# EXPRESSIONS FOR THE LOWER BOUND TO CONFIDENCE COEFFICIENTS

*By* SAIBAL KUMAR BANERJEE

*Indian Statistical Institute, Calcutta*

*SUMMARY.* A lower bound to the probability of sample estimate plus (and minus) $t$- times ($t > 1$) estimate of sampling error covering the population mean (or total) is derived for samples from non-normal populations. Extensions of the result to the case of ratio estimates and multistage designs are also considered.

## 1. Introduction

Estimate of sampling variance of estimate indicates variability of the estimate and if the parent population is normal, sample estimate $\bar{x}_n$ plus (and minus) $t$-times estimate of sampling error covers population mean $m$ in $\alpha$ per cent of the cases where $\alpha$ is defined as

$$\alpha = \frac{1}{\sqrt{n-1}} \cdot \frac{100}{B\left(\frac{n-1}{2}, \frac{1}{2}\right)} \int_{-t}^{+t} \left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}} dt.$$

For a sample from a non-normal population if $\hat{m}$ is an estimate of $m$ and $\hat{V}(\hat{m})$(computed from sample readings) an estimate of sampling variance of $\hat{m}$, an expression for the lower bound to the probability that $\hat{m} \pm t\sqrt{\hat{V}(\hat{m})}$ covers $m$, may be of some interest. In a paper (Banerjee, 1956) it was shown that if $x_1, x_2, \ldots, x_n$ be a sample of size $n$ drawn at random with replacement from a population with mean $m$ and $B_2$-coefficient $B_2$, then, for $t > 1$,

$$\text{prob.} \left\{ |\bar{x}-m| < t\sqrt{\frac{\sum_1^n (x_i-\bar{x})^2}{n(n-1)}} \right\} > \frac{1}{\frac{B_2-3}{n}+1+\frac{2}{(t^2-1)^2}\left\{\frac{t^4}{n-1}+1\right\}}.$$

An extension of the result to the case of pps sampling in stratified multistage design may be of some interest. Some of the extensions are indicated below which are all based upon a simple lemma. It is seen that the role of $B_2$ in sampling with equal probability is taken over by a similar parameter in sampling with unequal probability. Stratified design introduces further parameters. The case of ratio estimate (simple ratio for single stratum, and combined ratio for $k$ strata) is also touched up without assuming bivariate normal distribution of $y$ and $x$. A 'probability' inequality in $R$ $\left(\frac{\sum_1^k M_{iy}}{\sum_1^k M_{ix}} = \text{population ratio to be estimated}\right)$ of the same form as Fieller's inequality is derived. An extension to multistage design is also indicated.

1.1. Lemma : *Let $\phi(x_1, x_2, \ldots x_p)$ be a function of $p$ stochastic variates such that $E(\phi) > 0$ and $E(\phi^2)$ exist.*

Then

$$\text{prob.}\left\{\ \phi \geqslant 0\right\} \geqslant \frac{[E(\phi)]^2}{E(\phi^2)} \qquad \dots \quad (1.1.1)$$

1.2. *Proof:* Let a variate $y$ be defined as

$$y = 1, \quad \text{if} \quad \phi \geqslant 0,$$
$$= 0, \quad \text{if} \quad \phi < 0.$$

Obviously $$\phi\, y \geqslant \phi,$$

or, $$E(\phi y) \geqslant E(\phi),$$

or, $$[E(\phi y)]^2 \geqslant [E(\phi)]^2, \quad (\because E(\phi) > 0)$$

or, $$E(\phi^2)\ E(y^2) \geqslant [E(\phi y)]^2 \geqslant [E(\phi)]^2 \quad \text{(Schwarz's inequality)}$$

or, $$\text{prob}\ (\phi \geqslant 0) = E(y^2) \geqslant \frac{[E(\phi)]^2}{E(\phi^2)}.$$

## 2. One stratum pps selection

2.1. Let there be a finite population consisting of $N$ units. Let $y_i$ denote variate value $y$ of the $i$-th unit. Let $n$ units be selected with replacement from $N$ units, with probability proportional to some measure of the units. Let $p_1, p_2, \dots, p_N$ denote the probability for the different units to be selected in a particular draw.

2.2. Let $z_s$ be an estimate of population total $M$ as built up from $s$-th selected unit. Obviously

$$z_s = \frac{y_i}{p_i}$$

if the $s$-th selected unit happens to be the $i$-th unit of the population.

2.3. Let us define a function $L$ of estimators $z_1, z_2, \dots z_n$ and $M$ and $t^2 (t > 1)$ as

$$L = \frac{t^2\hat{\lambda}}{n} - (\bar{z} - M)^2 \qquad \dots \quad (2.3.1)$$

where $$\hat{\lambda} = \frac{\sum_1^n (z_s - \bar{z})^2}{n-1}; \text{ and } \bar{z} = \frac{\Sigma\, z_s}{n}.$$

