ON STRATIFICATION IN SAMPLING INVESTIGATION INVOLVING MORE THAN ONE CHARACTER

BY

P.C. GUPTA* AND G.R. SETH**

Rajasthan University, Jaipur and F.A.O., Rome

(Received : January, 1973)

1. INTRODUCTION

Determination of strata boundaries in case of stratified sampling was first considered by Dalenius [1] stratifying on the character itself he obtained equations whose solutions give the strata boundaries by minimising the variance of the stratified random sampling estimate, both in case of optimum and proportional allocation. Dalenius and Gurney [2] considered the problem of stratification on the basis of an auxiliary variable. Taga (1967) considered a general problem of optimum stratification based on concomitant variable (when strata are not necessarily intervals) in case of proportional allocation. Singh and Sukhatme [6] considered the problem of optimum stratification on an auxiliary variable x when the form of the regression of the estimation variable y and the form of the variance function $v(y/x)$ are known. They obtained the minimal equations giving optimum strata boundaries for Neyman and proportional allocation and suggested various approximate solutions as these equations were not directly solvable. Singh and Sukhatme [7] and Singh [9] further extended these results when the auxiliary variable was also used to construct the varying probability estimates in stratified sampling for all the three allocations namely, proportional, Neyman and equal. Singh and Prakash Dev. (8) studied the approximation suggested by Singh and Sukhatme [6] for equal allocation and suggested a cum $f\sqrt{\phi}$ rule for finding out the approximate optimum strata boundaries. Ghosh [4] has studied two-way stratification based on two characters under study and has discussed optimum boundary points which minimises the generalised variance of the unbiased linear

* At present at South Gujarat University, Surat.

** At present in New Delhi.

estimate. In the persent paper, stratification for the study of more than one characters on the basis of an auxiliary character has been investigated. The knowledge of the number of strata is assumed together with the nature of allocation of units among strata namely proportionate allocation.

2. Mathematical Formation

The following assumptions are made :

(*i*) Variates X, Y, and Z have a joint continuous probability density function $f(x, y, z)$ and first two moments of X, Y Z exist.

(*ii*) Population is infinite.

Though these assumptions will not, in general be satisfied, yet in practice they will be approximately satisfied.

Let $-\infty, x_1, x_2, \ldots x_{i-1}, x_i, x_{i+1}, \ldots\ldots x_{L-1}, +\infty$ denote the boundaries of 'L' strata where L is fixed in advance.

$$P_i = \int_{x_{i-1}}^{x_i} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y, z)\, dy\, dz\, dx$$

$=$Proportion of units falling in the ith stratum ...(1)

$$\mu_{iy} = \frac{1}{P_i} \int_{x_{i-1}}^{x_i} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x, y, z)\, dy\, dz\, dx$$

$=$mean of the character Y of the ith stratum ...(2)

$$\sigma^2_{iy} = \frac{1}{P_i} \int_{x_{i-1}}^{x_i} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (y-\mu_{iy})^2 f(x, y, z)\, dy\, dz\, dx$$

$=$Variance of the character Y of the ith stratum ...(3)

with similar expression for μ_{iz} and σ^2_{iz}

$$\sigma_{iyz} = \frac{1}{P_i} \int_{x_{i-1}}^{x_i} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (y-\mu_{iy})(z-\mu_{iz})\, f(x, y, z)$$

$dy\, dz\, dx =$Covariance between character Y and Z of the ith stratum ...(4)

$$g(x)=\int_{-\infty}^{+\infty}\int_{+\infty}^{+\infty} f(x, y, z)\, dy\, dz\, dx$$

$$=\text{marginal density function of } X \qquad \ldots(5)$$

$$E(Y \mid X=x_i)=\frac{1}{g(x_i)}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x_i, y, z)\, dy\, dz \qquad \ldots(6)$$

and

$$E(Z \mid X=x_i)=\frac{1}{g(x_i)}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} zf(x_i, y, z)\, dy\, dz \qquad \ldots(7)$$

Suppose a stratified sample of size n is drawn with simple random sample of n_i observations from the ith stratum of the population and the following sample statistics are evaluated. Let y_{ij} and z_{ij} denote the value of the ith unit in the ith stratum of character Y and Z respectively.

