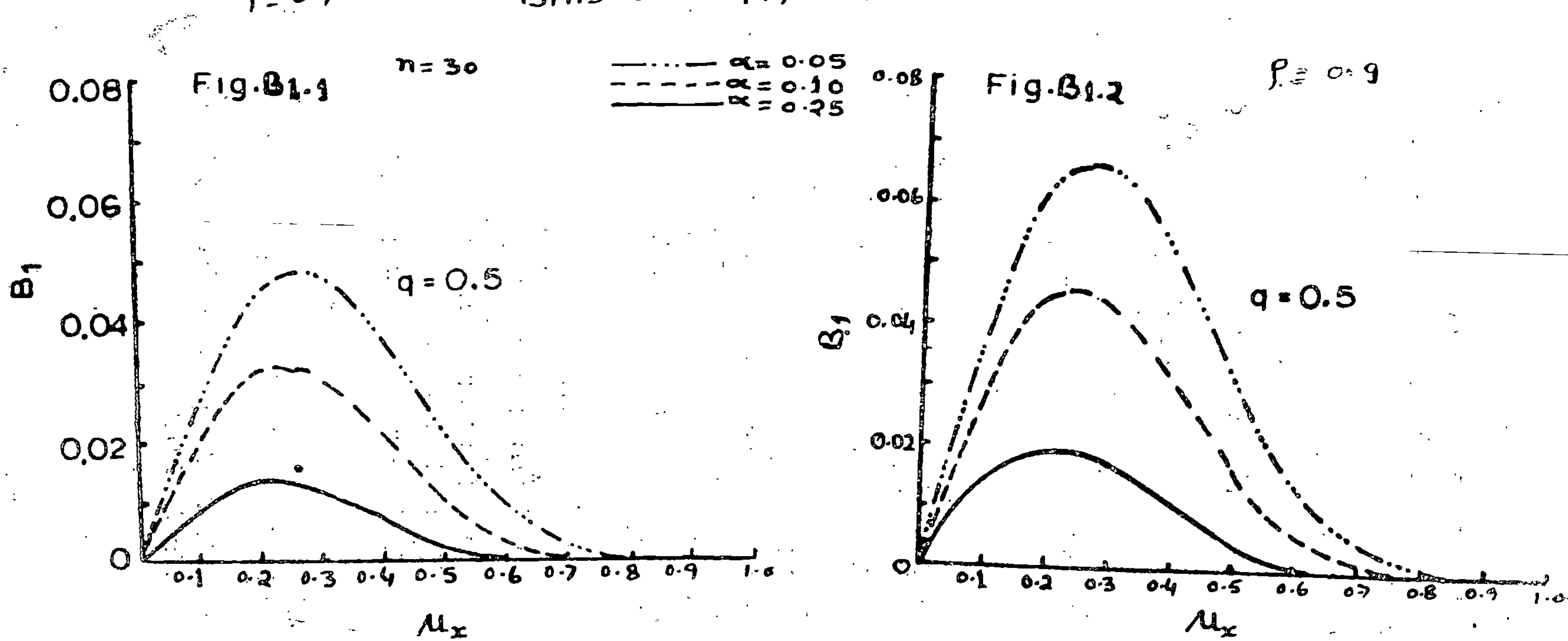
To obtain the expected value of $\hat{\mu}_{PTY}$, we proceed as follows :

$$\begin{aligned} E(\hat{\mu}_{PTY}) &= E\,[\hat{\mu}_{PTY} \mid \text{Accept } H_0]\, P_r\,(\text{Accept } H_0) \\ &\quad + E\,[\hat{\mu}_{PTY} \mid \text{Accept } H_1]\, P_r\,(\text{Accept } H_1) \\ &= E\,[T_1 \mid \; |\bar{x}| \leqslant Z_\alpha/\sqrt{n}]\, P_r\,(|\bar{x}| \leqslant Z_\alpha/\sqrt{n}) \\ &\quad + E\,[T_2 \mid \; |\bar{x}| > Z_\alpha/\sqrt{n}]\, P_r\,(|\bar{x}| > Z_\alpha/\sqrt{n}) \end{aligned} \qquad \ldots(2.3.2)$$

Where Z_α is $(1-\alpha/2)$ 100 per cent point on $N(0, 1)$ and α is level of significance. On derivation, the expression (2.3.2) takes the form as follows :

$$\begin{aligned} E(\hat{\mu}_{PTY}) &= \mu_y - c\rho E\,[\bar{x} \mid \; |\bar{x}| \leqslant Z_\alpha/\sqrt{n})\, P_r\,(|\bar{x}| \leqslant Z_\alpha/\sqrt{n}) \\ &= \mu_y - \frac{c\rho}{\sqrt{n}} \int_b^a Z\phi(z)\, dz - c\rho\mu_x \int_b^a \phi(z)\, dz \end{aligned} \qquad \ldots(2.3.3)$$

