ON ERRORS OF ESTIMATES IN VARIOUS TYPES OF DOUBLE SAMPLING PROCEDURE*

By K. C. SEAL
Statistical Laboratory, Calcutta

1. Introduction

The term double sampling has come to be applied to any sampling technique which involves two sampling investigations. As in many other kinds of sampling, reduction in cost and increase in accuracy are also the main advantages of this type of sampling. Neyman (1938) has given a method of sampling in which a first large sample of an auxiliary character is used to subdivide the population into groups in which a second (main) character varies little, so that if the characters are correlated, better estimates of the main character can be obtained from a rather small second sample.

Another familiar kind of double sampling is that in which a first sample of size n taken for both the characters is used to determine the regression of the main character y on the other x and a second sample of size N, observed only for the auxiliary character x is used to obtain an estimate of the main character y. This procedure is particularly applicable to situations in which the enumeration of the main character involves too much cost but an auxiliary variate correlated to it can be easily measured. Cochran (1939) has given examples of this type of sampling Formulae for variance of the estimates assuming linear regression have been given by Snedecor and King (1942) and C. Bose (1943). An expression for variance of estimates for a particular type of non-linear regression with one auxiliary variate was derived by Bose and Gayen (1946).

It might however be possible to increase the precision of the estimate in this kind of double sampling by including instead of one, a number of correlated auxiliary variates. B. Ghosh (1947) has indicated that for linear regression and for random sampling, the estimate is unbiased and has obtained an approximate formula for the variance of estimate based on many auxiliary variates.

Double sampling, as has been stated before whether with a single auxiliary variate or with many auxiliary variates is a technique in itself for increasing the accuracy of the sample. It may further be possible to couple with it other methods of sampling known for enhancing accuracy. Thus different modes can be adopted of choice of sample units for the first and second samples. For example, the first sample may be random and second systematic. Considerations of cost and of accuracy would again warrant using for the first sample sometimes a predetermined and sometimes a specially chosen set of values.

---

* Presented at the 38th Session of the Indian Science Congress, Bangalore, January, 1951.

In this paper expressions for estimates and for variances of the estimates have been obtained corresponding to various types of double sampling. Types with single auxiliary variate and with many auxiliary variates have been separately dealt with. In most cases linear regression has been assumed. Two instances have been considered assuming non-linear regression. A modified case has been studied in which the expected value of the auxiliary variable in the first sample is a constant multiple of the expected value of the auxiliary variate in the other sample.

The problem of optimum allocation of sample units in the first and second samples has been touched upon by deriving the optimum numbers for one of the types following the line adopted by Schumacher and Chapman (1942).

In the Appendices certain interesting results, which are believed to be new, viz., the joint distribution of regression coefficients, and the expected value of the typical element in the inverse of the sample dispersion matrix, have been derived for the multivariate normal population. The already known distribution of the partial regression coefficient has also been derived by an alternative method using rectangular coordinates.

2. MULTIVARIATE AUXILIARY SET—DIFFERENT SITUATIONS OF SAMPLING

(2.1) x_n—*Random*, y_n—*Random*, x_N—*Random*: When the auxiliary variates consist of a set $x_1, x_2, \ldots, x_k$, assuming *linear* regression, the estimate of the population mean value of *y for random* sampling is given by the relation

$$Y = \bar{y}_n + \sum_{i=1}^{k} b_{ni} (\bar{x}_{Ni} - \bar{x}_{ni}) \qquad \ldots \quad (2.11)$$

where $\bar{y}_n$, b_{ni} and $\bar{x}_{ni}$ are derived from the first sample and $\bar{x}_{Ni}$ from the second. Assuming a *multivariate normal population* and letting

$$E(y) = \eta, \qquad E(x_i) = \xi_i (i = 1, 2, \ldots, k)$$
$$V(y) = \sigma_y^2, \quad V(x_i) = \sigma_{ii} = \sigma_i^2$$
$$\text{cov}\,(x_i, x_j) = \sigma_{ij} = \zeta_{ij}\, \sigma_i \sigma_j$$
$$\text{cov}\,(y, x_i) = \sigma_{yi} = \zeta_{yi}\, \sigma_y \sigma_i$$
$$E(b_{ni}) = \beta_{ni} = -\frac{R_{yi}}{R_{yy}} \frac{\sigma_y}{\sigma_i},$$

R_{ij} being the cofactor of ζ_{ij} in the determinant $|\zeta_{ij}|$ $(i, j = y, 1, 2, \ldots, k)$, the following results are readily derived by the use of results in Appendix A.

$$V(b_{ni}) = \sigma^2_{y \cdot x} E(c_{ii}) = \frac{\sigma^2_{y \cdot x}\, \sigma^{ii}}{n-k-2} = \frac{\sigma^2_y (1 - R^2_{y \cdot 12 \cdots k}) \sigma^{ii}}{n-k-2}$$

$$\text{cov}(b_{ni}, b_{nj}) = \sigma^2_{y \cdot x} E(c_{ij}) = \frac{\sigma^2_{y \cdot x} \sigma^{ij}}{n-k-2} = \frac{\sigma^2_y (1 - R^2_{y \cdot 12 \cdots k}) \sigma^{ij}}{n-k-2}$$

where $\sigma^2_{y\cdot x}$ represents the residual variance of y when x's are kept fixed, $R_{y\cdot 12\cdots k}$ the multiple correlation coefficient between y and $x_1, x_2, \ldots, x_k$ and the matrix $((\sigma^{ij}))$ is the reciprocal matrix of $((\sigma_{ij}))$.

Thus for the case of double sampling in which the x's and y's in the first sample and the x's in the second sample are all chosen at random we have,

$$E(Y) = \eta + \sum_{i=1}^{k} \beta_{ni}(\xi_i - \xi_i)$$

$$= \eta \qquad \ldots \quad (2.12)$$

$$V(Y) = E\left\{(\bar{y}_n - \eta) + \sum_{i=1}^{k} b_{ni}(\bar{x}_{Ni} - \xi_i) - \sum_{i=1}^{k} b_{ni}(\bar{x}_{ni} - \xi_i)\right\}^2$$

$$= \frac{\sigma^2_y}{n} + \left[\sum_{i=1}^{k}\sum_{j=1}^{k} \sigma_{ij}\left\{\beta_{ni}\ \beta_{nj} + \frac{\sigma^2_{y\cdot x}\sigma^{ij}}{n-k-2}\right\}\right] \times \left(\frac{1}{n} + \frac{1}{N}\right) - 2\sum_{i=1}^{k} \beta_{ni}\frac{\sigma_{yi}}{n}$$

$$= \frac{\sigma^2_y}{n} + \left(\frac{1}{n} + \frac{1}{N}\right)\sum_{i=1}^{k}\sum_{j=1}^{k} \beta_{ni}\beta_{nj}\sigma_{ij}$$

$$+ \frac{\sigma^2_y(1 - R^2_{y\cdot 12\cdots k})}{n-k-2} \cdot \left(\frac{1}{n} + \frac{1}{N}\right)\sum_{i=1}^{k}\sum_{j=1}^{k} \sigma_{ij}\sigma^{ij} - 2\sum_{i=1}^{k}\beta_{ni}\frac{\sigma_{yi}}{n}$$

