A STUDY OF THE STRATIFIED RANDOM SAMPLING

BY HIROJIRO AOYAMA

(Received May 5, 1954)

CONTENTS

1 Introduction

Main problems in statistical surveys are the assignment of the population, i.e., fixing the sampling field, and the estimation of the universe parameter, such as the arithmetic mean, from data. Once the population is fixed, we can estimate its mean (which equals to the universe mean in value) from the sample mean or test statistical hypotheses in regard to some parameters. But for the present these estimation and testing are not always treated effectively from the same points of view. For example, we adopt always the method of stratification of the universe for the improvement of accuracy of estimation, but even in such cases we are sometimes obliged to use the method of testing hypothesis which is valid only for simple random sampling. In this paper we shall treat the problems concerning stratified random sampling from practical points of view.

2 Stratification for the improvement of accuracy

Suppose there is a universe of size N, each element of which has a characteristic X. We take a sample of size n from this universe by the simple random sampling. Let $x_1, x_2, \ldots, x_n$ be a sample of size n. Then the sample mean

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i$$

is an unbiased estimate of the population mean $\bar{X}$, and the variance of $\bar{x}$ is

$$D^2(\bar{x})_{\text{ran.}}=\frac{N-n}{N-1}\frac{\sigma^2}{n} \tag{2.1}$$

where σ^2 is the population variance and the suffix indicates the simple random sampling.

The improvement of accuracy, which means to make $D^2(\bar{x})_{\text{ran.}}$ smaller, is achieved by stratification, that is, by grouping homogeneous elements into R groups in regard to the characteristic. We call each of these groups "stratum" of size N_i when it contains N_i elements, where the suffix indicates the group and runs over $(1, 2, \ldots, R)$.

From each stratum we take a sample of size n_i by the simple random sampling ($i=1, 2, \ldots, R$) and define the sample mean as

$$\bar{x}=\frac{1}{N}\sum_{i=1}^{R} N_i\bar{x}_i \tag{2.2}$$

where $\bar{x}_i=\frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$. Then the variance of this sample mean is

$$D^2(\bar{x})_{\text{st.}}=\frac{1}{N^2}\sum_{i=1}^{R} N_i^2\frac{N_i-n_i}{N_i-1}\frac{\sigma_i^2}{n_i} \tag{2.3}$$

where σ_i^2 is the population variance in the i-th stratum and the suffix st. indicates that the variance is that in the case when stratification is carried out.

We get almost always

$$D^2(\bar{x})_{\text{ran.}}>D^2(\bar{x})_{\text{st.}} \tag{2.4}$$

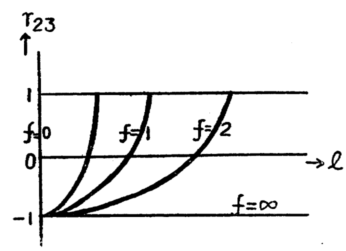
hence we can improve the accuracy of estimation by means of stratification.

2.1 Selection of controls for stratification

The effectiveness of stratification is such as mentioned above, but we must consider by what control we should perform stratification. Of course we cannot use the characteristic X, about which we have not any knowledge or knowledge enough to make use of before the survey. Usually we may, however, have some characteristics which correlate highly with X. Let these characteristics be $Y, Z, \ldots, W$. When we know only one of these characteristics, we take that one. But when we know more than two of them and we are allowed only to use two of them for the sake of simpleness of the procedure of stratification, we meet the problem "which two should we choose?"

Now suppose we have chosen two characteristics Y, Z, and let ρ be the multiple correlation coefficient of X to Y and Z, and r_{12}, r_{13} and r_{23} the simple correlation coefficients between X and Y, X and Z, and Y and Z, respectively. Then we have

$$\rho^2=\frac{r_{12}^2+r_{13}^2-2r_{12}r_{13}r_{23}}{1-r_{23}^2}=f \quad \text{(say)} \qquad (2.1.1)$$

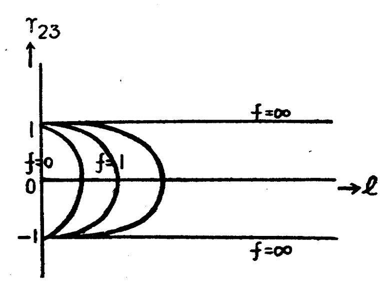


If we can get the greatest improvement of accuracy by means of two controls Y and Z, the value f then attain the maximum value. From what values of r_{12}, r_{13}, and r_{23} can we get the maximum value f ?