BIAS OF $\hat{\mu}_{PTY}$ - B_1



Where $Z=(\bar{x}-\mu_x)\sqrt{n}$, ϕ (.) is probability density function of $N(0, 1)$ and $a=Z_\alpha-\sqrt{n}\ \mu_x$ and $b=-Z_\alpha-\sqrt{n}\ \mu_x$. The bias of $\hat{\mu}_{PTY}$ is, therefore, given by

$$B_1=-c\rho/\sqrt{n}\int_b^a Z\,\phi(z)\,d_z-c\rho\mu_x\int_b^a \phi(z)\,d_z$$

$$=c\rho/\sqrt{n}\,[\phi(a)-\phi(b)]-c\rho\mu_x\,[\phi(a)-\phi(b)] \quad \ldots(2.3.4)$$

Where ϕ (.) is cumulative distribution function of $N(0, 1)$. It can be easily shown that $B_1=-c\rho\mu_x$ when $\alpha=0$, *i.e.*, null hypothesis H_0 is always accepted, and $B_1=0$ when $\alpha=1$, *i.e.*, H_1 is always accepted. The following properties of B_1 are noted :

(*i*) B_1 is antisymmetric about $\mu_x=0$. But $|B_1|$ is symmetric about $\mu_x=0$.

(*ii*) At $\mu_x=0$, both T_1 and T_2 are unbiased, hence $\hat{\mu}_{PTY}$ is unbiased.

(*iii*) As the $\lim\limits_{\mu_x\to\infty} Pr\,(H_0 \text{ Accepted})=0$, $\lim\limits_{\mu_x\to\infty} B_1=0$.

Thus, bias tend to zero when μ_x tends to infinity.

Behaviour of the bias with respect to μ_x is presented in the Figures—*B* 1.1 and *B* 1.2 for a set of values of n q α and ρ, *i.e.*, for $n=30$; $q=0.5$, $\alpha=0.05$, 0.10, 0.25 and $\rho=0.7$, 0.9. It is obvious from figures that the bias is zero at $\mu_x=0$. It first increases and then decreases as μ_x increases. The bias is very close to zero at $\mu_x=1$. It is perhaps so because of n being large. The bias found here is quite small almost in all cases.

Now, the *MSE* of $\hat{\mu}_{PTY}$ is given by

$$M_1=E\,(\hat{\mu}_{PTY}-\mu_y)^2$$

$$=E\,(\hat{\mu}^2_{PTY})-2\,\mu_y\,E\,(\hat{\mu}_{PTY})+\mu_y^2$$

M_1 may also be expressed as

$$M_1=E\,(T_2-\mu_y)^2-c^2\,\rho^2\,E\,[(\bar{x}-\mu_x)^2-\mu_x^2 \mid H_0 \text{ Accepted}]\,Pr\,(H_0 \text{ Accepted}).$$

Without deriving the actual expression of M_1 on the assumption of normality, the following results hold :

(*i*) If $Pr\,(H_0$ Accepted$)=1$, $\hat{\mu}_{PTY}=T_1$ and, therefore, $M_1=MSE.\,(T_1)$.

(*ii*) As $\mu_x \to \infty$, Pr (H_0 Accepted) $\to 0$, hence M_1 tends to $V(T_2)$.

However, $E(\hat{\mu}^2_{PTY}) = \mu_y^2 + \frac{c^2(1-\rho^2)}{m} + \frac{(1-c)^2}{u}$

$$+\frac{c^2\rho^2}{n} - \frac{c^2\rho^2}{n}\int_b^a z^2 \phi(z)\, dz$$

$$-\frac{2c\rho\mu_y}{n}\int_b^a Z\,\phi(z)\, dz - (2c\rho\,\mu_y\,\mu_x - c^2\rho^2\mu_x^2)\int_b^a \psi(z)\, dz \qquad \ldots(2.3.6)$$

After substituting the value of $E(\hat{\mu}^2_{P_TY})$ and $E\hat{\mu}_{PTY}$ from) (2.3.6) and (2.3.3) respectively, in (2.3.5), we obtain the *MSE* on simplification as follows :

$$M_1 = \frac{c^2(1-\rho^2)}{m} + \frac{(1-c)^2}{u} + \frac{c^2\rho^2}{n} - \frac{c^2\rho^2}{n}\int_b^a z^2\phi(z)\, dz$$

$$+ c^2\rho^2\mu_x^2 \int_b^a \phi(z)\, dz$$

$$= \frac{c^2(1-\rho^2)}{m} + \frac{(1-c)^2}{u} + \frac{c^2\rho^2}{n} + \frac{c^2\rho^2}{n}[a\,\phi(a) - b\,\phi(b)]$$

$$-c^2\rho^2\left(\frac{1}{n} - \mu_x^2\right)[\phi(a) - \phi(b)] \qquad \ldots(2.3.7)$$