Now $\quad \sum_{j=1}^{k} \sigma_{ij}\sigma^{ij} = 1 \qquad \therefore \quad \sum_{i=1}^{k}\sum_{j=1}^{k} \sigma_{ij}\sigma^{ij} = k$

$$\sum_{j=1}^{k} \beta_{nj}\sigma_{ij} = \sigma_{yi}$$

$$\sum_{i=1}^{k}\sum_{j=1}^{k} \beta_{ni}\beta_{nj}\sigma_{ij} = \sum_{i=1}^{k} \beta_{ni}\sigma_{yi} = \sum_{i=1}^{k} -\frac{R_{yi}}{R_{yy}}\ \frac{\sigma_y}{\sigma_i}\zeta_{yi}\sigma_i\sigma_y$$

$$= \frac{-\left(\sum_{i=1}^{k} \zeta_{yi}R_{yi}\right)}{R_{yy}}\ \sigma^2_y = \left(\frac{R - R_{yy}}{R_{yy}}\right)\sigma^2_y$$

$$= \sigma^2_y R^2_{y\cdot 12\cdots k}$$

Hence,

$$V(Y) = \frac{\sigma^2_y}{n} + \frac{k}{n-k-2}\ \sigma^2_y(1 - R^2_{y\cdot 12\cdots k})\left[\frac{1}{n} + \frac{1}{N}\right]$$

$$+ \sigma^2_y R^2_{y\cdot 12\cdots k}\left[\frac{1}{N} - \frac{1}{n}\right]$$

$$= \sigma^2_y(1 - R^2_{y\cdot 12\cdots k})\left[\frac{1}{n} + \frac{k}{n-k-2}\left(\frac{1}{n} + \frac{1}{N}\right)\right] + \frac{\sigma^2_y R^2_{y\cdot 12\cdots k}}{N} \qquad \ldots \quad (2.13)$$

We note that the equation (2.13) is independent of the variances σ_{ij}'s of the x's.

(2.2) x_n — *Random*, y_n — *Random*, x_N — *Systematic*: Considering now a more general case, *viz.*, when x's and y's in the *first* sample are chosen at *random* but in the *second* sample are chosen in a *systematic* manner, we have for the estimate Y the same relation as in (2.11).

$$E(Y) = \eta \qquad \ldots \quad (2.21)$$

$$E\left((\bar{x}_{Ni}-\xi_i)(\bar{x}_{Nj}-\xi_j)\right) = \frac{1}{N^2} E\left\{ \sum_{p=1}^{N} \sum_{q=1}^{N} (x_{Ni_p}-\xi_i)(x_{Nj_q}-\xi_j)\right\}$$

$$= \frac{\sum_{p=1}^{N} \sum_{q=1}^{N} \zeta^{(xx)}_{i_p j_q} \sigma_i \sigma_j}{N^2}$$

where $\zeta^{(xx)}_{i_p j_q}$ $(p = 1,\ldots,N; q = 1, \ldots N)$ stands for the correlation between x_{Ni_p} and x_{Nj_q} in the population.

$$V(Y) = E\left((\bar{y}_n-\eta) + \sum_{i=1}^{k} b_{ni}(\bar{x}_{Ni}-\xi_i) - \sum_{i=1}^{k} b_{ni}(\bar{x}_{ni}-\xi_i)\right)^2$$

$$= \frac{\sigma^2_y}{n} + \sum_{i=1}^{k} \sum_{j=1}^{k} \left[\beta_{ni}\beta_{nj} + \frac{\sigma^2_{y\cdot x}\sigma^{ij}}{n-k-2} \right] \left[\frac{\sum_{p=1}^{N} \sum_{q=1}^{N} \zeta^{(xx)}_{i_p j_q} \sigma_i \sigma_j}{N^2} + \frac{\sigma_{ij}}{n} \right] - 2 \sum_{i=1}^{k} \beta_{ni} \frac{\sigma_{yi}}{n}$$

$$= \frac{\sigma^2_y}{n} + \sum_{i=1}^{k} \sum_{j=1}^{k} \left[\beta_{ni}\beta_{nj} + \frac{\sigma^2_y(1-R^2_{y\cdot 12\cdots k})\sigma^{ij}}{n-k-2} \right] \times \left\{ \frac{\sum_{p=1}^{N} \sum_{q=1}^{N} \zeta^{(xx)}_{i_p j_q} \sigma_i \sigma_j}{N^2} \right\} + \frac{k\sigma^2_y(1-R^2_{y\cdot 12\cdots k})}{n(n-k-2)} - \frac{\sigma^2_y R^2_{y\cdot 12\cdots k}}{n}$$

$$= \frac{\sigma^2_y(1-R^2_{y\cdot 12\cdots k})}{n} \cdot \frac{n-2}{n-k-2} + \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{\sum_{p=1}^{N} \sum_{q=1}^{N} \zeta^{(xx)}_{i_p j_q} \sigma_i \sigma_j}{N^2} \times \left[\beta_{ni}\beta_{nj} + \frac{\sigma^2_y(1-R^2_{y\cdot 12\cdots k})}{n-k-2} \sigma^{ij} \right] \qquad \ldots \quad (2.22)$$

It is easily seen that the case considered in 2.1 is a particular case of this in which

$$\zeta^{(xx)}_{i_p j_q} = \zeta_{ij} \quad \text{for} \quad p = q$$

$$= 0 \quad \text{for} \quad p \neq q$$

(2.3) x_n—*Fixed*, y_n—*Random*, x_N—*Random*: Next we consider a third case where x's in the *first* sample are *fixed*, but x's in the *second* sample and y's in the *first* sample are chosen at *random*. Here again the estimate of the population mean value of y remains the same as (2.11). Assuming,

$$E(y_{np}) = \alpha_n + \sum_{i=1}^{k} \beta_{ni} x_{nip} \quad (p = 1, 2, \ldots, n)$$

so that

$$E(\bar{y}_n) = \alpha_n + \sum_{i=1}^{k} \beta_{ni} \bar{x}_{ni} \equiv \eta' \text{ (say)}$$

Hence

$$E(Y) = \eta' + \sum_{i=1}^{k} \beta_{ni}(\xi_i - \bar{x}_{ni}) = \eta \quad \ldots \quad (2.31)$$

$$V(Y) = E\{(\bar{y}_n - \eta) - \sum_{i=1}^{k} (\bar{x}_{ni} - \xi_i)(b_{ni} - \beta_{ni}) + \sum_{i=1}^{k} b_{ni}(\bar{x}_{Ni} - \xi_i)\}^2$$

$$= A + \frac{B}{n} + \frac{C}{N} \quad \ldots \quad (2.32)$$

where

$$A = \sigma^2_{y \cdot x} \left\{ \sum_{i=1}^{k} \sum_{j=1}^{k} (\bar{x}_{ni} - \xi_i)(\bar{x}_{nj} - \xi_j) c_{ij} \right\}$$

$$B = \sigma^2_{y \cdot x}$$

$$C = \sum_{i=1}^{k} \sum_{j=1}^{k} \sigma_{ij} \left\{ c_{ij} \sigma^2_{y \cdot x} + \beta_{ni} \beta_{nj} \right\}$$

Corollary: *If the total cost T is of the form:*

$$T = \alpha + \beta n + \gamma N \quad \ldots \quad (2.33)$$

where α, β and γ are parameters estimated from data, the optimum values of n and N for a given cost T with minimum variance should satisfy the equations

$$\frac{\partial V}{\partial n} + \lambda \frac{\partial T}{\partial n} = 0$$

$$\frac{\partial V}{\partial N} + \lambda \frac{\partial T}{\partial N} = 0$$

or,

$$\frac{\beta n}{\sqrt{B\beta}} = \frac{\gamma N}{\sqrt{C\gamma}} = \frac{T - \alpha}{\sqrt{B\beta} + \sqrt{C\gamma}}$$

so that

$$n = \frac{T - \alpha}{\beta} \frac{\sqrt{B\beta}}{\sqrt{B\beta} + \sqrt{C\gamma}}$$

and

$$N = \frac{T - \alpha}{\gamma} \frac{\sqrt{C\gamma}}{\sqrt{B\beta} + \sqrt{C\gamma}} \quad \ldots \quad (2.34)$$