2.4. It can be easily shown that

$$\left.\begin{aligned} E(\hat{\lambda}) &= \lambda \\ E\{(\bar{z}-M)^2\} &= \frac{\lambda}{n} \\ E(\hat{\lambda}^2) &= \frac{B_2\lambda^2 - 3\lambda^2}{n} + \frac{2\lambda^2}{n-1} + \lambda^2 \\ E[\hat{\lambda}(\bar{z}-M)^2] &= \frac{B_2\lambda^2 - 3\lambda^2}{n^2} + \frac{\lambda^2}{n} \\ E\{(\bar{z}-M)^4\} &= \frac{B_2\lambda^2 - 3\lambda^2}{n^3} + \frac{3\lambda^2}{n^2} \end{aligned}\right\} \qquad \dots \quad (2.4.1)$$

where
$$\lambda = \sum_1^N p_i \left( \frac{y_i}{p_i} - M \right)^2$$

and
$$B_2 = \frac{\sum_1^N p_i \left( \frac{y_i}{p_i} - M \right)^4}{\lambda^2}.$$

We have accordingly

$$\left. \begin{aligned} E(L) &= (t^2-1)\,\frac{\lambda}{n} \\ E(L^2) &= (t^2-1)^2 \frac{\lambda^2}{n^2} \left[ \frac{B_2-3}{n} + 1 + \frac{2}{(t^2-1)^2} \left\{ \frac{t^4}{n-1} + 1 \right\} \right] \end{aligned} \right\} \quad \ldots \quad (2.4.2)$$

From (1.1.1), (2.3.1) and (2.4.2),

$$\text{prob.} \quad (L > 0) > \frac{1}{\frac{B_2-3}{n} + 1 + \frac{2}{(t^2-1)^2} \left\{ \frac{t^4}{n-1} + 1 \right\}}$$

$$\text{or,} \quad \text{prob.} \left\{ z + \frac{t\hat{\lambda}}{n} > M > z - \frac{t\hat{\lambda}}{n} \right\} > \frac{1}{\frac{B_2-3}{n} + 1 + \frac{2}{(t^2-1)^2} \left\{ \frac{t^4}{n-1} + 1 \right\}} \ldots \quad (2.4.3)$$

2.5. Table 1 below gives numerical values of the lower bound to the probability $z + \frac{t\hat{\lambda}}{n} > M > z - \frac{t\hat{\lambda}}{n}$ for $t = 3$ and sample size $n = 4, 6, 8, 10, 12, 20, 30, 50, 100$ for different $B_2$—values.

TABLE 1. LOWER BOUND OF PROBABILITY OF THE INEQUALITY $z + \frac{t\hat{\lambda}}{n} > M > z - \frac{t\hat{\lambda}}{n}$

(values worked out from (2.4.3) taking $t = 3$)

| $B_2$-value | sample size $n$ = | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 4 | 6 | 8 | 10 | 12 | 20 | 30 | 50 | 100 |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| 1.0 | 0.727 | 0.830 | 0.875 | 0.899 | 0.914 | 0.939 | 0.951 | 0.959 | 0.964 |
| 2.0 | 0.615 | 0.729 | 0.789 | 0.825 | 0.849 | 0.897 | 0.921 | 0.941 | 0.953 |
| 3.0 | 0.533 | 0.650 | 0.718 | 0.762 | 0.793 | 0.859 | 0.894 | 0.923 | 0.946 |
| 4.0 | 0.471 | 0.587 | 0.659 | 0.708 | 0.744 | 0.823 | 0.868 | 0.907 | 0.937 |
| 5.0 | 0.421 | 0.535 | 0.609 | 0.661 | 0.700 | 0.791 | 0.844 | 0.891 | 0.929 |

From Table 1 it is seen that if $B_2$ be some thing like 4.0 (or less), working with three time the sampling error $z \pm 3\sqrt{\hat{V}(z)}$ will cover the true value in about 70.8 per cent of the cases (or more) if $n = 10$. If, however, $B_2$ be 3.0 or less (in pps sampling $B_2$ is likely to be small in general) working again with three times the sampling error $z \pm 3\sqrt{\hat{V}(z)}$ will cover the true value in about 76.2 per cent of the cases or more.

## 3. $K$ strata pps selection

3.1. Let there be $K$ finite populations consisting of $N_1, N_2 \ldots, N_k$ units. Let $y_{ij}$ denote variate value $y$ of the $j$-th unit of the $i$-th population. Let us consider a scheme of sampling where $n_i$ units are selected with replacement with probability proportional to some measure of the units from the $i$-th population ($i = 1, 2, \ldots, k$). Let $p_{ij}$ denote the probability that the $j$-th unit of the $i$-th population will appear in a particular selection while sampling for units from the $i$-th population. Obviously, $\sum\limits_{j=1}^{N_i} p_{ij} = 1$ (for $i = 1, 2, \ldots, k$).

3.2. Let $z_{is}$ be an estimate of population total $M_i$ of the $i$-th population as built up from the $s$-th selected unit, among $n_i$ units selected from the $i$-th population. Obviously

$$z_{is} = \frac{y_{ij}}{p_{ij}}$$

if the $s$-th selected unit happens to be $j$-th unit of the $i$-th population.