If we consider the *MSE* as a function μ_x, say $M_1(\mu_x)$; the following properties may be observed :

(*i*) *MSE* is symmetric about $\mu_x = 0$

(*ii*) At $\mu_x = 0$, the *MSE* becomes as follows :

$$M_1(0) = \frac{c^2(1-\rho^2)}{m} + \frac{(1-c)^2}{u} + \frac{c^2\rho^2}{n}\alpha + \frac{c^2\rho^2}{n}\,2\,Z_\alpha\,\phi(Z_\alpha)$$

Letting $\alpha = 0$, it reduces to

$$[M_1(0)]_{\alpha=0} = \frac{c^2(1-\rho^2)}{m} + \frac{(1-c)^2}{u} \qquad \ldots(2.3.8)$$

This result implies that if $\mu_x=0$ and $\alpha=0$, the H_0 is always accepted and thus, the estimator T_1 with MSE given in (2.3.8) is always used.

(*iii*) If $\mu_x \to \infty$, it can be easily shown that

$$\underset{\mu_x \to \infty}{\text{Limit}} \; M_1(\mu_x) = \frac{c^2(1-\rho^2)}{m} + \frac{(1-c)^2}{u} + \frac{c^2\rho^2}{n} \quad \ldots(2.3.9)$$

This result confirms our intuitive reasoning that as $\mu_x \to \infty$, the alternative hypothesis H_1 must be always accepted and, therefore, usual estimator T_2 $(=\hat{\bar{Y}}_2)$ with MSE given in (2.3.9) is always used.

To study the relative efficiency (RE) of $\hat{\mu}_{PTY}$ as compared to T_2, it would be worthwhile to determine an optimum value of c by minimising M_1. The optimum value of c is, thus given by

$$C_t = \frac{np}{(1-q^2\rho^2) - npq\rho^2\left(\frac{1}{n} - \mu_x^2\right)[\phi(a)-\phi(b)] + pq\rho^2[a\phi(a)-b\phi(b)]} \quad \ldots(2.3.10)$$

Obviously, c_t is a function of μ_x and α, the size of the test. Therefore, this approach of getting c_t does not look practicable. The optimum value of c is, thus taken by (2.1.2) which is obtained by minimizing the variance of T_2 for comparison purpose of $\hat{\mu}_{PTY}$ and T_2 $(=\hat{\bar{Y}}_2)$.

The MSE of $\hat{\mu}_{PTY}$ can be factorised as $M_1=M_{11}+M_{12}$ where M_{11} is variance of T_2. Therefore the relative efficiency (RE) is given by

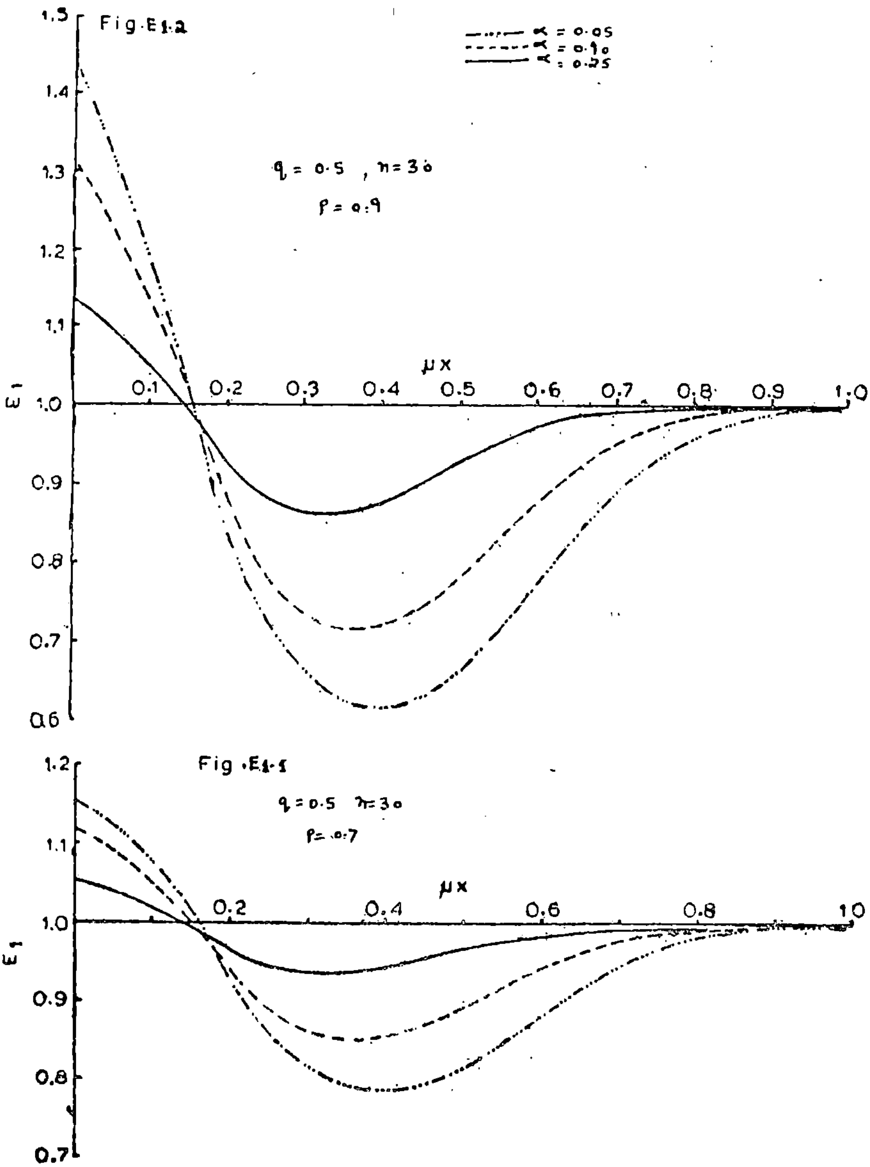
$$E_1 = \frac{M_{11}}{M_{11}+M_{12}} \quad \ldots(2.3.11)$$