(2.4) x_n—*Fixed*, y_n—*Random*, x_N—*Systematic*: Let us now take a fourth case in which x's are *fixed* in the *first* sample, but y's are *random* and x's in the *second* sample are *correlated*. Here again (2.11) holding good,

$$E(Y) = \eta + \sum_{i=1}^{k} \beta_{ni}(\xi_i - \bar{x}_{ni}) = \eta \qquad \dots \quad (2.41)$$

$$V(Y) = E\Big\{(\bar{y}_n - \eta) - \sum_{i=1}^{k} (\bar{x}_{ni} - \xi_i)(b_{ni} - \beta_{ni}) + \sum_{i=1}^{k} b_{ni}(\bar{x}_{Ni} - \xi_i)\Big\}^2$$

$$= \frac{\sigma^2_{y\cdot x}}{n} + \sum_{i=1}^{k}\sum_{j=1}^{k} E\left\{\frac{\sum_{p=1}^{N}\sum_{q=1}^{N}(x_{Ni_p} - \xi_i)(x_{Nj_q} - \xi_j)}{N^2}\right\} E(b_{ni}b_{nj})$$

$$+ \sum_{i=1}^{k}\sum_{j=1}^{k} (\bar{x}_{ni} - \xi_i)(\bar{x}_{nj} - \xi_j)c_{ij}\sigma^2_{y\cdot x}$$

$$= \frac{\sigma^2_{y\cdot x}}{n} + \frac{1}{N^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\Big[\beta_{ni}\beta_{nj} + c_{ij}\sigma^2_{y\cdot x}\Big]\left\{\sum_{p=1}^{N}\sum_{q=1}^{N}\zeta^{(xx)}_{i_p j_q}\,\sigma_i\sigma_j\right\}$$

$$+ \sum_{i=1}^{k}\sum_{j=1}^{k} (\bar{x}_{ni} - \xi_i)(\bar{x}_{nj} - \xi_j)c_{ij}\sigma^2_{y\cdot x} \qquad \dots \quad (2.42)$$

From (2.42) it is inferred that $V(Y)$ will be smaller according as x's in the first sample are wider apart, but having the means $\bar{x}_{ni}$ as near to the population means $\xi_i (i=1, 2, \dots, k)$ as possible.

Obviously, the result (2.32) follows from (2.42) by putting

$$\zeta^{(xx)}_{i_p i_q} = \zeta_{ij} \quad \text{for} \quad p = q$$

$$= 0 \quad \text{for} \quad p \neq q$$

3. Single auxiliary variate—different situations of sampling

(3.1) x_n—*Fixed*, y_n—*Systematic*, x_N—*Random*: We now consider certain special cases of double sampling when there is only one auxiliary variate. Let x's be fixed in the first sample but the y's correlated and x's chosen randomly in the second sample.

$$E(x_{Ni}) = \xi \;(i = 1, 2, \dots, N)$$

$$E(y_{ni}) = \alpha_n + \beta_n x_{ni} \;(i = 1, 2, \dots, n)$$

$$V(y_{ni}|x_{ni}) = \sigma^2_{y\cdot x}$$

$$\text{cov}\,(y_{ni}, y_{nj}|x_{ni}, x_{nj}) = \zeta_{ij}^{(yy)}\sigma^2_{y\cdot x}$$

We have,

$$Y = \bar{y}_n + b_n(\bar{x}_N - \bar{x}_n) \quad \dots \quad (3.11)$$

$$E(b_n) = E\left\{\frac{\sum_{i=1}^{n}(x_{ni}-\bar{x}_n)(y_{ni}-\bar{y}_n)}{\sum_{i=1}^{n}(x_{ni}-\bar{x}_n)^2}\right\}$$

$$= \beta_n$$

$$E(Y) = (\alpha_n+\beta_n\bar{x}_n)+\beta_n(\xi-\bar{x}_n)$$

$$= \alpha_n+\beta_n\xi \quad \dots \quad (3.12)$$

Further,

$$V(b_n) = \frac{\sigma^2_{y\cdot x}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ni}-\bar{x}_n)(x_{nj}-\bar{x}_n)\zeta_{ij}^{(yy)}}{\left[\sum_{i=1}^{n}(x_{ni}-\bar{x}_n)^2\right]^2}$$

$$\text{cov}(b_n, \bar{y}_n) = \frac{1}{n}E\left\{\frac{\sum_{i=1}^{n}y_{ni}(x_{ni}-\bar{x}_n)\sum_{i=1}^{n}(y_{ni}-\alpha_n-\beta_n x_{ni})}{\sum_{i=1}^{n}(x_{ni}-\bar{x}_n)^2}\right\}$$

$$= \frac{\sigma^2_{y\cdot x}\sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_{ij}^{(yy)}(x_{ni}-\bar{x}_n)}{n\sum_{i=1}^{n}(x_{ni}-\bar{x}_n)^2}$$

$$V(\bar{y}_n) = \frac{\sigma^2_{y\cdot x}}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_{ij}^{(yy)}$$

We then have,

$$V(Y) = E\{(\bar{y}_n-\alpha_n-\beta_n\bar{x}_n)+b_n(\bar{x}_N-\xi)+(b_n-\beta_n)(\xi-\bar{x}_n)\}^2$$

$$= \sigma^2_{y\cdot x}\left[\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_{ij}^{(yy)}}{n^2}+\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_{ij}^{(yy)}(x_{ni}-\bar{x}_n)(x_{nj}-\bar{x}_n)}{\left[\sum_{i=1}^{n}(x_{ni}-\bar{x}_n)^2\right]^2}\right.$$

$$\left.\times\left\{\frac{\sigma_x^2}{N}+(\xi-\bar{x}_n)^2\right\}+\frac{2(\xi-\bar{x}_n)\sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_{ij}^{(yy)}(x_{ni}-\bar{x}_n)}{n\sum_{i=1}^{n}(x_{ni}-\bar{x}_n)^2}\right]+\frac{\sigma_x^2\beta_n^2}{N} \quad \dots \quad (3.13)$$

A particular case of (3.13), *viz.*, when

$$\zeta_{ii}^{(yy)} = 1 \quad (i = 1, 2, ..., n)$$
$$\zeta_{ij}^{(yy)} = 0 \quad (i \neq j)$$

is (2.32) when $k = 1$.

In order to get an estimate of $V(Y)$, it is obvious that some assumptions will have to be made about the correlation coefficient $\zeta_{ij}^{(yy)}$, such as $\zeta_{ij} = \zeta_u$ for $|i-j| = u$.

(3.2) x_n — *Fixed*, y_n — *Systematic*, x_N — *Systematic*: We now consider the situation in which x's are *fixed* in the *first* sample but x's in the *second* sample as also in the *first* sample are *correlated*.

Here also equations (3.11) and (3.12) holding good and assuming

$$\text{cov}\,(x_{Ni}, x_{Nj}) = \zeta_{ij}^{(xx)}\sigma_x^2 \; (i, j = 1, 2, ..., N)$$

we have,

$$V(Y) = \sigma^2_{y \cdot x}\left[\frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}\zeta_{ij}^{(yy)}}{n^2} + \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}\zeta_{ij}^{(yy)}(x_{ni}-\bar{x}_n)(x_{nj}-\bar{x}_n)}{\left[\sum\limits_{i=1}^{n}(x_{ni}-\bar{x}_n)^2\right]^2}\right.$$

$$\times\left\{\frac{\sigma_x^2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\zeta_{ij}^{(xx)}+(\xi-\bar{x}_n)^2\right\}+\frac{2(\xi-\bar{x}_n)}{\sum\limits_{i=1}^{n}(x_{ni}-\bar{x}_n)^2}$$

$$\left.\times\frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}\zeta_{ij}^{(yy)}(x_{ni}-\bar{x}_n)}{n}\right]+\frac{\sigma_x^2\beta_n^2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\zeta_{ij}^{(xx)} \qquad \ldots \quad (3.21)$$

If

$$\zeta_{ii}^{(xx)} = 1 \;\; (i = 1, 2, ..., N)$$
$$\zeta_{ij}^{(xx)} = 0 \quad (i \neq j)$$

this formula reduces to (3.13).

(3.3) x_n — *Systematic*, y_n — *Systematic*, x_N — *Systematic*: Let us now take the more general case, in which x's and y's in the *first* sample and x's in the *second* sample all are *correlated* among themselves. In this case even the approximate formulæ for expectation and variance of Y are complicated. Here also (3.11) holds good.