Then,

$$\bar{y}_i=\frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} \quad =\text{sample mean of } i\text{th stratum for character } Y \qquad \ldots(8)$$

$$\bar{z}_i=\frac{1}{n_i}\sum_{j=1}^{n_i} z_{ij} \quad =\text{sample mean of } i\text{th stratum for character } Z. \qquad \ldots(9)$$

and

$$\bar{y}=\sum_{i=1}^{L} p_i\, \bar{y}_i, \quad \bar{z}=\sum_{i=1}^{L} P_i\, \bar{z}_i \quad \text{provide the conventional}$$

unbiased estimate of the population mean of character Y and Z respectively.

Variances and covariance of these estimates are given below

$$\sigma^2_{\bar{y}}=\text{Var}(\bar{Y})=\sum_{i=1}^{L} P_i^2\ \sigma^2_{iy}/n_i \qquad \ldots(10)$$

$$\sigma^2_{\bar{z}}=\text{Var}(\bar{z})=\sum_{i=1}^{L} P_i^2\ \sigma^2_{iz}/n_i \qquad \ldots(11)$$

$$\sigma_{\bar{y}\bar{z}}=\text{Cov}(\bar{Y}, \bar{Z})=\sum_{i=1}^{L} P_i^2\ \sigma_{iyz}/n_i \qquad \ldots(12)$$

and the generalised variance $|G|$ of the mean of a stratified random sample with simple random sampling of units within each stratum is given by

$$|G| = \begin{vmatrix} \sigma^2_{\bar{y}} & \sigma_{\bar{y}\bar{z}} \\ \sigma_{\bar{y}\bar{z}} & \sigma^2_{\bar{z}} \end{vmatrix} \quad ...(13)$$

3. Strata Boundaries for Proportionate Allocation

For the proportionate allocation with fixed sample size, that is for $n_i=nP_i$ variances and covariances of unbiased estimates of population mean of character Y and Z reduce to

$$\sigma^2_{\bar{y}} = \frac{1}{n}\sum_{i=1}^{L} P_i\, \sigma^2_{iy} \quad ...(14)$$

$$\sigma^2_{\bar{z}} = \frac{1}{n}\sum_{i=1}^{L} P_i\, \sigma_{iz} \quad ...(15)$$

$$\sigma_{\bar{y}\bar{z}} = \frac{1}{n}\sum_{i=1}^{L} P_i\, \sigma_{iyz} \quad ...(16)$$

The object is to find x_i's $(i=1, 2, \ldots L.1)$ which minimise the value of generalised variance given by $|G|$.

Minimisation of $|G|$ with respect to x_i's leads to

$$\sigma^2_{\bar{z}}\ (\mu_{i+1y}-\mu_{iy}) \left[E(Y \mid x_i) - \frac{\mu_{i+iy}+\mu_{iy}}{2} \right] + \sigma^2_{\bar{y}} (\mu_{i+iz}-\mu_{iz})$$

$$\left[E(Z \mid X_i) - \frac{\mu_{i+iz}+\mu_{iz}}{2} \right] - \sigma_{\bar{y}\bar{z}}[(E(Y \mid x_i)(\mu_{i+iz}-\mu_{1z}) + E(Z \mid x_i)$$

$$(\mu_{i+1y}-\mu_{iy}) - (\mu_{i+1z}\mu_{i+1y}-\mu_{iz}\mu_{iy})] = 0 \quad ...(17)$$

$$(i=1, 2, \ldots, L-1)$$

Solution of (17) will give boundary points which minimises $|G|$. It can be noted here that if coefficients of $\sigma^2_{\bar{y}}$ and $\alpha^2_{\bar{z}}$ are equated to zero, the coefficient $\sigma_{\bar{y}\bar{z}}$ becomes automatically zero. Its solution will be studied in section 4.