It can be again verified that E_1 is symmetric about $\mu_x=0$, *i.e.* $E_1(-\mu_x)=E_1(\mu_x)$. Thus, it is sufficient to consider the RE only for $\mu_x \geqslant 0$.

Behaviour of RE w.r.t. μ_x is presented in Figures—$E_{1.1}$ and $E_{1.2}$ for a set of values of n, q, α and ρ, *i.e.*, for $n=30$; $q=0.5$, $\alpha=0.05$, 0.10, 0.25 and $\rho=0.7$, 0.9. It is seen that RE is maximum at $\mu_x=0$, which is, in fact, expected. But it decreases below unity as μ_x increases and it is near to unity at $\mu_x=1$.

However, one would like to choose the estimators with high efficiency and it would be ideal if RE is always greater than unity. To minimise the loss in RE, the criterion for selecting α given by Han and Baucroft (1968) will be used. Let E be RE which is a function of α and $\delta=\mu_x/\sigma_x$. Denote this function by $E(\alpha, \delta)$. Then the criterion is given by

BEHAVIOUR OF RELATIVE EFFICIENCY - E_1



"If the experimenter does not know the size of δ *and is willing to accept an estimator which has RE of no less than* E_0*, then among the set of estimators with* $\alpha \in A$ *where* $A=[\alpha\ ;\ E(\alpha, \delta) \geqslant E_0$ *for all* $\delta)$*, the estimator is chosen to maximise* $E(\alpha, \delta)$ *for all* α *and* δ*. Since* max $\delta E(\alpha, \delta)=E(\alpha=0)$*, he selects* $\alpha \in A$ *(say* α^**) which*

maximise E (α, 0) (say, E). This criterion will guarantee that the RE of chosen estimator is at least E_0 and it may become as large as E*''.*

Table 1 gives the values of E_1^* and E_{10} which are maximum and minimum of E_1, respectively, obtained from this criterion. It will help us to choose proper α^* which ensures the minimum RE at a chosen level E_{10}. The corresonding RE may be high as E_1^*. Though the RE have been given in the table only for $\rho=0.7$ and 0.9, it has been observed that for small values of ρ the gain and loss in RE is very small. However, for larger values of ρ there are substantial gains but risk of more losses are also probable as E_{10} is small. The Table 1 also depicts a similar behaviour of RE for increasing q, *i.e.*, the gain and loss in RE is very small for large values of q.

When variance-covariance matrix is unknown

In this section, a preliminary test estimation procedure will be discussed when Σ is not known. Before this procedure is discussed the usual estimator of μ_y on the second occasion will be briefly reviewed under the assumption that Σ is unknown.

Usual estimator of μ_y when Σ is unknown

The sampling plan will be same as cited in the preceding section.

Define

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

$$s'^2_x = \frac{1}{m-1}\sum_{i=1}^{m}(x_i-\bar{x}')^2,$$

$$s'^2_y = \frac{1}{m-1}\sum_{i=1}^{m}(y_i-\bar{y}')^2$$

$$s'_{xy} = \frac{1}{m-1}\sum_{i=1}^{m}(x_i-\bar{x}')(y_i-\bar{y}')$$

and $$\hat{\beta} = \frac{s'_{xy}}{s'^2_x},$$

TABLE 1

Maximum and Minimum Values of E_1

(E_1^* =Maximum, E_{10}=Minimum)