Assuming a multivariate normal distribution for x's and y's in the first sample and letting

$$E(x_{ni}) = E(x_{Ni}) = \xi$$

$$E(y_{ni}) = \eta$$

$$\left.\begin{array}{l} \mathrm{cov}\,(x_{ni}, x_{nj}) = \zeta_{ij}^{(xx)}\sigma_x^2 \\ \mathrm{cov}\,(y_{ni}, y_{nj}) = \zeta_{ij}^{(yy)}\sigma_y^2 \\ \mathrm{cov}\,(x_{ni}, y_{nj}) = \zeta_{ij}^{(xy)}\sigma_x\sigma_y \end{array}\right\} \quad (i, j = 1, 2, .., n)$$

$$E(x_{ni}-\xi)(x_{nj}-\xi)(y_{nk}-\eta) = 0 \quad (i, j, k = 1, 2, ..., n)$$

so that

$$E(x_{ni}x_{nj}y_{nk}) = \{\xi^2+\zeta_{ij}^{(xx)}\sigma_x^2\}\eta+\xi\sigma_x\sigma_y\{\zeta_{ik}^{(xy)}+\zeta_{jk}^{(xy)}\},$$

we have,

$$E(b_n) \simeq \frac{E\left\{\sum_{i=1}^{n} x_{ni}y_{ni} - \frac{\sum_{i=1}^{n} x_{ni} \sum_{i=1}^{n} y_{ni}}{n}\right\}}{E\left\{\sum_{i=1}^{n} x^2_{ni} - \frac{\left(\sum_{i=1}^{n} x_{ni}\right)^2}{n}\right\}}$$

$$\simeq \frac{\sigma_y}{\sigma_x} \cdot \frac{\left\{n(n-1)\zeta - \sum_{i\neq j=1}^{n} \zeta_{ij}^{(xy)}\right\}}{\left\{n(n-1) - \sum_{i\neq j=1}^{n} \zeta_{ij}^{(xx)}\right\}} = \beta_n(say)$$

where

$$\zeta = \zeta_{ii}^{(xy)} \; (i=1,2,...,n)$$

Further,

$$E\left\{\sum_{i=1}^{n} x_{ni} \sum_{i=1}^{n} x_{ni}\, y_{ni}\right\} = \sum_{i=1}^{n} \sum_{j=1}^{n} E(x_{ni}\, y_{ni}\, x_{nj})$$

$$= n^2\xi^2\eta + \sum_{i=1}^{n} \sum_{j=1}^{n} \zeta_{ij}^{(xx)}\, \sigma_x^2\eta + n^2\zeta\xi\sigma_x\sigma_y + \sum_{i=1}^{n} \sum_{j=1}^{n} \zeta_{ij}^{(xy)}\, \xi\sigma_x\sigma_y$$

$$E\left\{\left(\sum_{i=1}^{n} x_{ni}\right)^2 \sum_{i=1}^{n} y_{ni}\right\} = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} E\,(x_{ni}\, x_{nj}\, y_{nk})$$

$$= n^3\xi^2\eta + n \sum_{i=1}^{n} \sum_{j=1}^{n} \zeta_{ij}^{(xx)}\, \sigma^2_x\eta + 2n \sum_{i=1}^{n} \sum_{j=1}^{n} \zeta_{ij}^{(xy)}\, \xi\sigma_x\sigma_y$$

Hence,

$$E(b_n\bar{x}_n) \simeq \frac{E\left[\sum_{i=1}^{n} x_{ni}y_{ni} \sum_{i=1}^{n} x_{ni} - \frac{\left(\sum_{i=1}^{n} x_{ni}\right)^2\left(\sum_{i=1}^{n} y_{ni}\right)}{n}\right]}{n\, E\left[\sum_{i=1}^{n} x^2_{ni} - \frac{\left(\sum_{i=1}^{n} x_{ni}\right)^2}{n}\right]}$$

$$\simeq \xi\frac{\sigma_y}{\sigma_x}\frac{\left\{n(n-1)\zeta - \sum_{i\neq j=1}^{n} \zeta^{(xy)}_{ij}\right\}}{\left\{n(n-1) - \sum_{i\neq j=1}^{n} \zeta^{(xx)}_{ij}\right\}}$$

$$\simeq E(\bar{x}_n)\, E(b_n)$$

$$\therefore \quad E(Y) \simeq \eta+\beta_u(\xi-\xi) = \eta \qquad \dots \quad (3.31)$$

Again,

$$E(b^2_n) \simeq \frac{E\left\{\sum_{i=1}^{n}(x_{ni}-\xi)(y_{ni}-\eta)\right\}^2}{E\left\{\sum_{i=1}^{n}(x_{ni}-\xi)^2\right\}^2}$$

$$\simeq \frac{E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} x'_{ni}\, x'_{nj}\, y'_{ni}\, y'_{nj}\right\}}{E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} x'^2_{ni}\, x'^2_{nj}\right\}}$$

Where x'_n, y'_n stand for the difference of the corresponding variate from its mean.

For $(i\neq j)$ $E(x'_{ni}\, x'_{nj}\, y'_{ni}\, y'_{nj})$ can be readily found out from the moment-generating function of the four variables (the distribution of which has been assumed to be multivariate normal) and is given by

$$E\,(x'_{ni}\, x'_{nj}\, y'_{ni}\, y'_{nj})$$

$$= \sigma^2_x\, \sigma^2_y\left\{ \zeta^{(xx)}_{ij}\, \zeta^{(yy)}_{ij} + \zeta^{(xy)}_{ij}\, \zeta^{(xy)}_{ji} + \zeta^2 \right\} \quad (i\neq j)$$

Similarly, for $(i=j)$

$$E\,(x'_{ni}\, x'_{nj}\, y'_{ni}\, y'_{nj}) = E\,(x'^2_{ni}\, y'^2_{ni})$$

$$= \sigma_x^2\sigma_y^2\,(1+2\zeta^2)$$

$$E\,(x'^2_{ni}\, x'^2_{nj}) = \sigma_x^4\left\{1 + 2\zeta^{(xx)^2}_{ij}\right\}$$

$$E\,(x'^4_{ni}) = 3\sigma_x^4$$

Hence,

$$E\,(b_n^2) \sim \frac{\sigma_y^2}{\sigma_x^2}\,\frac{\left\{n+n(n+1)\zeta^2+\sum\limits_{i\neq j=1}^{n}\zeta_{ij}^{(xx)}\zeta_{ij}^{(yy)}+\sum\limits_{i\neq j=1}^{n}\zeta_{ij}^{(xy)}\zeta_{ji}^{(xy)}\right\}}{\left\{n(n+2)+2\sum\limits_{i\neq j=1}^{n}\zeta_{ij}^{(xx)}\right\}^2}$$

$$\sim \beta'^2_n(\text{say})$$

$$\therefore\quad V(Y) \simeq E\{(\bar{y}_n-\eta)+b_n(\bar{x}_N-\xi)-b_n(\bar{x}_n-\xi)\}^2$$

$$\simeq E(\bar{y}_n-\eta)^2+\left(E(\bar{x}_N-\xi)^2+E(\bar{x}_n-\xi)^2\right)E(b_n^2)-2E(b_n)E(\bar{x}_n-\xi)(\bar{y}_n-\eta)$$

$$\simeq \frac{\sigma_y^2}{n^2}\left\{n+\sum_{i\neq j=1}^{n}\zeta_{ij}^{(yy)}\right\}-\frac{2\beta_n\sigma_x\sigma_y}{n^2}\left\{n\zeta+\sum_{i\neq j=1}^{n}\zeta_{ij}^{(xy)}\right\}$$

$$+\beta'^2_n\sigma_x^2\left\{\frac{1}{N}+\frac{1}{n}+\frac{1}{N^2}\sum_{i\neq j=1}^{n}\zeta_{ij}^{(xx')}+\frac{1}{n^2}\sum_{i\neq j=1}^{n}\zeta_{ij}^{(xx)}\right\} \quad \ldots \quad (3.32)$$

(3.4) *x_n—Random, y_n—Random, x_N—Random, assuming $E(\bar{x}_n)=\lambda E(\;)$*: We now turn to a different set up. In the *first* sample x's and y's are both *random*, so also x's in the *second* sample;
but

$$E(\bar{x}_n) = \lambda E(\bar{x}_N) = \xi$$

Let

$$E(\bar{y}_n) = \eta$$
$$V(x_n) = \sigma^2_{xn},\; V(x_N) = \sigma^2_{xN}$$
$$V(y_n) = \sigma^2_{yn},\; \text{cov}\,(x_n y_n) = \zeta x_n y_n \sigma_{xn}\sigma_{yn}$$

To give an instance from practice of this set up we may regard x to be the green weight of a crop yield and y the corresponding dry weight; and x and y to be both measured in the first sample for a particular structure of sampling unit while x in the the second sample is measured for another structure of sampling unit.