4. Stratification with a General Regression of Y and Z on X

Let the regression of Y as well as Z on X is of the same form given by

$$E(Y \mid X=x) \qquad \alpha_y+\beta_y \ \phi(x) \qquad \ldots(18)$$

$$E(Z \mid X=x) \qquad \alpha_z+\beta_z \ \phi(x) \qquad \ldots(19)$$

Then

$$\mu_{iy}=\alpha_y+\beta_y \ E(\phi(x)/x\epsilon A_i) \qquad \ldots(20)$$

$$\mu_{iz}=\alpha_z+\beta_z \ E(\phi(x)/x\epsilon A_i) \qquad \ldots(21)$$

where

$$E(\phi(x)/x\epsilon A_i)=\frac{1}{P_i}\int_{x_{i-1}}^{x_i} \phi(x)\, g(x)\, dx \qquad \ldots(22)$$

And $\qquad A_i$ is the interval (x_{i-1}, x_i).

Substituting these values in equation (17) giving the boundaries of the strata we get

$$\beta_y^2 \ \sigma_{\bar{z}}^2 \left[(E(\phi(x)/x\epsilon A_{i+1})-E(\phi(x)/x\epsilon A_i)\right]$$

$$\left[\phi(x_i)-\frac{E(\phi(x)/x\epsilon A_i)+E(\phi(x)/x\epsilon A_{i+1})}{2}\right]$$

$$+\beta_z^2 \ \sigma_{\bar{y}}^2 \left[E(\phi(x)/x\epsilon A_{i+1})-E(\phi(x)/x\epsilon A_i)\right]$$

$$\left[\phi(x_i)-\frac{E(\phi(x)/x\epsilon A_i)+E(\phi(x)/x\epsilon A_{i+1})}{2}\right]$$

$$-2\beta_y\beta_z\sigma_{\bar{y}\bar{x}}\left[E(\phi(x)/x\epsilon A_{i+1})-E(\phi(x)/x\epsilon A_i)\right]$$

$$\left[\phi(x_i)-\frac{E(\phi(x)/x\epsilon A_i)+E(\phi(x)/x\epsilon A_{i+1})}{2}\right]=0$$

or

$$\left[E(\phi(x)/x \ A_{i+1})-E(\phi(x)/x \ A_i)\right] \left[\beta_y^2 \ \sigma_{\bar{z}}^2+\beta_z^2 \ \sigma_y^2 -2\beta_y\beta_z\sigma_{\bar{y}\bar{z}}\right]$$

$$\times\left[\phi(x_i)-\frac{E(\phi(x)/x\epsilon A_i)+E(\phi(x)/x\epsilon A_{i+1})}{2}\right]=0 \qquad \ldots(23)$$

which gives either

$$E(\phi(x)/x\epsilon A_i)=E(\phi(x)/x\epsilon A_{i+1}) \qquad \ldots(24)$$

or

$$\phi(x_i)=\frac{E(\phi(x)/x\epsilon A_i)+E(\phi(x)/x\epsilon A_{i+1})}{2} \qquad \ldots(25)$$

$$(i=1, 2, 3, \ldots\ldots L-1)$$

because $\sigma^2_{\bar{z}}\,\beta^2_y+\sigma^2_{\bar{y}}\,\beta^2_z-2\sigma_{\bar{y}\bar{z}}\,\beta_y\beta_z\neq0$ as

$$\begin{vmatrix}\sigma^2_{\bar{y}} & -\sigma_{\bar{y}\bar{z}}\\ -\sigma_{\bar{y}\bar{z}} & \sigma^2_{\bar{z}}\end{vmatrix}$$ is positive definite

The solution given by (24) will give a poor stratification as in that case strata means for character Y and Z do not vary from one stratum to another. Thus the only meaningful solution which minimises $|G|$ is gives by equation (25).

4.1. Linear Regression

Here
$$E(Y/X=x)=\mu_y+\beta_{yx}(x-\mu_x) \quad ...(26)$$
$$E(Z/X=x)=\mu_z+\beta_{zx}(x-\mu_x) \quad ...(27)$$
and
$$\mu_{iy}=\mu_y+\beta_{yx}(\mu_{ix}-\mu_x) \quad ...(28)$$
$$\mu_{iz}=\mu_z+\beta_{zx}(\mu_{ix}-\mu_x) \quad ...(29)$$
where
$$\mu_{ix}=\frac{1}{P_i}\int_{x_{i-1}}^{x_i} xg(x)dx=\text{mean of character } X \text{ of } i\text{th stratum}$$

Putting these values in equation (17) we get
$$\mu_{i+1x}=\mu_{ix} \quad ...(30)$$
$$x_i=\frac{\mu_{ix}+\mu_{i+ix}}{2} \quad ...(31)$$

as $\sigma^2_{\bar{z}}\,\beta^2_{yx}+\sigma^2_{\bar{y}}\,\beta^2_{zx}-2\beta_{zx}\beta_{yx}\,\bar{y}\bar{z}\neq0$ by same argument given above.