3.3. Let us define a function $L$ of estimators $z_{is}$ ($s = 1, 2, \ldots n_i$ ; $i = 1, 2 \ldots, k$) and $M_1, M_2 \ldots M_k$ and $t^2 (t > 1)$ as

$$L = t^2 \sum_1^k \frac{\lambda_i}{n_i} - \left\{ \sum_1^k \bar{z}_i - \sum_1^k M_i \right\}^2 \qquad \ldots \quad (3.3.1)$$

where
$$\lambda_i = \frac{\sum\limits_{s=1}^{n_i} (z_{is} - \bar{z}_i)^2}{n_i - 1}$$

and
$$\bar{z}_i = \frac{\sum\limits_{s=1}^{n_i} z_{is}}{n_i} \qquad (i = 1, 2, \ldots k).$$

3.4. We have
$$L = \sum_1^k \left\{ t^2 \frac{\lambda_i}{n_i} - (\bar{z}_i - M_i)^2 \right\} + \sum_{\substack{i,j=1 \\ (i \neq j)}}^k (\bar{z}_i - M_i)(\bar{z}_j - M_j) \qquad \ldots \quad (3.4.1)$$

Hence
$$E(L) = (t^2 - 1) \sum_1^k \frac{\lambda_i}{n_i} \qquad \ldots \quad (3.4.2)$$

where
$$\lambda_i = \sum_{j=1}^{N_i} p_{ij} \left( \frac{y_{ij}}{p_{ij}} - M_i \right)^2 \qquad \text{(for } i = 1, 2, \ldots k).$$

3.5. From (3.4.1)

$$L^2 = \left( \sum_1^k l_i \right)^2 + \left\{ \sum_{\substack{i,j=1 \\ (i \neq j)}}^k (\bar{z}_i - M_i)(\bar{z}_j - M_j) \right\}^2 +$$
$$+ 2 \left( \Sigma\, l_i \right) \left\{ \sum_{\substack{i,j=1 \\ (i \neq j)}}^k (\bar{z}_i - M_i)(\bar{z}_i - M_j) \right\} \qquad \ldots \quad (3.5.1)$$

where
$$l_i = \frac{t^2 \lambda_i}{n_i} - (\bar{z}_i - M_i)^2 \quad \text{(for } i = 1, 2, \ldots k).$$

It can be easily shown

$$E\left\{\left(\sum_{1}^{k} l_i\right)^2\right\} = E\left[\sum_{1}^{k} l_i^2 + \sum_{\substack{i,j=1\\(i\neq j)}}^{k} l_i l_j\right]$$

$$E\left\{\sum_{1}^{k} l_i^2\right\} = (t^2-1)^2 \sum_{i}^{k}\left[\frac{\lambda_i^2}{n_i^2}\left\{\frac{B_{2i}-3}{n_i}+1+\frac{2}{(t^2-1)^2}\left(\frac{t^4}{n_i-1}+1\right)\right\}\right]$$

where
$$B_{2i} = \frac{\sum\limits_{j=1}^{N_i} p_{ij}\left(\dfrac{y_{ij}}{p_{ij}}-M_i\right)^4}{(\lambda_i)^2}. \qquad \ldots (3.5.2)$$

$$E\left\{\sum_{\substack{i,j=1\\(i\neq j)}}^{k} l_i l_j\right\} = (t^2-1)^2 \sum_{\substack{i,j=1\\(i\neq j)}}^{k} \frac{\lambda_i\lambda_j}{n_i n_j} = (t^2-1)^2\left\{\left(\sum_{1}^{k}\frac{\lambda_i}{n_i}\right)^2 - \sum_{1}^{k}\frac{\lambda_i^2}{n_i^2}\right\}. \qquad \ldots (3.5.3)$$

$$E\left\{\sum_{\substack{i,j=1\\(i\neq j)}}^{k} (z_i-M_i)(z_j-M_j)\right\}^2 = 4\sum_{\substack{i,j=1\\(i<j)}}^{k}\frac{\lambda_i\lambda_j}{n_i n_j} = 2\sum_{\substack{i,j=1\\(i\neq j)}}^{k}\frac{\lambda_i\lambda_j}{n_i n_j} = 2\left\{\left(\sum_{1}^{k}\frac{\lambda_i}{n_i}\right)^2 - \sum_{1}^{k}\frac{\lambda_i^2}{n_i^2}\right\}.$$
$$\ldots (3.5.4)$$

$$E\left[\left\{\sum_{1}^{k} l_i\right\}\left\{\sum_{\substack{i,j=1\\ i\neq j}}^{k} (z_i-M_i)(z_j-M_j)\right\}\right] = 0. \qquad \ldots (3.5.5)$$

3.6. From (3.5.1)–(3.5.5) it follows

$$E(L^2) = (t^2-1)^2\left[\sum_{1}^{k}\frac{\lambda}{n}(B_{2i}-3) + \left(\sum_{1}^{k}\frac{\lambda_i}{n_i}\right)^2 + \frac{2}{(t^2-1)^2}\left\{\left(\sum_{1}^{k}\frac{\lambda_i^2}{n_i^2(n_i-1)}\right)t^4 + \left(\sum_{1}^{k}\frac{\lambda_i}{n_i}\right)^2\right\}\right.$$

$$= \left(\sum_{1}^{k}\frac{\lambda_i}{n_i}\right)^2 (t^2-1)^2\left[\frac{\sum\limits_{1}^{k}\dfrac{\lambda_i^2}{n_i^3}(B_{2i}-3)}{\left(\sum\limits_{1}^{k}\dfrac{\lambda_i}{n_i}\right)^2} + 1 + \frac{2}{(t^2-1)^2}\left\{\frac{t^4\sum\limits_{1}^{k}\dfrac{\lambda_i^2}{n_i^2(n_i-1)}}{\left(\sum\limits_{1}^{k}\dfrac{\lambda_i}{n_i}\right)^2}+1\right\}\right]$$
$$\ldots (3.6.1)$$