Put $r_{12}=l\cos\theta$, $r_{13}=l\sin\theta$ (2.1.2)

Then we have

$$f=l^2\frac{1-kr_{23}}{1-r_{23}^2} \qquad (2.1.3)$$

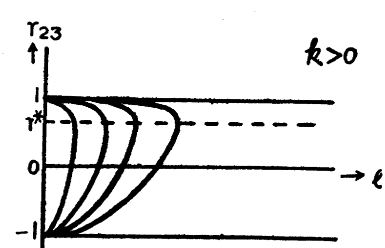


where $k=\sin 2\theta$. The right-hand illustrations show relation (2.1.3).

(i) When $k=1$

$$f=\frac{l^2}{1+r_{23}}$$

Hence the niveau lines $f=$const. are parabolas.



$$r^*=\frac{1}{k}-\sqrt{\frac{1}{k^2}-1}$$

(ii) When $k=-1$

$$f=\frac{l^2}{1-r_{23}}$$

Similar to case (i)

(iii) When $k=0$

$$f=\frac{l^2}{1-r_{23}^2}$$

Hence the niveau lines $f=$const. are ellipses.

(iv) General cases.

$$f=l^2\frac{1-kr_{23}}{1-r_{23}^2}$$

If $k \gtrless 0$, we have niveau curves with the maximum value l at the values of r_{23} near 1 or -1.

Hence when $k=\pm 1$, that is, $r_{12}=\pm r_{13}$, f attains its maximum value at $r_{23}=\mp 1$; otherwise f becomes larger according as r_{23} comes near ± 1.

Table 1.

r_{12}	r_{13}	r_{23}	$f=\rho^2$	ρ	k
.4	.4	.9 .4 .0 $-$.4	.1684 .2286 .3200 .5333	.41 .48 .57 .72	1
.4	.2	.9 .4 .0 $-$.4	.2947 .1619 .2000 .3143	.54 .40 .45 .55	$\neq 0, \pm 1$
.4	.0	$\pm$.9 $\pm$.4 .0	.8421 .1905 .1600	.92 .44 .40	0

Table 1 shows these circumferences. As is seen from these figures the correlation coefficient between Y and Z must be small, when Y and Z have the equal correlation coefficients with X, and on the contrary, it must be large when Y and Z have the different correlation coefficients with X.

If there are many characteristics which are able to be taken as controls for stratification, the conditions for adopting some of $Y, Z, \dots W$, become much complicated. If three controls, for example, such as Y, Z and W are adopted, we have the increase of the square of the multiple correlation coefficient as follows:

$$g\equiv\Delta\rho^2=\frac{\{r_{14}(1-r_{23}^2)-r_{24}(r_{12}-r_{13}r_{23})-r_{34}(r_{13}-r_{12}r_{23})\}^2}{1-r_{23}^2-r_{24}^2-r_{34}^2+2r_{23}r_{24}r_{34}} \tag{2.1.4}$$

Here the niveau surface $g=$const. becomes an ellipsoidal surface with variables r_{14}, r_{24} and r_{34}, and g attains its maximum value for many different combinations of r_{14}, r_{24} and r_{34}.

When $n-1$ controls have been already adopted and one control is newly adopted, the increase of the square of the multiple correlation coefficient is

$$g\equiv\Delta\rho^2=\frac{\{\alpha r_{1n}+L(r_{2n}, r_{3n}, \dots, r_{n-1,n})\}^2}{\beta+Q(r_{2n}, r_{3n}, \dots, r_{n-1,n})} \tag{2.1.5}$$

where α, β are constants, and L, Q a linear formula and a quadratic formula with regard to $r_{1n}, r_{2n}, \ldots, r_{n-1,n}$, respectively. So, the surface $g=$const. becomes a hyperellipsoidal surface.