Equation (31) provides the boundary points which minimises $|G|$. Incidently, these are the same as obtained by Dalenius, when he stratifies the population on the character X under study itself.

4.2. Expotential regression

Here
$$E(Y/X=x)=\alpha_y+\beta_y\,e^{-\gamma x} \qquad \gamma>0 \quad ...(32)$$
$$E(Z/X=x)=\alpha_z+\beta_z\,e^{-\gamma x} \qquad \gamma>0 \quad ...(33)$$
and
$$\mu_{iy}=\alpha_y+\beta_y\,E(e^{-\gamma x}/x\in A_i) \quad ...(34)$$
$$\mu_{iz}=\alpha_z+\beta_z\,E(e^{-\gamma x}/x\in A_i) \quad ...(35)$$

where

$$E(e)^{-\gamma x}/x\epsilon A_i) = \frac{1}{P_i}\int_{x_{i-1}}^{x_i} e^{-\gamma x}\ g(x)dx \qquad \ldots(36)$$

Substituting these values in (17) we have either

$$E\,(e^{-\gamma x}/x\epsilon A_i) = E\,(e^{-\gamma x}/x\epsilon A_{i+1}) \qquad \ldots(37)$$

or

$$e^{-\gamma x_i} = \frac{E(e^{-\gamma x}/x\epsilon A_i) + E\,(e^{-\gamma x}/x\epsilon A_{i+1})}{2} \qquad \ldots(38)$$

as $\sigma^2_{\bar{y}}\ \ \beta^2_y\ +\beta^2_y\ \ \sigma^2_{\bar{z}}\ -2\beta_y\ \beta_z\ \sigma_{\bar{y}\bar{z}} \neq 0$ by the same argument as above.

Again (37) provides poor startification for Y and Z with same reasoning as earlier. The optimum boundary points are given by (38).

5. Study of Individual Variances and Covariance

In this section we shall study an important property that this type of stratification not only minimises the generalised variance $/G/$ but their individual variances also, when the regression of Y and Z on character X on which startification is based, has the same form say as in (18) and (19).

We know that

$$P_i\sigma^2_{iy} = \int_{x_{i-1}}^{x_i} E\,(y^2/x)\ g\ (x)\ \ dx - \mu^2_{iy}$$

$$= \int_{x_{i-1}}^{x_i} \text{Var}\ (y/x)\ g\ (x)\ dx + \int_{x_{i-1}}^{x_i} E^2\ (y/x)\ g\ (x)\ dx - \mu^2_{iy}$$

But $$\int_{x_{i-1}}^{x_i} E^2\ (y/x)\ g\ (x)\ dx = \int_{x_{i-1}}^{x_i} (\alpha_y + \beta_y\ (\phi\ (x))^2\ g\ (x)\ dx$$

$$= \alpha^2_y\ \ P_i + 2\alpha_y\beta_y\, P_i E(\phi(x)/x\epsilon A_i) + \beta^2_y\ \ P_i\ E\ (\phi^2(x)/x\epsilon A_i)$$

or

$$P_i\,\sigma^2_{iy} = \int_{x_{i-1}}^{x_i} \text{Var}\,(y/x)\,g\,(x)\,dx + \beta_y^2\;P_i\,\sigma^2_{i\phi(x)}$$

where

$$\sigma^2_{i\phi(x)} = E\phi^2\,(x)/x\in A_i) - E^2\,(\phi\,(x)/x\in A_i \qquad \text{...(39)}$$

or

$$\sum_{i=1}^{L} P_i\,\sigma^2_{iy} = E\;(\text{Var}\;\;(Y/X)) + \beta_y^2 \sum_{i=1}^{L}\; P_i\,\sigma^2_{i\phi\,(x)} \qquad \text{...(40)}$$

where

$$E\,(\text{Var}\,(Y/X)) = \int_{-\infty}^{+\infty} \text{Var}\,(Y/X)\,g\,(x)\,dx \qquad \text{...(41)}$$