From (1.1.1), (3.3.1), (3.4.2) and (3.6.1),

$$\text{prob.}\ \ (L \geqslant 0) \geqslant \frac{\{E(L)\}^2}{E(L^2)}$$

$$\geqslant \frac{1}{\dfrac{\sum_{1}^{k} \frac{\lambda_i^2}{n_i^3}(B_{2i}-3)}{\left(\sum_{1}^{k} \frac{\lambda_i}{n_i}\right)^2}+1+\dfrac{2}{(t^2-1)^2}\left\{\dfrac{t^4 \sum_{1}^{k} \frac{\lambda_i^2}{n_i^2(n_i-1)}}{\left(\sum_{1}^{k} \frac{\lambda_i}{n_i}\right)^2}+1\right\}} \quad \ldots \quad (3.6.2)$$

or,
$$\sum_{1}^{k} \bar{x}_i + t\sqrt{\sum_{1}^{k} \frac{\lambda_i}{n_i}} \geqslant \sum_{1}^{k} M_i \geqslant \sum_{1}^{k} \bar{x}_i - t\sqrt{\sum_{1}^{k} \frac{\lambda_i}{n_i}}$$

with probability equal to or greater than the right hand most expression of (3.6.2).

3.7. If all the $n_i$'s are equal to $n$ the expression for the lower bound takes the form

$$\frac{1}{\dfrac{1}{n}\dfrac{\sum_{1}^{k} \lambda_i^2 (B_{2i}-3)}{\left(\sum_{1}^{k}\lambda_i\right)^2}+1+\dfrac{2}{(t^2-1)^2}\left\{\dfrac{t^4}{n-1}\cdot\dfrac{\sum_{1}^{k}\lambda_i^2}{\left(\sum_{1}^{k}\lambda_i\right)^2}+1\right\}} \quad \ldots \quad (3.7.1)$$

which is equal to
$$\frac{1}{\frac{\theta}{n}(\bar{B}_2-3)+1+\frac{2}{(t^2-1)^2}\left\{\frac{\theta t^4}{n-1}+1\right\}} \quad \ldots \quad (3.7.2)$$

where
$$\bar{B}_2 = \frac{\sum_{1}^{k} \lambda_i^2 B_{2i}}{\sum_{1}^{k} \lambda_i^2}\ ;\ \theta = \frac{\sum_{1}^{k} \lambda_i^2}{\left(\sum_{1}^{k}\lambda_i\right)^2} = \frac{\left\{\frac{C.V.(\lambda)}{100}\right\}^2+1}{K}.$$

Table 2 below gives numerical values of (3.7.2) for $K = 8$; $n = 4, 8, 12, 16$; $\bar{B}_2 = 2, 3, 4$, and $C.V.(\lambda) = 0.0, 25.0, 50.0, 75.0, 100.0, 125.0,$ and $150.0$.

TABLE 2. LOWER BOUND OF THE PROBABILITY

$$\sum_1^k \bar{z}_i + t\sqrt{\sum_1^k \frac{\hat{\lambda}_i}{n}} > \sum_1^k M_i > \sum_1^k \bar{z}_i - t\sqrt{\sum_1^k \frac{\hat{\lambda}_i}{n}}$$

FOR A STRATIFIED DESIGN OF 8 STRATA AND 4, 8, 12 AND 16 UNITS PER STRATUM FOR DIFFERENT VALUES OF CV($\lambda$)

(values worked out from (3.7.2) taking $t = 3$)

| $\beta_2$-value | CV($\lambda$) values as | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0.0 | 25.0 | 50.0 | 75.0 | 100.0 | 125.0 | 150.0 |
| (0) | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | | | number of units per stratum = 4 | | | | |
| 2.0 | .905 | .901 | .890 | .872 | .848 | .819 | .786 |
| 3.0 | .880 | .875 | .860 | .830 | .805 | .769 | .723 |
| 4.0 | .856 | .850 | .832 | .803 | .760 | .724 | .678 |
| | | | number of units per stratum = 8 | | | | |
| 2.0 | .943 | .941 | .936 | .928 | .917 | .903 | .887 |
| 3.0 | .929 | .927 | .919 | .908 | .892 | .872 | .849 |
| 4.0 | .916 | .913 | .903 | .888 | .867 | .842 | .814 |
| | | | number of units per stratum = 12 | | | | |
| 2.0 | .953 | .952 | .949 | .943 | .936 | .927 | .917 |
| 3.0 | .943 | .942 | .937 | .929 | .918 | .905 | .889 |
| 4.0 | .934 | .932 | .926 | .915 | .901 | .884 | .863 |
| | | | number of units per stratum = 16 | | | | |
| 2.0 | .957 | .957 | .954 | .951 | .945 | .939 | .931 |
| 3.0 | .950 | .949 | .946 | .940 | .932 | .921 | .909 |
| 4.0 | .943 | .942 | .937 | .929 | .918 | .905 | .889 |

From table 2 it is seen that for a design containing 8 strata and $n$ units per stratum ($n > 4$), confidence statement of the form $\sum_1^k \bar{z}_i + t\sqrt{\sum_1^k \frac{\hat{\lambda}_i}{n}} > \sum_1^k M_i > \sum_1^k \bar{z}_i - t\sqrt{\sum_1^k \frac{\hat{\lambda}_i}{n}}$ will be true in 70.6 per cent of the cases (or more) if $\beta_2 \leq 4$ and coefficient of variation of $\lambda$ values be less than or equal to 100. If the number of strata be increased, other parameters remaining the same, the probability (as judged by the expression for the lower bound) will increase.