2.2 Optimum stratification by means of one control

Let us consider the case when we stratify the units of the population by means of one control Y which is highly correlated with the characteristic X. Let the distribution function $\Phi(y)$ of Y be known. Then we divide the population into R strata from which we take sample by the proportionate sampling, and we want to estimate the total value of X. How should we stratify the population in order to minimize the variance of the estimate of the total value? The following idea comes from the paper of Hayashi and Maruyama [1]. In this case the variance of the estimated total value x is

$$D^2(x)=\sum_{i=1}^{R} N_i^2 \frac{N_i-n_i}{N_i-1} \frac{\sigma_{ix}^2}{n_i}, \tag{2.2.1}$$

where N_i, n_i and σ_{ix}^2 denote the population size, the sample size and the population variance of the i-th stratum, respectively. Because of the proportionate sampling, neglecting the finite population correction, (2.2.1) becomes

$$D^2(x)=\frac{N}{n}\sum_{i=1}^{R} N_i \sigma_{ix}^2. \tag{2.2.2}$$

If a linear regression $x=\alpha' y+\beta'$ holds between X and Y, we may put approximately with a constant α

$$\sigma_x^2 = \alpha^2 \sigma_y^2, \tag{2.2.3}$$

therefore we may put also

$$\sigma_{ix}^2 = \alpha^2 \sigma_{iy}^2. \tag{2.2.4}$$

Let $\overline{Y}_i$, σ_y^2 and $\overline{Y}$ be the population mean in the i-th stratum, the population variance and the population mean, respectively, and y_i the cutting points of these strata. That is,

$$\left.\begin{aligned} \sigma_y^2 &= \int_{-\infty}^{\infty} (y-\overline{Y})^2 d\Phi(y) \\ \overline{Y} &= \int_{-\infty}^{\infty} y d\Phi(y) \\ \overline{Y}_i &= \int_{y_{i-1}}^{y_i} y d\Phi(y)/(\Phi(y_i)-\Phi(y_{i-1})) \end{aligned}\right\} \tag{2.2.5}$$

From (2.2.3), (2.2.4) and the relation

$$N\sigma_y^2=\sum_i N_i\sigma_{iy}^2+\sum_i N_i(\overline{Y}_i-\overline{Y})^2 \tag{2.2.6}$$

follows

$$D^2(x)=\frac{N\alpha^2}{n}\sum_i N_i\sigma_{iy}^2=\frac{N\alpha^2}{n}\{N\sigma_y^2-\sum_i N_i(\overline{Y}_i-\overline{Y})^2\}$$

$$=\frac{N\alpha^2}{n}\{N\sigma_y^2+N\overline{Y}^2-\sum_i N_i\overline{Y}_i^2\}\ . \tag{2.2.7}$$

Hence, to minimize $D^2(x)$ is to maximize

$$f=\sum_{i=1}^{R} N_i\overline{Y}_i^2=N\sum_i\Big(\int_{y_{i-1}}^{y_i} y d\Phi(y)\Big)^2\Big/\Big(\Phi(y_i)-\Phi(y_{i-1})\Big)\,, \tag{2.2.8}$$

where $y_0=-\infty$, $y_R=+\infty$. When the number of these strata is R, we have from $\dfrac{\partial f}{\partial y_i}=0$

$$y_i=\frac{\overline{Y}_i+\overline{Y}_{i+1}}{2},\quad i=1,2,\ldots,R-1. \tag{2.2.9}$$

When the number of these strata increases, $D^2(x)$ decreases, as can easily be seen from the relation

$$N\sigma^2\geqq N_1\sigma_1^2+N_2\sigma_2^2 \qquad (N=N_1+N_2)$$

The positions of cutting points y_i are obtained by means of successive approximation, but the relative positions are as follows. From (2.2.9) we have

$$2y_i=\frac{\int_{y_{i-1}}^{y_i} y\,d\Phi}{\Phi(y_i)-\Phi(y_{i-1})}+\frac{\int_{y_i}^{y_{i+1}} y\,d\Phi}{\Phi(y_{i+1})-\Phi(y_i)}$$

and by the mean value theorem

$$(y_i-y_{i-1})(1-\theta)=(y_{i+1}-y_i)\theta' \tag{2.2.10}$$

where $0<\theta<1$, $0<\theta'<1$, or if we can assume the existence of Φ', we have approximately

$$(y_i-y_{i-1})\frac{\Phi'(y_{i-1})}{2\Phi'(y_{i-1}^*)}=(y_{i+1}-y_i)\Big(1-\frac{\Phi'(y_i)}{2\Phi'(y_i^*)}\Big) \tag{2.2.11}$$

where $y_{i-1}<y_{i-1}^*<y_i<y_i^*<y_{i+1}$. If the variation of the density function $\Phi'(y)$ is assumed to be small, we have

$$y_i-y_{i-1}\fallingdotseq y_{i+1}-y_i\,,$$

that is, equal intervals are approximately optimum for our purpose.