Similarly we get,

$$\sum_{i=1}^{L} P_i\,\sigma^2_{iz} = E(\text{Var}(Y/X)) + \beta_z^2 \sum_{i=1}^{L} P_i\,\sigma^2_{i\,\phi\,(x)} \qquad \text{...(42)}$$

and

$$\sum_{i=1}^{L} P_i\,\sigma_{iyz} = E\;(\text{Cov}\;(Y,\;Z/X)) + \beta_y\beta_z \sum_{i=1}^{L} P_i\,\sigma^2_{i\,\phi\,(x)} \qquad \text{...(43)}$$

where

$$E\,(\text{Var}(Z/X)) = \int_{-\infty}^{+\infty} \text{Var}\,(Z/X)\,g\,(x)\,dx \qquad \text{...(44)}$$

$$E(\text{Cov.}\,(Y,Z/X)) = \int_{-\infty}^{+\infty} \text{Cov}\,(Y,Z/X)\,g\,(x)\,dx \qquad \text{...(45)}$$

Now with the help of (42), (44), (45) G can be written as

$$G=\frac{1}{n^2}\begin{bmatrix} E(\text{Var } (Y/X))+\beta_y^2 \sum_{i=1}^{L} P_i\, \sigma^2_{i\phi(x)} & E(\text{Cov } (Y, Z/X)) + \beta_y\beta_z \sum_{i=1}^{L} P_i\, \sigma^2_{i\phi(x)} \\ E(\text{Cov. } (Y,Z/X) + \beta_y\beta_z \sum_{i=1}^{L} P_i\, \sigma^2_{i\phi(x)} & E(\text{Var } (Z/X)) + \beta_z^2 \sum_{i=1}^{L} P_i\, \sigma^2_{i\phi(x)} \end{bmatrix}$$

$$=\frac{1}{n^2}\begin{bmatrix} E(\text{Var } (Y/X)) & E(\text{Cov. } (Y, Z)/X) \\ E(\text{Cov. } (Y, Z/X)) & E(\text{Var } (Z/X)) \end{bmatrix}$$

$$+\sum_{i=1}^{L} P_i\, \sigma^2_{i\phi(x)} \Big[\beta_y^2\ E(\text{Var}(Z/X)) + \beta_z^2\ E(\text{Var } (Y/X))$$

$$-2\, \beta_y\beta_z.\ E(\text{Cov. } (Y, Z/X)) \Big] \qquad \ldots(46)$$

Thus to minimise $|G|$ is the same as minimising

$$\sum_{i=1}^{L} P_i\ \sigma^2_{i\phi(x)}$$

It can be also seen that

$$\sum_{i=1}^{L} P_i\ \sigma^2_{iy}$$

and

$$\sum_{i=1}^{L} P_i\ \sigma^2_{iz}$$

are also simultaneously minimised when

$$\sum_{i=1}^{L} P_i \, \sigma^2_{i\phi(x)}$$

is minimised. The latter is minimised when we stratify the population with regard to the x with boundary points given by expression (25). In case $\phi(x)$ is linear, the stratification on the character X reduces to optimum stratification for the Character X as well.

This result has an important implication. That is under the conditions of that character Y and Z has the same form of regression on X, it is possible to fix up stratification which minimises the variance estimates of the population mean of Y, Z as well the generalised variance leading to the result that one can get estimates of Y and Z having the smallest concentration ellipsoid.