3.8. With respect to stratified designs having a constant number of $n$ units per stratum there is another method of estimation of sampling error and allied confidence interval for the population mean (or total). This method at times may be operationally easy and thus less costly in large scale tabulations. Hence in this context it may not be out of place to discuss that method. Denoting as before by $z_{is}$ ($s = 1, 2, \ldots n$; $i = 1, 2, \ldots, k$) estimate of population total $M_i$ as built up from the $s$-th selected unit from $n$ units selected from the $i$-th population, a set of estimators and a function $L_1$ may be defined as

$$a_s = \sum_{i=1}^{k} z_{is} \qquad \ldots \quad (3.8.1)$$

$$L_1 = \frac{t^2 \sum\limits_{s=1}^{n} (a_s-\bar{a})^2}{n(n-1)} - (\bar{a}-\sum_1^k M_i)^2 \qquad \ldots \quad (3.8.2)$$

where

$$\bar{a} = \frac{\sum\limits_{s=1}^{n} a_s}{n} .$$

It can be easily shown that

$$E(L_1) = (t^2-1) \sum_1^k \frac{\lambda_i}{n},$$

and

$$\frac{\{E(L_1)\}^2}{E(L_1^2)} = \frac{1}{\dfrac{\theta(\bar{B}_2-3)}{n} +1+\dfrac{2}{(t^2-1)^2}\left\{ \dfrac{t^4}{n-1} +1 \right\}} \qquad \ldots \quad (3.8.3)$$

where $\lambda_i$, $\bar{B}_2$ and $\theta$ are as defined earlier in paras 3.4 and 3.7.

Since $\theta < 1$, comparing (3.8.3) with (3.7.2) it is seen that (3.7.2) will always be greater than (3.8.3). Hence if judged only by this criterion (viz. the expression for the lower bound of probability of confidence statement being true) confidence statement of the form $\bar{a} \pm t \sqrt{\dfrac{\Sigma (a_s-\bar{a})^2}{n(n-1)}}$ is not to be preferred over $\sum\limits_1^k \hat{z}_i \pm t \sqrt{\sum\limits_1^k \dfrac{\lambda_i}{n}}$. If, however, number of units per stratum is large (something like 16 or more) the second method may be used in preference over the first.

## 4. ONE STRATUM PPS SELECTION RATIO ESTIMATE

4.1. One stratum, pps selection, ratio estimate : Let there be a finite population consisting of $N$ units. Let $y_i$, $x_i$ denote respectively variate values of character $y$ and $x$ of the $i$-th unit. Let $n$ units be selected with replacement from $N$ units with probability proportional to some measure of the units. Let $p_1, p_2 \ldots p_N$ denote the probability for the different units to be selected in a particular draw.

4.2. Let $z_s$ and $w_s$ be respectively estimates of population totals $M_y$ and $M_x$ of character $y$ and $x$ as built up from the $s$-th selected unit. Obviously

$$z_s = \frac{y_i}{p_i} \; ; \; w_s = \frac{x_i}{p_i}$$

if the $s$-th selected happens to be the $i$-th unit of the population.

4.3. Let us define a function $L$ of estimators $z_1, z_2, \ldots z_n$, $w_1, w_2, \ldots w_n$ and $M_y$ and $M_x$ and $t^2 (t > 1)$ as

$$L \equiv \frac{t^2}{n(n-1)} \left[ \sum_1^n \left\{ z_i - Rw_i - (\bar{z} - R\bar{w}) \right\}^2 \right] - (\bar{z} - R\bar{w})^2 \quad \ldots \quad (4.3.1)$$

$$= \frac{t^2}{n} \left\{ \lambda_z + R^2 \lambda_w - 2Rr\sqrt{\lambda_z \lambda_w} \right\} - (\bar{z} - R\bar{w})^2 \quad \ldots \quad (4.3.2)$$

where $$R = \frac{M_y}{M_x} \; ; \; \bar{z} = \frac{\sum_1^n z_i}{n} \; ; \; \bar{w} = \frac{\sum_1^n w_i}{n} \; ;$$

$$\lambda_z = \frac{\sum_1^n (z_i - \bar{z})^2}{n-1} \; ; \; \lambda_w = \frac{\sum_1^n (w_i - \bar{w})^2}{n-1} \; ; \; r\sqrt{\lambda_z \lambda_w} = \frac{\sum_1^n z_i w_i - n\bar{z}\bar{w}}{n-1} .$$

4.4. Treating $z_i - Rw_i$ as a variate and taking mathematical expectations of $L$ and $L^2$ it can be shown that probability $\{L \geqslant 0\} \geqslant P_0$,

where $$P_0 = \frac{1}{\dfrac{B_2(z, w) - 3}{n} + 1 + \dfrac{2}{(t^2-1)^2} \left\{ \dfrac{t^4}{n-1} + 1 \right\}} \quad \ldots \quad (4.4.1)$$

where $$B_2(z, w) = \frac{\sum_1^N p_i \left( \dfrac{y_i - Rx_i}{p_i} \right)^4}{\left\{ \sum_1^N p_i \left( \dfrac{y_i - Rx_i}{p_i} \right)^2 \right\}^2} . \quad \ldots \quad (4.4.2)$$

Hence we have from (1.1.1), (4.3.1) and (4.4.1),

$$\frac{t^2}{n} \left\{ \lambda_z + R^2 \lambda_w - 2Rr\sqrt{\lambda_z \lambda_w} \right\} \geqslant (\bar{z} - R\bar{w})^2 \quad \ldots \quad (4.4.3)$$

with probability greater than (or equal to) $P_0$.