In this case an unbiased estimate of the population mean η of y will be given by

$$Y = \bar{y}_n+b_n(\lambda\bar{x}_N-\bar{x}_n)$$
$$E(Y) = \eta+\beta(\xi-\xi) = \eta \quad \ldots \quad (3.41)$$

$$V(Y) = E\left((\bar{y}_n-\eta)+b_n\lambda\left(\bar{x}_N-\frac{\xi}{\lambda}\right)-b_n(\bar{x}_n-\xi)\right)^2$$

$$= \frac{\sigma_{yn}^2}{n}+\left(\frac{\lambda^2\sigma^2_{xN}}{N}+\frac{\sigma^2_{xn}}{n}\right)\frac{\sigma^2_{yn}}{\sigma^2_{xn}}\left\{\frac{1-\zeta^2_{x_n y_n}}{n-3}+\zeta^2_{x_n y_n}\right\}-\frac{2\zeta^2_{x_n y_n}\sigma^2_{yn}}{n}$$

$$= \frac{\sigma^2_{yn}(1-2\zeta^2_{x_n y_n})}{n}+\frac{\sigma^2_{yn}}{\sigma^2_{xn}}\left\{\frac{\lambda^2\sigma^2_{xN}}{N}+\frac{\sigma^2_{xn}}{n}\right\}\left\{\frac{1+(n-4)\zeta^2_{x_n y_n}}{n-3}\right\} \quad \ldots \quad (3.42)$$

If now $\sigma_{x_n}^2 = \mu^2\sigma_{x_N}^2$, (3.42) reduces to

$$V(Y) = \frac{\sigma^2_{y_n}}{n}(1-2\zeta^2_{x_n y_n})+\sigma^2_{y_n}\left\{\frac{\lambda^2}{\mu^2}\cdot\frac{1}{N}+\frac{1}{n}\right\}\left\{\frac{1+(n-4)\zeta^2_{x_n y_n}}{n-3}\right\} \quad \ldots \quad (3.43)$$

If, further, the coefficient of variations of x's in the first and second samples are equal *i.e.*, when $\lambda = \mu$, (3.43) reduces to

$$V(Y) = \frac{\sigma^2_{y_n}}{n}\cdot\frac{n-2}{n-3}(1-\zeta^2_{x_n y_n})+\frac{\sigma^2_{y_n}}{N(n-3)}\left[1+\zeta^2_{x_n y_n}(n-4)\right] \quad \ldots \quad (3.44)$$

It is interesting to compare (3.44) with (2.13) when $k = 1$.

(3.5) *x_n—Fixed, y_n—Random, x_N—Random, for non-linear regression*: Hitherto we have dealt with cases in which a linear relation between x and y was assumed to hold good. Taking the relationship to be *non-linear* and confining to a *second degree parabola* we may write

$$y = \alpha+\beta(x-\bar{x}_n)+\gamma(x^2-\overline{x_n^2})$$

where $\bar{x}_n$ and $\bar{x}_n^2$ denote the means of x and x^2's in the first sample *i.e.*

$$\bar{x}_n = \frac{\sum_{i=1}^{n} x_{ni}}{n} \qquad \text{and} \qquad \overline{x_n^2} = \frac{\sum_{i=1}^{n} x_{ni}^2}{n}$$

The normal equations for estimating α, β, γ will then be

$$\bar{y}_n = a_n$$

$$\Sigma(x-\bar{x}_n)y = b_n\Sigma(x-\bar{x}_n)^2+c_n\Sigma(x-\bar{x}_n)(x^2-\overline{x_n^2})$$

$$\Sigma(x^2-\bar{x}_n^2)y = b_n\Sigma(x-\bar{x}_n)(x^2-\overline{x_n^2})+c_n\Sigma(x^2-\overline{x_n^2})^2$$

where a_n, b_n and c_n are the estimates of α, β and γ respectively.

From above we may write

$$b_n = c_{11}\Sigma(x-\bar{x}_n)y+c_{12}\Sigma(x^2-\bar{x}_n^2)y$$

$$c_n = c_{21}\Sigma(x-\bar{x}_n)y+c_{22}\Sigma(x^2-\bar{x}_n^2)y$$

where c_{11}, c_{12}, c_{22} can be readily found out.

$$V(a_n) = \frac{\sigma^2_{y\cdot x}}{n},\ \text{cov}\,(a_n, b_n) = \text{cov}\,(a_n, c_n) = 0$$

$$V(b_n) = c_{11}\sigma^2_{y\cdot x},\ \text{cov}\,(b_n, c_n) = c_{12}\sigma^2_{y\cdot x}$$

$$V(c_n) = c_{22}\sigma^2_{y\cdot x},$$

when x's in the first sample are assumed to be fixed but y's in the first sample and x's in the second sample are assumed to be random. In this case the estimate of the population mean of y will be given by

$$Y = \bar{y}_n+b_n(\bar{x}_N-\bar{x}_n)+c_n(\overline{x_N^2}-\overline{x_n^2}) \quad \ldots \quad (3.51)$$

where $\bar{x}_N$ and $\overline{x_N^2}$ have similar meanings to $\bar{x}_n$ and $\overline{x_n^2}$.

Let
$$E(\bar{x}_N) = \mu_1', \; E(\overline{x_N^2}) = \mu_2'$$
$$E(Y) = \alpha+\beta(\mu_1'-\bar{x}_n)+\gamma(\mu_2'-\overline{x_n^2}) = \eta \qquad \ldots \quad (3.52)$$

and

$$V(Y) = E[(\bar{y}_n-\alpha)+(b_n-\beta)(\bar{x}_N-\mu_1')-(b_n-\beta)(\bar{x}_n-\mu_1')$$
$$+\beta(\bar{x}_N-\mu_1')+(c_n-\gamma)(\overline{x_N^2}-\mu_2')-(c_n-\gamma)(\overline{x_n^2}-\mu_2')+\gamma(\overline{x_N^2}-\mu_2')]^2$$
$$= \frac{\sigma^2_{y\cdot x}}{n}+c_{11}\sigma^2_{y\cdot x}\left\{\frac{\mu_2'-\mu_1'^2}{N}+(\bar{x}_n-\mu_1')^2\right\}+\beta^2\frac{\mu_2'-\mu_1'^2}{N}$$
$$+c_{22}\sigma^2_{y\cdot x}\left\{\frac{\mu'_4-\mu'^2_2}{N}+(\overline{x_n^2}-\mu'_2)^2\right\}+\gamma^2\frac{\mu'_4-\mu'^2_2}{N}$$
$$+2c_{12}\sigma^2_{y\cdot x}\left\{\frac{\mu'_3-\mu'_1\mu'_2}{N}+(\bar{x}_n-\mu'_1)(\overline{x_n^2}-\mu'_2)\right\}$$
$$+2\beta\gamma\,\frac{\mu'_3-\mu'_1\mu'_2}{N} \qquad \ldots \quad (3.53)$$

where
$$\mu'_r = E\left\{\frac{\sum_{i=1}^{n} x^r_{ni}}{n}\right\}$$

(3.5 a) The above result in (3.5) can be generalised when the non-linear relation between y and x is taken to be a parabola of pth degree *i.e.*

$$y = \alpha+\beta^1(x-\bar{x}_n)+\beta^2(x^2-\overline{x_n^2})+\ldots+\beta^p(x^p-\overline{x_n^p})$$

where $\overline{x_n^k} = \sum_1^n \frac{x_n^k}{n}$ (The numerals over β's are not powers).