For one thing stratification on the characters Y and Z themselves is not possible in general. Secondly unless there is linear regression of Y on Z or Z on Y, it is not possible to fix points of stratification for which

$$\sum_{i=1}^{L} P_i \, \sigma^2_{iz} \, , \quad \sum_{i=1}^{L} P_i \, \sigma^2_{iy}$$

based on character Y or Z are simultaneously minimised. For instance if we stratify on the basis of character Z and assuming that regression Y on Z is given by

$$E(Y/Z=z)=\alpha_y+\beta_y \Psi(z) \qquad \ldots(47)$$

by the same dispersion matrix of the estimates of $\bar{Y}$ and $\bar{Z}$ of population mean of Y and Z is given by

$$|G|=\begin{bmatrix} \text{EVar}\,(Y/Z)+\beta_y^2 \sum_{i=1}^{L} P_i \, \sigma^2_{i\phi(x)} & \beta_y \sum_{i=1}^{L} P_i \, \text{Cov}\,(Z, \psi(Z)/z\epsilon A_i) \\ \beta_y \sum_{i=1}^{L} P_i \, \text{Cov}\,(Z, \psi(Z)/z\epsilon A_i) & \sum_{i=1}^{L} P_i \, \sigma^2_{iz} \end{bmatrix} \qquad (\ldots 48)$$

and the optimum boundary points are given by

$$\beta_y^2\,\sigma_{\bar{z}}^2\,[E\,(\psi(Z)/z\epsilon A_{i+1})-E\,(\psi\,(Z)/z\epsilon A_i)]$$

$$\left[\psi(z_i)-\frac{E\,(\psi(Z)/z\epsilon A_{i+1})+E(\psi(Z)/z\epsilon A_i)}{2}\right.$$

$$\left.+\sigma_{\bar{y}}^2\,(\mu_{i+1z}-\mu_{iz})\left(z_i-\frac{\mu_{i+1z}+\mu_{iz}}{2}\right)-\beta_y\,\sigma_{\bar{y}\bar{z}}\,(\mu_{i+1z}-\mu_{iz})\right]$$

$$\left((z_i)-\frac{E(\Psi(Z)/z\epsilon A_{i+1})+E(\Psi(Z)/z\epsilon A_i}{2}\right)$$

$$+(E(\Psi(Z)/z\epsilon A_{i+1})-E(\Psi(R)/z\epsilon A_i).$$

$$\left(z_i-\frac{\mu_{i+1z}+\mu_{iz}}{2}\right)=0 \qquad ...(49)$$

It is obvious from (48) that simultaneous minimization of

$$\sum_{i=1}^{L} P_i\sigma_{iy}^2\ ,\ \sum_{i=1}^{L} P_i\sigma_{iz}^2 \text{ is not possible.}$$

All the results developed in the sections 4 and 5 can, however, be generalised for any number of charaeters under study. Let $y_1......y_p$ are the characters under study and x be the stratifying character, their generalised variance $|G|$ is given by

$$|G| = \frac{1}{n^p}\begin{bmatrix} \sigma_{y_1}^2 & \sigma_{y_1y_2}\cdots\cdots\sigma_{y_1y_p} \\ \sigma_{y_2y_1} & \sigma_{y_2}^2\cdots\cdots\ \sigma_{y_2y_p}^2 \\ \cdots\cdots & \cdots\cdots\cdots\cdots \\ \cdots\cdots & \cdots\cdots\cdots\cdots \\ \sigma_{y_py_1} & \sigma_{y_py_2}\cdots\cdots\sigma_{y_p}{}^2 \end{bmatrix} \qquad ...(50)$$

When G is minimise with respect to x_i's it gives the same expression as given in (25) under the assumption that regression of $Y_1..............Y_p$ on X is of the same form, namely