4.5. Following Fieller (1940) "confidence limits" for $R$ can be derived from (4.4.3). In brief the method may be indicated as under. From (4.4.3) a quadratic equation in $R$ and a quadratic inequality in $R$ can be derived as under :

Quadratic equation : $$R^2(\bar{w}^2 - p\lambda_w) - 2R(\bar{z}\bar{w} - pr\sqrt{\lambda_z \lambda_w}) + (\bar{z}^2 - p\lambda_z) = 0. \quad (4.5.1)$$

Quadratic inequality : $$R^2(\bar{w}^2 - p\lambda_w) - 2R(\bar{z}\bar{w} - pr\sqrt{\lambda_z \lambda_w}) + (\bar{z}^2 - p\lambda_z) \leqslant 0. \quad \ldots \quad (4.5.2)$$

where $$p = \frac{t^2}{n}.$$

Actual numerical values of $R$ which satisfy (4.5.1) for clarity of exposition, may be considered under three heads :

(a) $\bar{w}^2 - p\lambda_w > 0$ ; (b) $\bar{w}^2 - p\lambda_w = 0$ ; (c) $\bar{w}^2 - p\lambda_w < 0$.

For (a), roots of equation (4.5.1) are real as the discriminant

$$D = 4\{(\bar{z}\bar{w}-pr\sqrt{\lambda_z\lambda_w})^2-(\bar{w}^2-p\lambda_w)(\bar{z}^2-p\lambda_z)\}$$
$$= 4\bar{z}^2\bar{w}^2\{p^2(C_{ww}-C_{zw})^2+p(1-pC_{ww})(C_{ww}+C_{zz}-2C_{zw})\} > 0$$

where
$$C_{zz} = \frac{\lambda_z}{\bar{z}^2}\ ;\ C_{ww} = \frac{\lambda_w}{\bar{w}^2}\ ;\ C_{zw} = \frac{r\sqrt{\lambda_z\lambda_w}}{\bar{z}\bar{w}}.$$

Hence from (4.5.2) limits for $R$ are

$$R_1 \leqslant R \leqslant R_2 \qquad \ldots \quad (4.5.3)$$

where $R_1$ and $R_2$ are the roots of (4.5.1) such that $R_2 > R_1$.

For (b), limits for $R$ are derivable from the relation

$$\bar{z}^2-p\lambda_z \leqslant 2R(\bar{z}\bar{w}-pr\sqrt{\lambda_z\lambda_w}). \qquad \ldots \quad (4.5.4)$$

Under (c) there can arise the sub-cases

$$\text{(c.1)} \quad \bar{z}^2-p\lambda_z \geqslant 0$$
$$\text{(c.2)} \quad \bar{z}^2-p\lambda_z < 0.$$

For (c.1) the discriminant of the equation (4.5.1) is positive and as such if $R_1$ and $R_2$ be the roots of the equation (4.5.1), limits for $R$ will be

$$R \leqslant R_1 \quad \text{or,} \quad R \geqslant R_2 \qquad \ldots \quad (4.5.5)$$

where $R_1$ and $R_2$ will satisfy the relation

$$R_2 > 0 > R_1. \qquad \ldots \quad (4.5.6)$$

For (c.2) depending upon the numerical values of $t$ and $r$ for given $\bar{z}^2-p\lambda_z$ and $\bar{w}^2-p\lambda_w$ the discriminant will be

either (c.21) positive
or, (c.22) negative.

For (c.21) limits for $R$ will be of the nature (4.5.5). For (c.22) as the roots of (4.5.1) will be imaginary any numerical value of $R$ will satisfy (4.5.2) and as such limits for $R$ derivable from (4.4.3) will be $\infty \geqslant R \geqslant -\infty$.

4.6. Limits of the nature (4.5.5) are practically useless and considering the very nature of the limits, the limits derivable from (4.4.3) cannot strictly be called confidence limits. Such limitations, however, apply equally to the bi-variate approach of Fieller as well.

## 5. $K$ STRATA PPS SELECTION COMBINED RATIO ESTIMATE

5.1. $K$ strata, pps selection, combined ratio estimate: Let there be $K$ finite populations consisting of $N_1, N_2 \ldots N_k$ units. Let $y_{ij}, x_{ij}$ denote respectively variate values of character $y$ and $x$ of the $j$-th unit of the $i$-th population. Let us consider a scheme of sampling where $n_i$ units are selected with replacement with probability proportional to some measure of the units from the $i$-th population ($i = 1, 2, \dots k$). Let $p_{ij}$ denote the probability that the $j$-th unit of the $i$-th population will appear in a particular selection while sampling for units from the $i$-th population.

Obviously $$\sum_{j=1}^{N_i} p_{ij} = 1 \text{ for } (i = 1, 2, \dots, K).$$

5.2. Let $z_{is}$ and $w_{is}$ be respectively estimates of population totals $M_{iy}$ and $M_{ix}$ of character $y$ and $x$ of the $i$-th population as built up from the $s$-th selected unit among $n_i$ units selected from the $i$-th population. Obviously

$$z_{is} = \frac{y_{ij}}{p_{ij}}; \; w_{is} = \frac{x_{ij}}{p_{ij}}$$

if the $s$-th selected unit happens to be the $j$-th unit of the $i$-th population.