The coefficients $\alpha, \beta^1, \beta^2, \ldots, \beta^p$ can be estimated by a_n, b_n^1, b_n^2, b_n^p by solving the $p+1$ normal equations and we can write

$$b_n^k = c_{k1}\Sigma(x-\bar{x}_n)y+c_{k2}\Sigma(x^2-\overline{x_n^2})y+\ldots+c_{kp}\Sigma(x^p-\overline{x_n^p})y \; (k = 1, 2, \ldots, p)$$

Here $V(a_n) = \frac{\sigma^2_{y\cdot x}}{n}$, $\text{cov}(b_n^i, b_n^j) = c_{ij}\sigma^2_{y\cdot x}(i, j, = 1, 2, \ldots, p)$

An unbiased estimate of the population mean of y's when we take another independent sample of x's of size N, will be given by

$$Y = \bar{y}_n+b^1_n(\bar{x}_N-\bar{x}_n)+b_n^2(\overline{x_N^2}-\overline{x_n^2})+\ldots+b_n^p(\overline{x_N^p}-\overline{x_n^p}) \qquad \ldots \quad (3.51a)$$
$$E(Y) = \eta \qquad \ldots \quad (3.52a)$$
$$V(Y) = \frac{\sigma^2_{y\cdot x}}{n}+\sigma^2_{y\cdot x}\sum_{i=1}^{p}\sum_{j=1}^{p}c_{ij}\left[\frac{\mu'_{i+j}-\mu'_i\mu'_j}{N}+(\overline{x_n^i}-\mu'_i)(\overline{x_n^j}-\mu'_j)\right]$$
$$+\sum_{i=1}^{p}\sum_{j=1}^{p}b_n^ib_n^j\,\frac{\mu'_{i+j}-\mu'_i\mu'_j}{N} \qquad \ldots \quad (3.53a)$$

(3.6) *x_n—Random, y_n—Random, x_N—Random, for non-linear regression*: If now we have the x's in the *first* sample *also random* instead of being taken as fixed, then also,

$$Y = \bar{y}_n + b_n(\bar{x}_N - \bar{x}_n) + c_n(\overline{x_N^2} - \overline{x_n^2}) \qquad \dots \quad (3.61)$$

Letting

$$E(\bar{x}_n) = E(\bar{x}_N) = \mu'_1$$

$$E(\overline{x_n^2}) = E(\overline{x_N^2}) = \mu'_2$$

$$E(\bar{y}_n) = \eta$$

and assuming large sample approximation (when b_n is independent of $\bar{x}_n$ and c_n is independent of $\overline{x_n^2}$) the expected value of Y will be *approximately* given by

$$E(Y) \simeq \eta + E(b_n)\, E(\bar{x}_N - \bar{x}_n) + E(c_n) E(\overline{x_N^2} - \overline{x_n^2}) \simeq \eta \qquad \dots \quad (3.62)$$

Also

$$V(b_n) = \sigma^2_{y\cdot x}\, E(c_{11}) = \sigma_{bb} \quad \text{(say)}$$

$$V(c_n) = \sigma^2_{y\cdot x}\, E(c_{22}) = \sigma_{cc} \quad \text{(say)}$$

$$\text{cov}\,(b_n, c_n) = \sigma^2_{y\cdot x} E(c_{12}) = \sigma_{bc} \quad \text{(say)}$$

where

$$c_{11} = \frac{\Sigma(x^2 - \overline{x_n^2})^2}{\begin{vmatrix} \Sigma(x - \bar{x}_n)^2 & \Sigma(x^2 - \overline{x_n^2})(x - \bar{x}_n) \\ \Sigma(x^2 - \overline{x_n^2})(x - \bar{x}_n) & \Sigma(x^2 - \overline{x_n^2}) \end{vmatrix}}$$

and likewise for c_{22} and c_{12}.

Now

$$E\{\Sigma(x - \bar{x}_n)^2\} = E\left\{\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right\}$$

$$= (n-1)(\mu'_2 - \mu_1'^2)$$

$$E\{\Sigma(x^2 - \overline{x_n^2})^2\} = E\left\{\Sigma x^4 - \frac{(\Sigma x^2)^2}{n}\right\}$$

$$= (n-1)(\mu'_4 - \mu_2'^2)$$

$$E\{\Sigma(x - \bar{x}_n)(x^2 - \overline{x_n^2})\} = E\left\{\Sigma x^3 - \frac{(\Sigma x)(\Sigma x^2)}{n}\right\}$$

$$= (n-1)(\mu_3' - \mu_1'\mu_2')$$

$$E(c_{11}) \simeq \frac{\textit{Expectation of numerator}}{\textit{Expectation of denominator}}$$

$$\simeq \frac{1}{(n-1)} \frac{(\mu_4' - \mu_2'^2)}{\mu_2(\mu_4' - \mu_2'^2) - (\mu_3' - \mu_1'\mu_2')^2}.$$

$$E(c_{22}) \simeq \frac{1}{(n-1)} \frac{\mu_2}{\mu_2(\mu_4' - \mu_2'^2) - (\mu_3' - \mu_1'\mu_2')^2}$$

$$E(c_{12}) \simeq -\frac{1}{(n-1)} \frac{(\mu_3' - \mu_1'\mu_2')}{\mu_2(\mu_4' - \mu_2'^2) - (\mu_3' - \mu_1'\mu_2')^2}$$

where
$$\mu_2 = \mu_2' - \mu_1'^2$$

$$V(Y) \simeq E\{(\bar{y}_n - \eta) + b_n(\bar{x}_N - \mu_1') - b_n(\bar{x}_n - \mu_1')$$
$$+ c_n(\bar{x}_N^2 - \mu_2') - c_n(\bar{x}_n^2 - \mu_2')\}^2$$
$$\simeq \frac{\sigma_y^2}{n} + \left(\frac{1}{N} + \frac{1}{n}\right)\{\mu_2(\sigma_{bb} + \beta^2) + (\mu_4' - \mu_2'^2)(\sigma_{cc} + \gamma^2)$$
$$+ 2(\mu_3' - \mu'_1\mu'_2)(\sigma_{bc} + \beta\gamma)\} \quad \ldots \quad (3.02)$$

I am indebted to Dr. C. R. Rao for his kind help and advice in the course of my investigation. I have also to thank Mr. A. Matthai for going through the manuscript and making some useful suggestions.

APPENDIX A

Let $((S_{ij}))$ denote the sample dispersion matrix for a k-variate normal population and $((c_{ij}))$ the corresponding reciprocal matrix. We then have the following results.

$$\left.\begin{aligned} E(c_{ij}) &= \frac{\sigma^{ij}}{n-k-2} \quad (i \neq j) \\ E(c_{ii}) &= \frac{\sigma^{ii}}{n-k-2} \end{aligned}\right\} \quad \ldots \quad (A.1)$$

Proof : Starting with the element c_{kk} of the matrix $((c_{ij}))$

$$E(c_{kk}) = E(S^{kk}) = E\left\{\begin{vmatrix} S_{11} & S_{12} & \cdots & S_{1,k-1} \\ S_{21} & S_{22} & \cdots & S_{2,k-1} \\ \cdots & \cdots & \cdots & \cdots \\ S_{k-1,1} & S_{k-1,2} & \cdots & S_{k-1,k-1} \end{vmatrix} \div \Delta\right\}$$

where

$$\Delta = \begin{vmatrix} S_{11} & S_{12} & \cdots & S_{1k} \\ S_{21} & S_{22} & \cdots & S_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ S_{k1} & S_{k2} & \cdots & S_{kk} \end{vmatrix}$$

Thus
$$E(c_{kk}) = \frac{1}{n} E\left\{\frac{(t_{11}t_{22}\ldots t_{k-1,k-1})^2}{(t_{11}t_{22}\ldots t_{kk})^2}\right\}$$
$$= \frac{1}{n} E\left(\frac{1}{t^2_{kk}}\right)$$

where t's represent the rectangular coordinates (Mahalanobis, Bose and Roy (1937)).