$$E\,(Y/X)=\alpha_{yj}+\gamma_{yj}\,\phi(x) \qquad ...(51)$$

under the above condition the dispersion matrix of $Y_1......Y_p$ is given by

$$\frac{1}{n}\begin{bmatrix} E\,(\mathrm{Var}\,(Y_1/X))+\beta^2{}_{y_1}\sum_{i=1}^{L} P_i\;\sigma^2_{i\phi(x)} & E\,(\mathrm{Cov}\,(Y_1,\,Y_2/X)+\beta_{y1}\,\beta_{y2}\sum_{i=1}^{L} P_i\;\sigma^2_{i\phi(x)} \\ & E\,(\mathrm{Cov}\,(Y_1,\;Y_p/X)+\beta_{y_1}\,\beta_{y_p}\sum_{i=1}^{L} P_i\;\sigma^2_{i\phi(x)} \\ \cdots\cdots\cdots & \cdots\cdots\cdots \\ \cdots\cdots\cdots & \cdots\cdots\cdots \\ E\,(\mathrm{Cov}\,(Y_1,\,Y_p)/X))+\beta_{y_1}\,\beta_{y_p}\sum_{i=1}^{L} P_i\;\sigma^2_{i\phi(x)} & E\,(\mathrm{Cov}\,(Y_p\,Y_2/X)+\beta_{y_2}\,\beta_{y_p}\sum_{i=1}^{L} P_i\;\sigma^2_{i\phi(x)} \\ & E\,(\mathrm{Var}\,(Y_p/X))\;\beta^2_{y_p}\sum_{i=1}^{L} P_i\;\sigma^2_{i\phi(x)} \end{bmatrix} \qquad \ldots(52)$$

In case we choose points of stratification on X such that

$$\sum_{i=1}^{L} P_i \, \sigma^2_{i\phi(x)}$$

is minimised, then

$$\sum_{i=1}^{L} P_i \, \sigma^2_{iy1} \ldots\ldots \sum_{i=1}^{L} P_i \, \sigma^2_{iyp}$$

are simultaneously minimised as also in the concentration ellipsoid of the estimates $y_1, \ldots\ldots y_p$ is the smallest among the class of stratified sample estimates based on any other stratification on the character X.

7. Particular Cases

In this section points of stratification in special cases will be obtained.

7.1. If the joint distribution of X, Y and Z is normal, then regression of both Y and Z on X is linear

$$E\,((Y/X)x) = \mu_y + \beta_{yx}\,(x - \sigma_x)$$

$$E\,((Z/X)x) = \mu_z + \beta_{zx}\,(x - \mu_x)$$

and the optimum boundary points on the character X are given by

$$x_i = \frac{\mu_{i+1x} + \mu_{ix}}{2}$$

$$i = 1, 2, \ldots\ldots L-1$$

The points of stratification in this case have been tabulated by Sethi (1963) for N (0,1).

7.2. Let form of distribution of Y, Z and X be such that regression of Y and Z on X is of exponential form

$$E\,(Y/X=x) = 1 + \frac{\alpha_1}{2} - \alpha_1\, e^{-x} \qquad \ldots(53)$$

$$E\,(Z/X=x) = 1 + \frac{\alpha_2}{2} - \alpha_2\, e^{-x} \qquad \ldots(54)$$

$$\alpha_1, \alpha_2 \neq 0$$

$$-1 \leqslant \alpha_1, \alpha_2 \leqslant 1$$

The optimum strata boundary points based on character X are given by

$$e^{-x_i} = \frac{1}{2}\left[e^{-x_{i-1}} + e^{-x_{i+1}} \right] i=1, 2,\ldots\ldots, L-1 \quad \ldots(55)$$

In words the inverse of the exponential of strata boundaries follows an arithmetical progression. The strata boundaries in this case for different values of L are given in Table 1.

7.3. If the joint distribution of Y, Z and X is uniform such that the regression Y and Z on X is of same form and is given by

$$E\,(Y/X=x) = \frac{b+a}{2} + \frac{\alpha_1 x}{3} - \frac{(b+a)\,(b-a)\,\alpha_1}{3} \quad \ldots(56)$$

$$E\,(Z/X=x) = \frac{b+a}{2} + \frac{\alpha_2 x}{3} - \frac{(b+a)\,(b-a)\,\alpha_2}{3} \quad \ldots(57)$$

$$\alpha_1,\ \alpha_2 \neq 0$$

$$-1 \leqslant \alpha_1,\ \alpha_2 \leqslant 1$$

and optimum boundary points on the character X are given by

$$x_i = \frac{x_{i-1}+x_{i+1}}{2} \qquad i=1, 2,\ldots\ldots, L-1 \quad \ldots(58)$$

In words, in this case, the strata boundaries are in an arithmetical progression, *i.e.*, equal strata width. The boundary points for different values of L are tabulated in Table II.