2.3 Optimum stratification by means of two controls

Let us estimate the total value Z with which two controls X and Y are highly correlated. As in section 2.2, we adopt proportional allocation and we have

$$D^2(z)=\frac{N}{n}\sum_{i=1}^{R} N_i\sigma_{iz}^2 . \tag{2.3.1}$$

Now assume the existence of a linear regression $z=\alpha' x+\beta' y+\gamma'$ and let ρ_{ixy} be the correlation coefficient between X and Y in the i-th stratum. Then we get

$$\sigma_{iz}^2=\alpha^2\sigma_{ix}^2+\beta^2\sigma_{iy}^2+2\alpha\beta\rho_{ixy}\sigma_{ix}\sigma_{iy} \tag{2.3.2}$$

where

$$\left.\begin{aligned} N\sigma_x^2&=\sum_i N_i\sigma_{ix}^2+\sum_i N_i(\overline{X}_i-\overline{X})^2\\ N\sigma_y^2&=\sum_i N_i\sigma_{iy}^2+\sum_i N_i(\overline{Y}_i-\overline{Y})^2\\ \sum_i N_i\rho_{ixy}\sigma_{ix}\sigma_{iy}&=\sum_i N_i\overline{X}_i\overline{Y}_i+N\rho_{xy}\sigma_x\sigma_y-N\overline{X}\,\overline{Y} \end{aligned}\right\} \tag{2.3.3}$$

Considering (2.3.2) and (2.3.3), we have from (2.3.1)

$$\begin{aligned} D^2(z)=\frac{N}{n}\{&\alpha^2(N\sigma_x^2+N\overline{X}^2-\sum_i N_i\overline{X}_i^2)+\beta^2(N\sigma_y^2+N\overline{Y}^2-\sum_i N_i\overline{Y}_i^2)\\ &+2\alpha\beta(N\rho_{xy}\sigma_x\sigma_y-N\overline{X}\,\overline{Y}+\sum_i N_i\overline{X}_i\overline{Y}_i)\} \end{aligned} \tag{2.3.4}$$

In order to minimize this $D^2(z)$ with fixed R we have only to maximize

$$f=\alpha^2\sum_i N_i\overline{X}_i^2+\beta^2\sum_i N_i\overline{Y}_i^2-2\alpha\beta\sum_i N_i\overline{X}_i\overline{Y}_i \tag{2.3.5}$$

Let $\Phi(x)$ and $\Psi(y)$ be the absolutely continuous distribution functions of X and Y, respectively. From $\dfrac{\partial f}{\partial x_i}=0$, $\dfrac{\partial f}{\partial y_i}=0$ it can be seen that cutting points x_i and y_i satisfy the following relations

$$x_i=\frac{\alpha(\overline{X}_i^2-\overline{X}_{i+1}^2)}{2\{\alpha(\overline{X}_i-\overline{X}_{i+1})-\beta(\overline{Y}_i-\overline{Y}_{i+1})\}} \tag{2.3.6}$$

$$y_i=\frac{\beta(\overline{Y}_i^2-\overline{Y}_{i+1}^2)}{2\{\beta(\overline{Y}_i-\overline{Y}_{i+1})-\alpha(\overline{X}_i-\overline{X}_{i+1})\}} \tag{2.3.7}$$

Eliminating α and β from these equations, we have

$$y_i=-\frac{\overline{Y}_i+\overline{Y}_{i+1}}{\overline{X}_i+\overline{X}_{i+1}}\left(x_i-\frac{\overline{X}_i+\overline{X}_{i+1}}{2}\right) \tag{2.3.8}$$

or

$$x_i = -\frac{\overline{X}_i + \overline{X}_{i+1}}{\overline{Y}_i + \overline{Y}_{i+1}}\left(y_i - \frac{\overline{Y}_i + \overline{Y}_{i+1}}{2}\right) \tag{2.3.9}$$

From (2.3.8) or (2.3.9) we can see that boundary points (x_i, y_i) lie on the straight line which passes through the points $\left(\frac{\overline{X}_i + \overline{X}_{i+1}}{2},\ 0\right)$ and $\left(0,\ \frac{\overline{Y}_i + \overline{Y}_{i+1}}{2}\right)$, respectively, as in section 2.2. Therefore, each stratum is a strip bounded by these straight lines.