The distribution of t_{kk} when other t's are integrated out is

$$\text{const. } e^{-\frac{n}{2}[T^{kk}t_{kk}^2]}(t^2_{kk})^{\frac{n-k-1}{2}}dt_{kk}$$

where the constant is such that

$$\text{const.} \int_{-\infty}^{\infty} e^{-\frac{n}{2}[T^{kk}t^2_{kk}]}(t_{kk}^2)^{\frac{n-k-2}{2}}d(t^2_{kk}) = 1$$

$$\therefore\ E\left(\frac{1}{t^2_{kk}}\right) = \text{const.} \int_{-\infty}^{\infty} e^{-\frac{n}{2}[T^{kk}t^2_{kk}]}(t^2_{kk})^{\frac{n-k-4}{2}}d(t^2_{kk})$$

$$= \frac{\left(\frac{n}{2}T^{kk}\right)^{\frac{n-k}{2}}}{\Gamma\left(\frac{n-k}{2}\right)} \frac{\Gamma\left(\frac{n-k-2}{2}\right)}{\left(\frac{n}{2}T^{kk}\right)^{\frac{n-k-2}{2}}}$$

$$= \frac{n}{n-k-2}\,T^{kk}$$

$$= \frac{n}{n-k-2}\,\sigma^{kk}$$

Thus $$E(c_{kk}) = \frac{\sigma^{kk}}{n-k-2}$$ and in general

$$E(c_{ii}) = \frac{\sigma^{ii}}{n-k-2}$$

Again,

$$E(c_{k,k-1}) = E\left\{\begin{vmatrix} S_{11} & \cdots & S_{1,k-2} & S_{1k} \\ S_{21} & \cdots & S_{2,k-2} & S_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ S_{k-1,1} & \cdots & S_{k-1,k-2} & S_{k-1,k} \end{vmatrix} \div \Delta\right\}$$

$$= -\frac{1}{n}E\left\{\frac{t_{k-1,k}(t_{11}t_{22}\cdots t_{k-2,k-2})(t_{11}t_{22}\cdots t_{k-1,k-1})}{(t_{11}t_{22}\cdots t_{kk})^2}\right\}$$

$$= -\frac{1}{n}E\left\{\frac{t_{k-1,k}}{t_{k-1,k-1}t_{kk}^2}\right\}$$

Let us denote

$$A_{11} = \frac{n}{2} T^{kk}, \qquad A_{12} = A_{21} = \frac{n}{2} T^{k-1,k}$$

$$A_{22} = \frac{n}{2} T^{k-1,k-1}, \quad t_{kk} = x_1, \ t_{k-1,k-1} = x_2, \ t_{k-1,k} = x_3$$

Now the joint-distribution of t_{kk}, $t_{k-1,k}$ and $t_{k-1\,k-1}$ is given by

$$C \int \int \int \exp. -\frac{n}{2} [T^{kk} t^2_{kk} + (T^{k-1,k-1} t^2_{k-1,k-1} + 2T^{k-1,k} t_{k-1,k-1} t_{k-1,k} + T^{kk} t^2_{k-1,k})] \times$$

$$t_{kk}^{n-k-1} t^{n-k}_{k-1,k-1} dt_{kk} dt_{k-1,k-1} dt_{k-1,k}$$

where the constant C is given by

$$C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp. -\frac{n}{2} \Big[\ldots\ldots \Big] t_{kk}^{n-k-1} t^{n-k}_{k-1,k-1} dt_{kk} dt_{k-1,k-1} dt_{k-1,k} = 1$$

or, integrating over t_{kk},

$$C_1 \frac{\Gamma\left(\frac{n-k}{2}\right)}{A_{11}^{\frac{n-k}{2}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp. -[A_{22} x_2^2 + 2A_{21} x_2 x_3 + A_{11} x_3^2]\, x_2^{n-k} dx_2 dx_3 = 1$$

or, $$C_1 \frac{\Gamma\left(\frac{n-k}{2}\right)}{A_{11}^{\frac{n-k}{2}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp. -\Big[A_{11} \left(x_3 + \frac{A_{21}}{A_{11}} x_2\right)^2 + A_{33} x_2^2\Big] x_2^{n-k} dx_2 dx_3 = 1 \quad \ldots \text{(A.11)}$$

where $$A_{33} = \frac{A_{11} A_{22} - A_{21}^2}{A_{11}}$$

putting $$y_2 = x_3 + \frac{A_{21}}{A_{11}} x_2$$

$$y_3 = x_2$$

so that $$\frac{\partial(y_2, y_3)}{\partial(x_2, x_3)} = 1$$

Hence (A.11) reduces to

$$C_2 \frac{\Gamma\left(\frac{n-k}{2}\right)}{A_{11}^{\frac{n-k}{2}}} \frac{\Gamma\left(\frac{1}{2}\right)}{A_{11}^{\frac{1}{2}}} \frac{\Gamma\left(\frac{n-k+1}{2}\right)}{A_{33}^{\frac{n-k+1}{2}}} = 1$$

or $$C_2 = \frac{\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix}^{\frac{n-k+1}{2}}}{\Gamma\left(\frac{n-k}{2}\right) \Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{n-k+1}{2}\right)}$$

Now
$$E\left\{-\frac{l_{k\cdot 1k}}{l_{k\cdot 1k\cdot 1}l^2_{kk}}\right\} = E\left\{-\frac{x_2}{x_2 x_1^2}\right\}$$

$$=C_2\frac{\Gamma\left(\frac{n-k-2}{2}\right)}{A_{11}^{\frac{n-k-2}{2}}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-A_{11}y_2^2-A_{33}y_3^2}\, y_3^{n-k-1}\left(\frac{A_{21}}{A_{11}}\, y_3-y_2\right)dy_2dy_3$$

$$=C_2\frac{\Gamma\left(\frac{n-k-2}{2}\right)}{A_{11}^{\frac{n-k-2}{2}}}\left\{\frac{A_{21}}{A_{11}}\,\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-k+1}{2}\right)}{A_{11}^{1/2}\,A_{33}^{\frac{n-k+1}{2}}}+0\right\}$$

since the second term is an odd function and the range is from $-\infty$ to ∞,

$$=C_2\frac{\Gamma\left(\frac{n-k-2}{2}\right)}{\left(A_{11}A_{33}\right)^{\frac{n-k+1}{2}}}\,\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-k+1}{2}\right)A_{21}$$

$$=\frac{\Gamma\left(\frac{n-k-2}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right)}A_{21}$$

$$=\frac{n}{n-k-2}T^{k-1\cdot k}$$

$$=\frac{n}{n-k-2}\,\sigma^{k\cdot k-1}$$

$$\therefore\quad E(c_{k\cdot k-1})=\frac{\sigma^{k\cdot k-1}}{n-k-2}$$

Hence in general

$$E\,(c_{ij})=\frac{\sigma^{ij}}{n-k-2}\left(\begin{array}{c} i=1,2,\ldots,k \\ j=1,2,\ldots,k\end{array}\right)$$