n	α^*	E	$q=0.3$		$q=0.4$		$q=0.5$		$q=0.6$		$q=0.7$	
			$\rho=0.7$	0.9	0.7	0.9	0.7	0.9	0.7	0.9	0.7	0.9
	0.50	E_1^*	1.022	1.043	1.018	1.038	1.014	1.032	1.010	1.027	1.007	1.021
		E_{10}	0.975	0.953	0.980	0.959	0.984	0.964	0.988	0.970	0.992	0.976
	0.25	E_1^*	1.809	1.185	1.070	1.159	1.154	1.134	1.039	1.109	1.025	1.084
20		E_{10}	0.900	0.823	0.917	0.842	0.934	0.860	0.961	0.881	0.968	0.904
	0.10	E_1^*	1.197	1.461	1.153	1.382	1.115	1.313	1.081	1.248	1.052	1.185
		E_{10}	0.788	0.559	0.821	0.688	0.855	0.719	0.890	0.754	0.925	0.797
	0.05	E_1^*	1.269	1.688	1.207	1.556	1.154	1.444	1.108	1.345	1.068	1.252
		E_{10}	0.700	0.541	0.742	0.580	0.788	0.616	0.836	0.658	0.886	0.711
	0.50	E_1^*	1.022	1.043	1.018	1.038	1.014	1.032	1.010	1.027	1.007	1.021
		E_{10}	0.975	0.952	0.979	0.958	0.984	0.964	0.988	0.969	0.992	0.976
	0.25	E_1^*	1.089	1.185	1.070	1.159	1.054	1.134	1.059	1.109	1,102	1.084

1	2	3	4	5	6	7	8	9	10	11	12	13
30		E_{10}	0.901	08.86	0.919	0.844	0.936	0.863	0.952	0.883	0.908	0.906
	0.10	E_1^*	1,197	1.451	1.153	1.382	1.115	1.313	1.081	1.248	1.052	1.185
		E_{10}	0.787	0.658	0.821	0.688	0.855	0.719	0.890	0.754	0.925	0.796
	0.05	E_1^*	1.269	1.688	1.207	1.556	1.154	1.444	1,108	1.345	1.068	1.252
		E_{10}	0.699	0.547	0.742	0.580	0.787	0.616	0.835	0.658	0,886	0.710
	0.50	E_1^*	1.022	1.043	1.018	1.038	1.014	1.032	1.010	1.037	1.007	1.021
		E_{10}	0.976	0.955	0.981	0.960	0.985	0.966	0.984	0.971	0.993	0.977
	0.25	E_1^*	1.089	1.185	1.070	1.159	1.054	1.134	1.039	1.109	1.025	1.084
40		E_{10}	0.901	0.825	0.913	0.843	0.935	0.862	0.952	0.882	0.968	0.904
	0.10	E_1^*	1.197	1.461	1.153	1.382	1.115	1.313	1.081	1.248	1.052	1,185
		E_{10}	0.783	0.653	0.877	0.685	0.852	0.714	0.888	0.750	0.923	0,792
	0.05	E_1^*	1.269	1.688	1.207	1.556	1.154	1.444	1.808	1.345	1.068	1.258
		E_{10}	0.710	0.560	0.751	0.592	0.796	0.628	0.842	0.669	0.891	0.720

the sample regression coefficient of Y on X. Thus, the usual estimator of μ_y is given by

$$\hat{\bar{Y}}'_2 = c[\bar{Y}' + \hat{\beta}(\bar{x} - \bar{x}')] + (1-c)\bar{Y}'' \qquad \text{...(3.1.1)}$$

Variance of $\hat{\bar{Y}}'_2$ due to Narain (1953) is as follows :

$$v(\hat{\bar{Y}}'_2) = \frac{c^2\sigma_y^2\,(1-\rho^2)}{m}\left[1 + \left(1 - \frac{m}{n}\right) - \frac{1}{m-3}\right] + \frac{(1-c)^2\,\sigma_y^2}{u} + \frac{c^2\rho^2\sigma_y^2}{n} \qquad \text{...(3.1.2)}$$

with the optimum value of c as given below :

$$c'_0 = \frac{m-3}{(m-2)(1-q^2\rho^2)-(1-q^2)} \qquad \text{...(3.1.3)}$$

Test statistic and preliminary test estimator (PTE)

To make a preliminary test of size α for testing the hypothesis

$$H_o : \mu_x = 0$$

versus

$$H_o : \mu_x \neq 0 \qquad \text{...(3.2.1)}$$

when σ_x^2 is unknown, the well known t-statistic is available and given by

$$t = \frac{\sqrt{n}\,\bar{x}}{s_x} \qquad \text{...(3.2.2)}$$

the statistic-t follows the t-distribution with $(n-1)$ degrees of freedom. Therefore, a preliminary test estimator subsequent to testing the hypothesis (3.2.1) is defined as follows :

$$\mu^*_{PTY} = \begin{cases} T'_1 = c(\bar{y}' - \hat{\beta}\bar{x}') + (1-c)\bar{y}'' & \text{If } H_0 \text{ is accepted} \\ T'_2 = (c[\bar{y}' + \hat{\beta}(\bar{x} - \bar{x}'')] + (1-c)\,\bar{y}'' & \text{If } H_1 \text{ is accepted} \end{cases} \qquad \text{...(3.2.3)}$$

where c is some arbitrary constant. The bias and MSE of it will now be derived.