TABLE 1

l.	*Strata boundary points*	*Var (Y)*	*Var (Z)*
2.	0.69	$\frac{1}{n}\left(1-\frac{\alpha_1^2}{16}\right)$	$\frac{1}{n}\left(1-\frac{\alpha_2^2}{16}\right)$
3.	0.40, 1.10	$\frac{1}{n}\left(1-\frac{2\alpha_1^2}{27}\right)$	$\frac{1}{n}\left(1-\frac{\alpha_2^2}{27}\right)$
4.	0.29, 0.69, 1.39	$\frac{1}{n}\left(1-\frac{5\alpha_1^2}{64}\right)$	$\frac{1}{n}\left(1-\frac{5\alpha_2^2}{64}\right)$
5.	0.22, 0.51, 0.92, 1.61	$\frac{1}{n}\left(1-\frac{2\alpha_1^2}{25}\right)$	$\frac{1}{n}\left(1-\frac{5\alpha_2^2}{64}\right)$
L	$\log_e L - \log_e L-1$	$\log_e L - \log_e L-2, \ldots\ldots\ldots\ldots\ldots, \log_e L$	

TABLE 2

$a=1,\ b=2$

L	Strata boundary points	Var (Y)	Var (Z)
2.	1.5	$\frac{1}{12n}\left(1-\frac{\alpha_1^2}{12}\right)$	$\frac{1}{12n}\left(1-\frac{\alpha_2^2}{12}\right)$
3.	1.33, 1.66	$\frac{1}{12n}\left(1-\frac{8\alpha_1^2}{81}\right)$	$\frac{1}{12n}\left(1-\frac{8\alpha_2^2}{81}\right)$
4.	1.25, 1.50, 1.75	$\frac{1}{12n}\left(1-\frac{15\alpha_1^2}{144}\right)$	$\frac{1}{12n}\left(1-\frac{15\alpha_2^2}{144}\right)$
5.	1.20, 1.40, 1.60, 1.80	$\frac{1}{12n}\left(1-\frac{18\alpha_1^2}{225}\right)$	$\frac{1}{12n}\left(1-\frac{108\alpha_2^2}{225}\right)$
L	$1.0+\frac{1}{L},\ 1.0+\frac{2}{L},\ 1.0+\frac{3}{L},\ldots\ldots,1.0+\frac{L-1}{L}$		

SUMMARY

Optimum stratification which minimises the generalised variance of estimates of the means of more than one character, say $y_1,\ldots\ldots y_p$, under study based on an auxiliary character X has been studied under the assumption of proportionate allocation. It has been shown that if the regression of each of Y_i $(i=1,\ldots\ldots p)$ on X is of the same form, the points of stratification based on X yields estimates which has the smallest concentration ellipsoid among all the points of stratification based on X.

REFERENCES

[1] Dalenius T. (1950) : 'The problem of optimum stratification', *Skand. Akt.* Vol. 33, pp. 203-213.

[2] Dalenius T. and Gurney M. (1951) : 'The problem of optimum stratification II, *Skand Akt.* Vol. 54, pp. 133-148.

[3] Dalenius T. and Hodges, J.L. (1957) : 'The choice of stratification points', *Skand Akt.* Vol. 40, pp. 198-203.

[4] Ghosh. S.P. (1963) : 'The optimum stratification with two characters'. *Ann. Math. Stat.* Vol. 134, pp. 866-872.

[5] Sethi. V.K. (1963) : 'A note on optimum stratification of population mean, *Austral. Jour. Statist.* Vol. 5, pp. 20-33.

[6] Singh R. and Sukhatme B.V. (1969) : 'Optimum stratification'. *Ann. Inst. Statist. Maths.* Vol. 21, pp. 515-528.

[7] Singh R. and Sukhatme B.V. (1972) : 'Optimum stratification in sampling with varying probabilities'. *Ann. Inst. Statist. Maths.* Vol. 24, pp. 485-494.

[8] Singh R. and Dev Prakash (1975) : 'Optimum stratification for equal allocation'. Vol. 27, (1975) pp. 273-280.

[9] Singh R. (1975) : 'A note on optimum stratification in sampling with varying probabilities', *Austral. Jour. Statist.* Vol. 17, pp. 12-21.