APPENDIX B

The joint distribution of the regression coefficients (b_{yi}) of y on x_i $(i=1,2,\ldots,k$ when x's follow a multivariate normal distribution is given by

$$\frac{\Gamma\left(\frac{n}{2}\right)|\sigma^{ij}|^{\frac{n-1}{2}}(\sigma^{yy})^{k/2}}{\Gamma\left(\frac{n-k}{2}\right)\pi^{k/2}}\;\frac{\prod_{i=1}^{k} db_{yi}}{\left|\sigma^{ij}+(b_{yi}-\beta_{yi})(b_{yj}-\beta_{yj})\sigma^{yy}\right|^{n/2}}$$

Proof: When x_i's $(i = 1,2,..,k)$ are constant the multivariate distribution of b_{yi}'s is given by

$$\frac{|S_{ij}|^{1/2}}{(\sigma_{y\cdot x}\sqrt{2\pi})^k} e^{-\frac{1}{2\sigma^2_{y\cdot x}}\sum_{i=1}^{k}\sum_{j=1}^{k} S_{ij}(b_{yi}-\beta_{yi})(b_{yj}-\beta_{yj})} \prod_{i=1}^{k} db_{yi}$$

$$\left[\text{where} \quad \sigma^2_{y\cdot x} = \frac{1}{\sigma^{yy}}\right]$$

While the distribution of x's follows the Wishart distribution

$$\frac{|\sigma^{ij}|^{\frac{n-1}{2}}}{2^{(n-1)k/2}\pi^{k(k-1)/4}} \cdot \frac{1}{\prod_{i=1}^{k}\Gamma\left(\frac{n-i}{2}\right)} e^{-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\sigma^{ij}S_{ij}} |S_{ij}|^{\frac{n-k-2}{2}} \prod_{i=1}^{k}\prod_{j=1}^{k} dS_{ij}$$

Hence the distribution of b_{yi}'s when x's vary will be given by

$$\text{const.} \prod_{i=1}^{k} db_{yi} \int\int\cdots\int e^{-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left\{\sigma^{ij}+\frac{(b_{yi}-\beta_{yi})(b_{yj}-\beta_{yj})}{\sigma_{y\cdot x}^2}\right\}S_{ij}} |S_{ij}|^{\frac{n-k-1}{2}} \prod_{i=1}^{k}\prod_{j=1}^{k} dS_{ij}$$

$$= C\frac{\prod_{i=1}^{k} db_{yi}}{|\gamma^{ij}|^{n/2}}$$

where

$$\gamma^{ij} = \sigma^{ij}+\frac{(b_{yi}-\beta_{yi})(b_{yj}-\beta_{yj})}{\sigma_{y\cdot x}^2}$$

$$= \sigma^{ij}+(b_{yi}-\beta_{yi})(b_{yj}-\beta_{yj})\sigma^{yy}$$

and

$$C = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right)} \frac{|\sigma^{ij}|^{\frac{n-1}{2}}(\sigma^{yy})^{k/2}}{\pi^{k/2}}$$

APPENDIX C

The distribution of $b_{k\cdot k-1\cdot 12\cdots k-2}$ (*i.e.* regression coefficient of x_k on x_{k-1}) when $x_1 x_2 \ldots x_k$ follow a multivariate normal distribution is given by

$$\frac{1}{a\,B\left(\frac{n-k+1}{2}, \frac{1}{2}\right)}\left\{1+\frac{(b_{k\cdot k-1\cdot 12\cdots k-2}-\beta_{k\cdot k-1\cdot 12\cdots k-2})^2}{a^2}\right\}^{-\frac{n-k+2}{2}} \times db_{k\cdot k-1\cdot 12\cdots k-2} \quad \text{(C.1)}$$

where

$$a = \frac{\begin{vmatrix} \sigma^{kk} & \sigma^{k,k-1} \\ \sigma^{k-1,k} & \sigma^{k-1,k-1} \end{vmatrix}^{2}}{\sigma^{kk}}$$

Proof,

$$b_{k\cdot k-1\cdot 12\cdots k-2} = \frac{t_{k-1\cdot k}}{t_{k-1\cdot k-1}}$$

The joint distribution of $t_{k-1\cdot k-1}$ and $t_{k-1\cdot k}$ is given by

$$\text{const. exp.}\left\{-\frac{n}{2}\left[T^{k-1\cdot k-1}t^2_{k-1,k-1} + 2T^{k-1\cdot k}t_{k-1,k-1}\,t_{k-1,k} + T^{kk}\,t^2_{k-1,k}\right]\right\} t^{n-k}_{k-1,k-1}\,dt_{k-1,k-1} \times dt_{k-1\cdot k} \quad \ldots \quad \text{(C.11)}$$

Let $u = \frac{l_{k-1 \cdot k}}{l_{k-1 \cdot k-1}}$ and $v = l_{k-1 \cdot k} \, l_{k-1 \cdot k-1}$

so that the jacobian of the transformation is given by

$$\frac{\partial(u, v)}{\partial(l_{k-1 \cdot k-1}, l_{k-1 \cdot k})} = 2 \frac{l_{k-1 \cdot k}}{l_{k-1 \cdot k-1}} = 2u$$

From (C.11) the joint distribution of u and v is given by

$$\text{const. exp.} \left\{ -\frac{nv}{2} \left[T^{kk} u + 2T^{k-1 \, k} + \frac{T^{k-1 \, k-1}}{u} \right] \right\} \times v^{\frac{n-k+2}{2} - 1} u^{-\frac{n-k+2}{2}} du \, dv \quad \dots \quad (C.12)$$

Integrating over v from 0 to ∞ and noting that $T^{ij} = \sigma^{ij}$ and

$$E(b_{k \cdot k-1 \cdot 12 \cdots k-2}) = \beta_{k \cdot k-1 \cdot 12 \cdots k-2} = -\frac{\sigma^{k \cdot k-1}}{\sigma^{kk}}$$

the distribution of u (*i.e.* $b_{k \cdot k-1 \cdot 12 \cdots k-2}$) is given by

$$\text{const.} \quad \left\{ 1 + \frac{(b_{k \cdot k-1 \cdot 12 \cdots k-2} - \beta_{k \cdot k-1 \cdot 12 \cdots k-2})^2}{a^2} \right\}^{-\frac{n-k+2}{2}} b_{k \cdot k-1 \cdot 12 \cdots k-2}$$

where

$$a = \frac{\begin{vmatrix} \sigma^{kk} & \sigma^{k, \, k-1} \\ \sigma^{k-1, \, k} & \sigma^{k-1, \, k-1} \end{vmatrix}^{\frac{1}{2}}}{\sigma^{kk}}$$

and the const. is found on integration between $-\infty$ to ∞ to be equal to

$$\frac{1}{a \; B\left(\frac{n-k+1}{2}, \frac{1}{2}\right)}$$

References

Bose, C. (1943): Note on sampling error in the method of double sampling. *Sankhyā*, **6**, 329-30.

Bose, C. and Gayen, A. K. (1946): Note on the expected discrepancy in the estimation (by double sampling) of a variate in terms of a concomitant variate when there exists a nonlinear regression between the two variates. *Sankhyā*, **8**, 73-74.

Cochran, W. G. (1939): The use of analysis of variance in enumeration by sampling. *J. Amer. Stat. Assocn.*, **34**, 494-96.

Ghosh, B. (1947): Double sampling with many auxiliary variates. *Cal. Stat. Assocn. Bull.*, **1**, 91-93.

Mahalanobis, P. C., Bose, R. C. and Roy, S. N. (1937): Normalisation of statistical variates and the use of rectangular coordinates in the theory of sampling distributions. *Sankhyā*, **3**, 8-40.

Neyman, J (1938): Contribution to the theory of sampling human populations. *J. Amer. Stat. Assocn.*, **33**, 101-116.

Schumacher, F. X. and Chapman, R. A. (1942): Sampling methods in forestry and range management. Durham, North Carolina, 186-189.

Snedecor, G. W. and King, A. J. (1942): Recent developments in sampling for agricultural statistics. *J. Amer. Stat. Assocn.*, **37**, 99-100.

Paper received : November, 1950.