Evaluation of Bias, MSE and RE of $\hat{\mu}^*_{PTY}$

Taking the expectation of $\hat{\mu}^*_{PTY}$, we obtain

$$E(\hat{\mu}^*_{PTY})=E\left[T_1' \;\middle|\; |t| \leqslant t_\alpha\right] Pr(|t| \leqslant t_\alpha)$$

$$+E\left[T_2' \;\middle|\; |t| > t_\alpha\right] Pr(|t| > t_\alpha) \quad ...(3.3.1)$$

where t_α is 100 $(1-\alpha/2)$ per cent point of t-distribution with $(n-1)$ degree of freedom. The joint distribution of $(\bar{x}, \bar{x}', \bar{y}')$ in this case also is trivariate normal and independent of $(s_x^2 \;,\; s'^2_x \;,\; s'_{xy})$. In practice, usually in the large scale surveys the sample size n is large Therefore, s_x^2 tends to σ_x^2 in probability and then s_x may be replaced by σ_x in the expression for 't' as given in (3.2.2). Assuming n to be large and putting σ_x in place of s_x in 't', the expectation of $\hat{\mu}^*_{PTY}$ may be obtained as follows :

$$E(\hat{\mu}^*_{PTY})=E\left[T_1' \;\middle|\; |\bar{x}| \leqslant \frac{t_\alpha \sigma_x}{\sqrt{n}}\right] Pr(|\bar{x}| \leqslant t_\alpha\sigma_x/\sqrt{n})$$

$$E\left[T_2' \;\middle|\; |\bar{x}| > t_\alpha\sigma_x/\sqrt{n}\right] Pr(|\bar{x}| > t_\alpha\sigma_x/\sqrt{n})$$
$$...(3.3.2)$$

Since $\hat{\beta}$ is independently distributed with the sample means and is also unbiased estimator of population regression coefficient β of Y on X, under the normality assumption, then the (3.3.2) reduces to

$$E(\hat{\mu}^*_{PTY})=\mu_y-c\beta\mu_x+c\beta E\left[\bar{x} \;\middle|\; |\bar{x}| > t_\alpha\sigma_x/\sqrt{n}\right]$$
$$Pr\left(|\bar{x}| > \frac{t_\alpha\sigma_x}{\sqrt{n}}\right)$$

which after simplification on previous lines may be obtained as follows :

$$E(\hat{\mu}^*_{PTY})=\mu_y+\frac{c\rho\sigma_y}{\sqrt{n}}[\phi(d')-\phi(d)]-c\rho\sigma_y\delta\,[\phi(d')-\phi(d)]$$
$$...(3.3.3)$$

where $d'=t_\alpha-\sqrt{n}\,\delta$ and $\alpha=-t_\alpha-\sqrt{n}\,\delta$. Thus, bias is given by

$$B_2=E(\hat{\mu}^*_{PTY})-\mu_y$$

$$=\frac{c\rho\sigma_y}{\sqrt{n}}[\phi(d')-\phi(d)]-c\rho\sigma_y\delta[\varphi(d')-\phi(d)] \quad \text{...(3.3.4)}$$

The relative bias of $\hat{\mu}^*_{PTY}$ is given by

$$B_2^*=B_2/\sigma_y=\frac{c\rho}{\sqrt{n}}[\phi(d')-\phi(d)]-c\rho\delta[\phi(d')-\phi(d)] \quad \text{...(3.3.5)}$$

B_2^* behaves w.r.t. δ in the same manner as B_1 behaves w.r.t. μ_x in the previous section.

The mean-square-error of $\tilde{\mu}^*_{PTY}$ is defined to be

$$M_2=E\,(\hat{\mu}^*_{PTY})-2\mu_y\quad \hat{E}\mu^*_{PTY})+\mu_y^2 \quad \text{...(3.3.6)}$$

Proceeding on the same line as done in the previous section, the *MSE* is found as follows :

$$M_2=\frac{c^2\,\sigma_y^2\,(1-\rho^2)}{m}\left[1+\left(1-\frac{m}{n}\right)\frac{1}{m-3}\right]$$

$$+\frac{(1-c^2)\,\sigma_y^2}{u}+\frac{c^2\rho^2\,\sigma_y^2}{n}$$

$$+\frac{c^2\,\sigma_y^2}{n}\left(\rho^2-\frac{1-\rho^2}{m-3}\right)[d'\,\phi\,(d')-d\,\phi\,(d)]$$

$$+\left[c^2\,\rho_y^2\,\delta^2\left(\rho^2+\frac{1-\rho^2}{m-3}\right)-\frac{c^2\sigma_y^2}{n}\left(\rho^2-\frac{1-\rho^2}{m-3}\right)\right]$$

$$\left[\phi\,(d')-\phi\,(d)\right]-\frac{2c^2\,\sigma_y^2}{n}\,\frac{1-\rho^2}{m-3}[\phi(d')-\phi(d)] \quad \text{...(3.3.7)}$$