2.4 Case when we adopt Neyman's method of allocation

In the above mentioned allocation we adopted the method of proportional allocation, but in this section we adopt Neyman's method of allocation. In this case the variance of the estimated total value x under the stratification by control Y is

$$D^2(x) \doteqdot \frac{1}{n}\left(\sum_i N_i \sigma_{ix}\right)^2 \tag{2.4.1}$$

As in section 2.2, we assume the existence of a linear regression between X and Y. Then

$$D^2(x) \doteqdot \frac{\alpha^2}{n}\left(\sum_i N_i \sigma_{iy}\right)^2 \tag{2.4.2}$$

and we can find cutting points y_i which minimize

$$f = \sum_i N_i \sigma_{iy} = N \sum_i \sqrt{(\Phi(y_i) - \Phi(y_{i-1}))\int_{y_{i-1}}^{y_i} (y - \overline{Y}_i)^2 d\Phi} \tag{2.4.3}$$

From $\frac{\partial f}{\partial y_i} = 0$ we get y_i satisfying the relation

$$\sigma_{iy}\left\{1 + \left(\frac{y_i - \overline{Y}_i}{\sigma_{iy}}\right)^2\right\} = \sigma_{i+1,y}\left\{1 + \left(\frac{y_i - \overline{Y}_{i+1}}{\sigma_{i+1,y}}\right)^2\right\} \tag{2.4.4}$$

which have been got by Dalenius, [2] from another point of view. By the mean value theorem (2.4.4) becomes

$$(\theta'' - \theta)(y_i - y_{i-1})\left(1 + \left(\frac{1-\theta}{\theta'' - \theta}\right)^2\right) = (\theta''' - \theta')(y_{i+1} - y_i)\left(1 + \left(\frac{\theta'}{\theta''' - \theta'}\right)^2\right) \tag{2.4.5}$$

where $0 < \theta,\ \theta',\ \theta'',\ \theta''' < 1$. If we can assume the variation of $\Phi'(y)$ is small in each stratum, we have

$$y_i - y_{i-1} \doteqdot y_{i+1} - y_i$$

which means that the intervals of the strata are nearly of equal size. When the intervals of the strata become smaller, we can easily prove that $\theta, \theta' \rightarrow \frac{1}{2}$, and $\theta''-\theta,\ \theta'''-\theta' \rightarrow \sqrt{3}/6$, therefore we obtain intervals of equal size as before.

2.5 General cases

Let us estimate the total value X as in the previous section. Select Y as the control of the stratification and we have

$$D^2(x) \doteqdot \sum_i N_i^2 \frac{\sigma_{ix}^2}{n_i} = \alpha^2 \sum_i N_i^2 \frac{\sigma_{iy}^2}{n_i}$$
$$= \alpha^2 N^2 \sum_i \frac{\Phi(y_i)-\Phi(y_{i-1})}{n_i} \int_{y_{i-1}}^{y_i} (y-\overline{Y}_i)^2 d\Phi \tag{2.5.1}$$

In order to obtain y_i which minimize this $D^2(x)$, we get from $\frac{\partial D^2}{\partial y_i}=0$

$$\frac{N_i}{n_i}(\sigma_{iy}^2+(y_i-\overline{Y}_i)^2) = \frac{N_{i+1}}{n_{i+1}}(\sigma_{i+1,y}^2+(y_i-\overline{Y}_{i+1})^2) \tag{2.5.2}$$

From this equation we can calculate y_i by means of successive approximation.

3 Influence of the stratification to the estimate

In the ordinary sampling survey we are obliged to take only a few controls, for we are ignorant of the controls or of complicated procedures. In such cases we perform stratification at first by one control B, and after we have got data, we stratify anew by another control A. But owing to the latter stratification we have a biased estimate.

For example, consider the case when we stratify the schools by the number of pupils in the school survey where we want to estimate certain parameters in the groups which are stratified by the number of classes.

In this case, suppose we want to estimate the population mean in regard to a characteristic X. At first we stratify the population by control B and take samples by the proportionate sampling. Then we stratify these samples anew by control A. Hence the formulas of double sampling can be used. Sample mean $\bar{x}$, which is an estimate of the population mean $\overline{X}$, is

$$\bar{x} = \frac{1}{n}\sum_j \sum_k x_{jk} = \frac{1}{n}\sum_j \sum_k \sum_i x_{ijk} \tag{3.1}$$

where j denotes the j-th stratum owing to control A, i denotes the i-th stratum owing to control B and n is the sample size. In this case $\bar{x}$ is of course an unbiased estimate of the population mean, that is,

$$E(\bar{x})=E\Big(\frac{1}{n}\sum_i\sum_k x_{ik}\Big)=\frac{1}{n}\sum_i\frac{n_i}{N_i}\sum_{k=1}^{N_i}X_{ik}=\frac{1}{N}\sum_i\sum_k X_{ik}=\bar{X} \tag{3.2}$$

and the variance of $\bar{x}$ is

$$D^2(\bar{x})=\frac{\sigma^2}{n}-\frac{1}{n}\sum_i P_i(\bar{X}_i-\bar{X})^2 \tag{3.3}$$

where $P_i=N_i/N$, which was mentioned by the author in [3].