5.3. Let us define a function $L$ of estimators $z_{is}$ and $w_{is}$ ($s = 1, 2, \dots n_i$; $i = 1, 2, \dots k$) and $M_{iy}$ and $M_{ix}$ ($i = 1, 2, \dots k$) and $t^2 (t > 1)$ as

$$L \equiv t^2 \sum_{i=1}^{k} \frac{\sum_{s=1}^{n} (u_{is} - \bar{u}_i)^2}{n_i(n_i - 1)} - \left( \sum_{1}^{k} \bar{u}_i \right)^2 \quad \dots \quad (5.3.1)$$

$$= t^2 \sum_{i=1}^{k} \frac{\sum_{s=1}^{n_i} \left\{ z_{is} - \bar{z}_i - R(w_{is} - \bar{w}_i) \right\}^2}{n_i(n_i - 1)} - \left\{ \sum_{1}^{k} \bar{z}_i - R \left( \sum_{1}^{k} \bar{w}_i \right) \right\}^2 \quad \dots \quad (5.3.2)$$

$$= t^2 \sum_{i=1}^{k} \frac{1}{n_i} \left\{ \lambda_{iz} + R^2 \lambda_{iw} - 2Rr_i \sqrt{\lambda_{iz} \lambda_{iw}} \right\} - \left\{ \sum_{i=1}^{k} \bar{z}_i - R \sum_{i=1}^{k} \bar{w} \right\}^2 \quad \dots \quad (5.3.3)$$

where $$u_{is} = z_{is} - M_{iy} - R(w_{is} - M_{ix}) \; ; \; \bar{u}_i = \frac{\sum_{s=1}^{n_i} u_{is}}{n_i}$$

$$\bar{z}_i = \frac{\sum_{s=1}^{n_i} z_{is}}{n_i} \; ; \; \bar{w}_i = \frac{\sum_{s=1}^{n_i} w_{is}}{n_i} \; ;$$

$$\lambda_{iz} = \frac{\sum_{s=1}^{n_i} (z_{is} - \bar{z}_i)^2}{n_i - 1} \; ; \; \lambda_{iw} = \frac{\sum_{s=1}^{n_i} (w_{is} - \bar{w}_i)^2}{n_i - 1} \; ;$$

$$r_i \sqrt{\lambda_{iz} \lambda_{iw}} = \frac{\sum_{s=1}^{n_i} z_{is} w_{is} - n_i \bar{z}_i \bar{w}_i}{n_i - 1} \; ; \; R = \frac{\sum_{i=1}^{k} M_{iy}}{\sum_{1}^{k} M_{ix}}.$$

5.4. It can be easily shown that

$$E(L) = (t^2-1) \sum_{1}^{k} \frac{\lambda(u_i)}{n_i} \qquad \ldots \quad (5.4.1)$$

and

$$E(L^2) = (t^2-1)^2 \left\{ \sum_{1}^{k} \frac{\lambda(u_i)}{n_i} \right\}^2 \left[ \frac{\sum_{1}^{k} \frac{\lambda^2(u_i)}{n_i^3} \left\{ B_{2i}-3 \right\}}{\left( \sum_{1}^{k} \frac{\lambda(u_i)}{n_i} \right)^2} + 1 + \frac{2}{(t^2-1)^2} \left\{ \frac{t^4 \sum_{1}^{k} \frac{\lambda^2(u_i)}{n_i^2(n_i-1)}}{\left( \sum_{1}^{k} \frac{\lambda(u_i)}{n_i} \right)^2} + 1 \right\} \right] \qquad \ldots \quad (5.4.2)$$

where

$$\lambda(u_i) = \sum_{j=1}^{N_i} p_{ij} \left\{ \frac{y_{ij}-R\,x_{ij}}{p_{ij}} - (M_{iy} - R\,M_{ix}) \right\}^2 \qquad \ldots \quad (5.4.3)$$

and

$$B_{2i} = \frac{\sum_{j=1}^{N_i} p_{ij} \left\{ \frac{y_{ij}-R\,x_{ij}}{p_{ij}} - (M_{iy} - R\,M_{ix}) \right\}^4}{\{\lambda(u_i)\}^2} \qquad \ldots \quad (5.4.4)$$

Hence from (1.1.1), (5.3.3), (5.4.1) and (5.4.2),

$$\text{prob.} \quad \{L \geqslant 0\} \geqslant P_0,$$

where

$$P_0 = \frac{\{E(L)\}^2}{E(L^2)} = \frac{1}{\frac{\sum_{1}^{k} \frac{\lambda^2(u_i)}{n_i3} (B_{2i}-3)}{\left( \sum_{1}^{k} \frac{\lambda(u_i)}{n_i} \right)^2} + 1 + \frac{2}{(t^2-1)^2} \left\{ \frac{t^4 \sum_{1}^{k} \frac{\lambda^2(u_i)}{n_i^2(n_i-1)}}{\left( \sum_{1}^{k} \frac{\lambda(u_i)}{n_i} \right)^2} + 1 \right\}} \qquad \ldots \quad (5.4.5)$$

5.5. We have from (5.3.3) and (5.4.5)

$$t^2 \sum_{1}^{k} \left\{ \frac{\lambda_{iz}}{n_i} + R^2 \frac{\lambda_{iw}}{n_i} - 2R \frac{r_i}{n_i} \sqrt{\lambda_{iz} \lambda_{iw}} \right\} \geqslant \left\{ \sum_{1}^{k} \bar{z}_i - R \sum_{1}^{k} \bar{w}_i \right\}^2 \qquad \ldots \quad (5.5.1)$$

with probability greater than (or equal to) $P_0$ where $P_0$ is given by (5.4.5). From (5.5.1) limits for $R$ can be derived on the same lines as discussed earlier for the case of a single stratum.