TABLE 2

Maximum and Minimum Value of E_2

(E_2^* =Maximum, E_{20}=Minimum)

n	α^*	E	$q=0.3$		$q=0.4$		$q=0.5$		$q=0.6$		$q=0.7$	
			$p=0.7$	0.9	0.7	0.9	0.7	0.9	0.7	0.9	0.7	0.9
	0.50	E_2^*	1.020	1.043	1.016	1.037	1.011	1.031	1.007	1.025	1.003	1.017
		E_{20}	0.974	0.952	0.979	0.958	0.984	0.964	0.989	0.971	0.993	0.978
	0.20	E_2^*	1.109	1.261	1.083	1.218	1.059	1.178	1.037	1.137	1.017	1.091
20		E_{20}	0.858	0.762	0.883	0.785	0.908	0.812	0.933	0.841	0.959	0.879
	0.10	E_2^*	1.188	1.500	1.140	1.405	1.098	1.320	1.061	1.240	1.027	1.156
		E_{20}	0.761	0.623	0.799	0.660	0.838	0.695	0.880	0.787	0.923	0.934
	0.05	E_2^*	1.255	1.743	1.187	1.584	1.129	1.450	1.079	1.329	1.035	1.208
		E_{20}	0.658	0.507	0.705	0.541	0.757	0.580	0.815	0.630	0.879	0.699
	0.50	E_2^*	1.020	1.043	1.016	1.037	1.012	1.031	1.008	1.025	1.004	1.018
		E_{20}	0.974	0.953	0.979	0.938	0.984	0.964	0.989	0.971	0.993	0.978
	0.20	E_2^*	1.111	1.258	1.085	1.217	1.062	1.178	1.041	1.139	1.021	1.098

1	2	3	4	5	6	7	8	9	10	11	12	13
20		E_{20}	0.864	0.769	0.888	0.793	0.911	0.817	0.936	0.845	0.960	0.880
	0.10	E_2^*	1.191	1.493	1.144	1.402	1.103	1.321	1.067	1.244	1.035	1.167
		E_{20}	0.761	0.627	0.798	0.659	0.837	0.693	0.878	0.735	0.921	0.786
	0.05	E_2^*	1.259	1.733	1.193	1.580	1.136	1.452	1.087	1.336	1.045	1.224
		E_{20}	0.673	0.524	0.719	0.557	0.769	0.596	0.823	0.642	0.881	0.705
	0.50	E_2^*	1.021	1.043	1.016	1.037	1.012	1.031	1.008	1.025	1.005	1.019
		E_{20}	0.974	0.952	0.979	0.957	0.984	0.964	0.988	0.970	0.993	0.997
	0.20	E_2^*	1.112	1.256	1.086	1.216	1.063	1.178	1.043	1.140	1.024	1.101
30		E_{20}	0.866	0.774	0.897	0.797	0.913	0.821	0.936	0.848	0.959	0.880
	0.10	E_2^*	1.193	1.488	1.147	1.399	1.106	1.320	1.070	1.246	1.039	1.172
		E_{20}	0.769	0.636	0.805	0.667	0.843	0.701	0.883	0.741	0.923	0.790
	0.05	E_2^*	1.262	1.726	1.197	1.577	1.140	1.452	1.092	1.340	1.051	1.232
		E_{20}	0.673	0.521	0.718	0.555	0.768	0.593	0.822	0.639	0.880	0.700

Further, letting M_2 as function of δ and denoting it by M_2 (δ), it is found symmetric about $\delta=0$, *i.e.*, M_2 $(-\delta)=M_2$ (δ). If $\alpha=0$, *i.e.*, H_0 is always accepted which implies that $\delta=0$, then it can be easily verified that the expression (3.3.7) reduces to the *MSE* of T_1' which is the case when one always uses T_1'. Similarly, if $\alpha=1$, *i.e.*, H_1 is always accepted and T_2' is always used, then (3.3.7) becomes the variance of T_2' which confirms our intuitive reasoning that T_2' is always used. When δ tends to infinity, we again find that *MSE* of $\tilde{\mu}^*_{PTY}$ simplified to the variance of T_2'.