The estimated mean in each stratum in regard to control B is unbiased, that is, when

$$\bar{x}_i=\frac{1}{n_i}\sum_j\sum_k x_{ijk}=\frac{1}{n_i}\sum_k x_{ik} \tag{3.4}$$

we have

$$E(\bar{x}_i)=\bar{X}_i \tag{3.5}$$

But the estimated mean

$$\bar{x}_j=\frac{1}{n_j}\sum_i x_{ij}=\frac{1}{n_j}\sum_i\sum_k x_{ijk} \tag{3.6}$$

in each new stratum owing to control A is biased.

In general, for two random variables X and Y, it holds

$$\frac{X}{Y}=\frac{E(X)}{E(Y)}+\frac{XE(Y)-YE(X)}{E^2(Y)}+R \tag{3.7}$$

where R denotes the error term. We have first

$$E(\sum_k\sum_i x_{ijk})=\sum_i n_{ij}\sum_{(n_{ij}\cdots n_{ij''})}P_{ij}^{n_{ij}}P_{ij'}^{n_{ij'}}\cdots P_{ij''}^{n_{ij''}}\frac{n_i!}{n_{ij}!\,n_{ij'}!\cdots n_{ij''}!}\cdot\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}}X_{ijk}$$

$$=\sum_i\frac{n_i}{N_i}X_{ijk}=\frac{n}{N}X_j \tag{3.8}$$

where $P_{ij}=N_{ij}/N_i$ and the summation runs over all combinations $(n_{ij}\cdots n_{ij''})$ which satisfy $n_{ij}+\cdots+n_{ij''}=n_i$.

In a similar manner we have

$$E(n_j)=\frac{n}{N}N_j \tag{3.9}$$

Hence from (3.7), (3.8) and (3.9) we have in the sense of the first approximation

$$E(\bar{x}_j) \doteqdot \frac{E(\sum_k \sum_i x_{ijk})}{E(n_j)} = \frac{X_j}{N_j} = \bar{X}_j \tag{3.10}$$

and

$$E(R) \doteqdot \frac{E(X)}{E^3(Y)}(E(Y^2) - E^2(Y)) = \frac{E(X)D^2(Y)}{E^3(Y)} \tag{3.11}$$

Similarly, we have

$$D^2(n_{ij}) = E(n_{ij}^2) - E^2(n_{ij}) = \frac{nN_j}{N} - \frac{n}{N}\sum_i \frac{N_{ij}^2}{N_i} \tag{3.12}$$

therefore, in our case, from (3.11)

$$E(R) \doteqdot \frac{N}{n}\frac{\bar{X}_j}{N_j}\left(1 - \sum_i \frac{N_{ij}^2}{N_i N_j}\right) = O\left(\frac{1}{n}\right) \tag{3.13}$$

Hence (3.10) becomes

$$E(\bar{x}_j) = \bar{X}_j + O\left(\frac{1}{n}\right) \tag{3.14}$$

As for the variance of $\bar{x}_j$, we have in general

$$D^2\left(\frac{X}{Y}\right) \doteqdot \frac{E\left(X - \frac{E(X)}{E(Y)}Y\right)^2}{E^2(Y)} \tag{3.15}$$

If we can assume that the correlation coefficient between x_j and n_j is 0 (this assumption is moderate when the correlation coefficient is positive), we have

$$D^2(\bar{x}_j) = \frac{1}{N_j^2}\sum_i \frac{N_{ij}^2 \sigma_{ij}^2}{n_{ij}} + \frac{N}{nN_j^2}\sum_i N_{ij}\left(1 - \frac{N_{ij}}{N_i}\right)\left(\bar{X}_{ij}^2 + \bar{X}_j^2 + \frac{\sigma_{ij}^2}{n_{ij}}\right) \tag{3.16}$$

Further, we must increase the sample size in order to get approximately unbiased estimate when we adopt a new stratification after the sampling procedure.