## 6. Extension to multistage designs

6.1. One stratum, two stage design, pps selection : Let there be a finite population consisting of $K$ first stage units, where the $i$-th unit contains $N_i$ second stage units. Let $y_{ij}$ denote value $y$ of the $j$-th second stage unit of the $i$-th first stage unit ($j = 1, 2, \ldots, N_i$, $i = 1, 2, \ldots k$). Let us consider a two stage sampling scheme where $n$ first stage units are selected with replacement from $k$ first units with probability proportional to some measure of the units. Let $p_1, p_2, \ldots p_k$ denote the probability for the different first stage unit to be selected in a particular draw. Within each selected first stage unit let us select $n_1$ or $n_2$ or ... $n_k$ second stage units (according as the selected first stage unit happens to be the 1st or 2nd——— or $k$-th first stage unit) with replacement with probability proportional to some measure of the second stage units. Let $p_{ij}$ denote the probability that the $j$-th second stage unit (of the $i$-th first stage unit) having variate value $y_{ij}$ will appear in a particular selection while sampling for second stage units after the $i$-th first stage unit has been selected ($j = 1, 2, \ldots N_i$, $i = 1, 2, \ldots k$). Oviously $\sum_{j=1}^{N_i} p_{ij} = 1$, for $i = 1, 2, \ldots k$

6.2. Let $z_s$ be an estimate of population total $M$ as built up from the $s$-th selected first stage unit. Obviously,

$$z_s = \frac{1}{p_i} \cdot \frac{1}{n_i} \sum_{l=1}^{n_i} \frac{y_i(l)}{p_i(l)}$$

if the $s$-th selected first stage unit happens to be the $i$-th first stage unit and and $y_{i(1)}$, $y_{i(2)}, \ldots y_{i(n_i)}$ happen to be $y$-values of $n_i$ selected second stage unit within the $i$-th first stage unit.

6.3. Let us define a function $L$ of estimators $z_1, z_2, z_3, \ldots z_n$, $M$ and $t^2 (t > 1)$ as

$$L = t^2 \frac{\sum_{s=1}^{n} (z_s - \bar{z})^2}{n(n-1)} - (\bar{z} - M)^2 \qquad \ldots \quad (6.3.1)$$

where

$$\bar{z} = \frac{\sum_{s=1}^{n} z_s}{n}$$

6.4. We have from (1.1.1) and (6.3.1)

$$\text{prob.} \quad \left\{ \bar{z} + t\sqrt{\frac{\Sigma(z_s-\bar{z})^2}{n(n-1)}} > M > \bar{z} - t\sqrt{\frac{(z_s-\bar{z})^2}{n(n-1)}} \right\}$$

$$\geqslant \frac{1}{\frac{B_2(z)-3}{n} + 1 + \frac{2}{(t^2-1)^2}\left\{ \frac{t^4}{n-1} + 1 \right\}} \qquad \ldots \quad (6.4.1)$$

where $B_2(z)$ for the scheme of sampling considered is given as $\frac{\mu_4(z)}{(\mu_2(z))^2}$, where

$$\mu_2(z) = \sum p_i \left( \frac{M_i}{p_i} - M \right)^2 + \sum \frac{1}{n_i p_i} \sum p_{ij} \left( \frac{y_{ij}}{p_{ij}} - M_i \right)^2$$

$$\mu_4(z) = \sum p_i \left( \frac{M_i}{p_i} - M \right)^4 + \sum \frac{1}{(n_i p_i)^3} \sum p_{ij} \left( \frac{y_{ij}}{p_{ij}} - M_i \right)^4 +$$

$$+3 \sum \frac{1}{(n_i p_i)^3} (n_i - 1) \left\{ \sum p_{ij} \left( \frac{y_{ij}}{p_{ij}} - M_i \right)^2 \right\}^2 +$$

$$+4 \sum \frac{1}{(n_i p_i)^2} \left( \frac{M_i}{p_i} - M \right) \sum p_{ij} \left( \frac{y_{ij}}{p_{ij}} - M_i \right)^3 +$$

$$+6 \sum \frac{1}{n_i p_i} \left( \frac{M_i}{p_i} - M \right)^2 \sum p_{ij} \left( \frac{y_{ij}}{p_{ij}} - M_i \right)^2.$$

when $M_i$ stands for total of $y$-values of the $i$-th first stage unit i.e. $M_i = \sum_{j=1}^{N_i} y_{ij}$, $(i = 1, 2, \ldots k)$.

## REFERENCES

BANERJEE, S. K. (1958) : A Lower Bound to the Probability of Student's Ratio. *Sankhyā*, 18, 391–394.

COCHRAN, W. G. (1953) : *Sampling Techniques*, John Wiley and Sons, New York.

FIELLER, E. C. (1940) : Biological Standardization of Insulin. *Suppl. J. Roy. Stat. Soc.*, 7, 1-64.

——— (1954) : Symposium on interval estimation. *J. Roy. Stat. Soc.*, B, 16, 175-222.

*Paper received : August, 1957.*