Now, the *RE* of $\tilde{\mu}^*_{PTY}$ as compared to the usual estimator $\hat{\bar{Y}}_2'$ $(=T_2')$ is given by

$$E_2=\frac{M_{21}}{M_{21}+M_{22}} \qquad \ldots(3.3.8)$$

as $M_2=M_{21}+M_{22}$ and M_{21} is the variance of T_2'. With the argument similar to that of given in the previous section, the value of c given in (3.3.3) will be used for obtaining *RE*. Since M_2 $(-\delta)=M_2$ (δ), E_2 may be considered as a function of δ and hence E_2 $(-\delta)=E_2$ (δ). Therefore, E_2 is symmetrie about $\delta=0$. Thus investigation, for $\delta\geqslant 0$ is only needed. The behaviour of *RE* of $\hat{\mu}^*_{PTY}$ w.r.t. δ is similar to *RE* of $\hat{\mu}$ w.r.t. μ_x. It is observed that E_2 is maximum at $\delta=0$, it exceeds unity in the neighbourhood of $\delta=0$, but decreases below unity as δ increases. It is again very close to unity at $\delta=1$.

Table 2 gives the values of E_2^* and E_{20} which are maximum and minimum of E_2, respectively, obtained from the criterion stated in the previous section. With the help of this table, one can choose α^* which ensures at least a minimum *RE* at a given level E_{20}. Though the Table 2 contains the values of *RE* for only $\rho=0.7$ and 0.9, it is here also observed that gain and loss in *RE* is very small for small values of ρ. For large value of ρ, the gains are appreciable but possibility of incurring more losses are also there as E_{20} is small. On the other hand, one can also see from this table that the trend of *RE* w.r.t. q indicates that the gain and loss in *RE* is small for large values of q. In fact, it has been observed but not reported in this table.

Conclusion :

In both the cases when variance-covariance matrix are known and unknown, the preliminary test estimators of the population mean on the second occasion are defined. It is found that the performance of the preliminary test estimator depends upon the various parameters *viz.*, n, ρ, α, q and δ. The values of n, α and q are at our disposal but the values of ρ and δ may not be known. However, from the criterion given by Han and Bancroft (1968), we have determined the values of *RE* and these are given in Tables 1 and 2. From both the tables, it is clear that for large values of ρ there are substantial gains in *RE* but risk of incurring more losses are also probable. This suggests that the preliminary test estimators should not be used for small values of ρ. It has also been observsd that the gain and loss in *RE* are very small for large values of q. Thus, for preliminary test estimator, one should not take very large value of q. Therefore, for a given set of ρ and level of minimum *RE*, a proper choice of α and q can be made so that the gain and loss in *RE* remain within reasonable limits.

It may be remarked that the *RE* given in Table 1 are not comparable to those given in Table 2 because the preliminary test estimators are compared to the estimators which are falling under H_1 and they are different in both the cases.

ACKNOWLEDGEMENT

The author is thankful to Dr. D. Singh, Director, I.A.S.R.I., New Delhi, under whose supervision this work was carried out. The author is also thankful to Dr. A.K. Srivastava, Scientist, S-2 for his valuable suggestion and painstaking cooperation during the preparation of this paper. The author thanks the refrees whose comments brought considerable improvement in the paper.

REFERENCES

Bancroft, T.A. (1944) : 'On bias in estimation due to the use of preliminary test of significance'. *Annals of Mathematical Statistics*, Vol. 15, 190-204.

Bennet, B.M. (1952) : 'Estimation of means on the basis of preliminary test of significance'. *Ann. Instt. Stat. Maths.*, Vol. 4, 31-43.

Cochran, W.G. (1977) : '*Sampling Technique*'. rd. ed., John Wiley and Sons, New York.

Han, C.P. and Bancroft T.A. (1968) : 'On pooling means when variance is unknown'. *Jour. Amer. Stat. Assoc.* Vol. 63, 1333-1342.

Han, C.P. (1973) : 'Double sampling with partial information on auxiliary variables'. *Jour. Amer. Stat. Assoc.* Vol. 68, 914-918.

Jessen, R.J. (1942) : 'Statistical investigation of sample survey for obtaining farm facts.' *Iowa Agril. Exp. Res. Bull.* 304.

Kitagava, T. (1963) : 'Estimation after preliminary test of significance'. *University of California Publication in Statistics*, Vol. 3, 147-186.

Narain, R.D. (1953) : 'On the recurrence formula in sampling on successive occasions'. *Jour. Indian Soc. Agril. Stat.*, Vol. 5, 96-99.

Patterson, H.D. (1950) : Sampling on successive occasions with partial replacements of units. *Jour. Roy. Stat. Soc.* Series B., Vol. 12, 241-255.

Sisodia, B.V.S. (1978) : 'On preliminary test estimation procedures and bias in repeated surveys'. *Uupublished Ph.D. Thesis*, IARI, New Delhi.

Tikkiwal, B.D. (1951) : Theory of successive sampling, *unpublished Diploma Thesis*, IARS (ICAR), New Delhi.

Tikkiwal, B.D. (1956) : A further contribution to the theory of univariate sampling on successive occasions. *Jour. Indian Soc. Agril. Stat.* Vol. 8, 84-90.

Yates, F. (1949) : *Sampling methods for census and sruveys*, Charles Griffin and Co. Ltd., London.