4 Practical systematic sampling

Systematic sampling method which has been introduced by Madow, W. G. and Madow, L. H. [4] is a very convenient procedure for many practical sampling surveys. In our country we ordinarily use this procedure in large sample survey where the lists of the universe elements are the voters' list, the residents' list and so on. But these lists contain often non-universe elements in which we are not interested.

We have treated in [5] the problem about this circumstance. To repeat here what we have done in [5], they are as follows:

1. For samples of size n by systematic sampling from the population of size $N \neq kn$, for instance $N=kn-r$, $0<r<k$, $n>50$, we have obtained the expectation, the variance and the mean square error of sample mean $\bar{x}$.
2. For the frame including non-universe elements and with size $N=kn$, we have obtained the expectation of the variance of sample mean in regard to the randomization of the universe.
3. To the case when we should take a sample of size exactly n, and when non-universe elements enter the sample obtained, we have got a better estimate $\bar{x}$ by substituting suitable universe elements for the non-universe elements.

Now we want to prove a more general theorem than that in [5].

THEOREM *Consider the things of two sorts, the numbers of which are Np and Nq, respectively, and divide these things into k sets of equal size. Put*

$$N=Np+Nq \tag{4.1}$$

and

$$M=N/k \tag{4.2}$$

If we denote by x_i the number of things concerning p in the i-th set, we have

$$E(x_i)=Mp \tag{4.3}$$

and

$$D^2(x_i)=\frac{M(M-1)Np(Np-1)}{N(N-1)}+Mp-(Mp)^2 \tag{4.4}$$

PROOF: If p_i is the percentage of the elements of (p) in the i-th set, we have a system of percentages $p_1, \ldots, p_k$ with the probability

$$\Pr\{p_1, \ldots, p_k\}=\frac{\binom{Np}{Mp_1}\binom{Nq}{Mq_1}\binom{Np-Mp_1}{Mp_2}\binom{Nq-Mq_1}{Mq_2}\cdots\binom{Mp_k}{Mp_k}\binom{Mq_k}{Mq_k}}{\binom{N}{M}\binom{N-M}{M}\binom{N-2M}{M}\cdots\binom{M}{M}}$$

$$=\frac{\prod_{i=1}^{k}\binom{M}{Mp_i}}{\binom{N}{Np}} \tag{4.5}$$

Then putting $Mp_i=x_i$, we have

$$E\{x_i\}=\frac{M}{\binom{N}{Np}}\sum{}'\binom{M}{x_1}\cdots\binom{M-1}{x_i-1}\cdots\binom{M}{x_k}=Mp, \tag{4.6}$$

where $\sum'$ denotes the summation over all x_i with the condition $x_1+\cdots+(x_i-1)+x_{i+1}+\cdots+x_k=Np$ and also

$$E\{x_i(x_i-1)\}=\frac{M(M-1)Np(Np-1)}{N(N-1)} \tag{4.7}$$

hence we have

$$D^2(x_i)=E\{x_i^2\}-E^2\{x_i\}=\frac{M(M-1)Np(Np-1)}{N(N-1)}+Mp-(Mp)^2 \tag{4.8}$$

Equations (4.6) and (4.7) are also obtained by the approximation formula of the (4.5) as follows.

If we can assume that p and p_i are very small but Np and Mp_i are constant for large N and M, we are able to use the formula of Poisson distribution, that is,

$$\binom{M}{Mp_i}p^{Mp_i}p^{Mq_i}\sim\frac{e^{-Mp}(Mp)^{Mp_i}}{(Mp_i)\,!}$$

and

$$\binom{N}{Np}p^{Np}q^{Nq}\sim\frac{e^{-Np}(Np)^{Np}}{(Np)\,!}$$

Therefore, we have approximately

$$\Pr\{p_1,\ldots,p_k\}\sim\frac{(Np)\,!}{(Mp_1)\,!\,(Mp_2)\,!\cdots(Mp_k)\,!}\left(\frac{1}{k}\right)^{Np} \tag{4.9}$$

and

$$E(x_i)=Mp\,, \tag{4.10}$$

$$E\{x_i(x_i-1)\}=\left(\frac{M}{N}\right)^2Np(Np-1)\,. \tag{4.11}$$

Therefore, in the section 2 of [5] if N' is the size of universe elements, and N_i' is that of universe elements of the i-th column, we have

$$E(N_i')=N'/k=\bar{N}'\,, \tag{4.12}$$

$$D^2(N_i')=\frac{(N-k)\bar{N}'\left(\bar{N}'-\frac{1}{k}\right)}{N-1}+\bar{N}'-\bar{N}'^2$$

$$=\bar{N}'\left(\frac{N-k}{N-1}\,\frac{\bar{N}'k-1}{k}+1-\bar{N}'\right) \tag{4.13}$$

For the example of voters' lists of 4-chome, Ikebukuro, which was treated in the example of [5],

$$N'=2476=096\,N$$

$$\therefore\quad \frac{\tau^2}{N'^2}=\left(\frac{2580-60}{2580-1}\,\frac{2476-1}{60}+1-41.3\right)\frac{1}{41.3}=0.000153\,.$$

5 Method of the analysis in the stratified random sampling

In the stratified random sampling we always estimate the population parameters from the weighted sample data which are got by means of simple random sampling from each stratum, so that we shall have often biased estimates of the population parameters. Nevertheless, we can get the linear unbiased estimate of the population mean because of the special construction of estimators. Often we want to estimate other population parameters such as correlation coefficients, or to test the hypothesis by means of the chi-square distribution. In such cases, however, we must establish the analyzing formula for the stratified random sampling.

If we have much information in every stratum, we need not the formulation such as in what follows, but we have never much information enough to make conclusions on the whole without using the precise information in every stratum.

Now we divide the universe elements into R strata at random, and denote the expectation in this randomization by the notation ε and distinguish it from the ordinary expectation E in the sampling field of each stratum. Then we obtain a modified Tchebycheff's inequality as follows.

THEOREM (TCHEBYCHEFF'S INEQUALITY) *Let X be a random variable with mean $E(X)$ and variance $D^2(X)$. Divide the universe elements into R strata at random. We denote expectation and variance concerning the latter randomization by ε and ϑ^2, respectively. Then for any $k>0$ it holds*

$$\Pr\{|X-\varepsilon E(X)|\geqq k\sqrt{D^2(X)+\{E(X)-\varepsilon E(X)\}^2}\}\leqq\frac{1}{k^2} \qquad (5.1)$$

PROOF: By the relation

$$X-\varepsilon E(X)=X-E(X)+E(X)-\varepsilon E(X)$$

we can easily prove the theorem as the ordinary Tchebycheff's inequality.

If we can here put $\varepsilon E(X)=E(X)$, that is, $E(X)$ is unrelated to the stratification, we have the ordinary Tchebycheff's inequality.

COROLLARY *We have approximately from (5.1)*

$$\Pr\{|X-\varepsilon E(X)|\geqq k\sqrt{\varepsilon D^2(X)+\vartheta^2 E(X)}\}\leqq\frac{1}{k^2} \tag{5.2}$$

or

$$\Pr\{|X-\varepsilon E(X)|\geqq k\sqrt{\varepsilon^* D^2(X)}\leqq\frac{1}{k^2} \tag{5.3}$$

with

$$\varepsilon^* D^2(X)=\varepsilon D^2(X)+\vartheta^2 E(X) \tag{5.4}$$

PROOF: Taking the expectation ε in the root sign of (5.1) we have (5.2).

We use practically this method of approximation when estimating variances.

For the sample mean $\bar{x}$ in the stratified random sampling, we have

$$E(\bar{x})=\frac{1}{N}\sum_{i=1}^{R}N_i\bar{X}_i=\bar{X}$$

which is unrelated to the stratification, so that

$$\varepsilon E(\bar{x})=E(\bar{x}) \tag{5.5}$$

and we can use the ordinary Tchebycheff's inequality.

If we wish to prove (5.5) directly, we may proceed as follows: in the randomization in regard to the stratification we have

$$\frac{N!}{N_1!\,N_2!\cdots N_R!}=k \qquad \text{(say)} \tag{5.6}$$

ways of partition, hence

$$\varepsilon(\bar{X}_i)=\varepsilon\left(\frac{1}{N_i}\sum_{j=1}^{N_i}X_{ij}\right)=\frac{1}{N_i}\sum_{i=1}^{k}\frac{1}{k}X_{(i)}=\frac{1}{kN_i}\frac{XkN_i}{N}=\frac{X}{N}=\bar{X}$$

where $X=X_1+X_2+\cdots+X_N$, $X_{(i)}$ is a sum of X in the i-th stratum in regard to these randomization. Therefore, we have

$$\varepsilon E(\bar{x})=\varepsilon\left(\frac{1}{N}\sum_{i=1}^{R}N_i\bar{X}_i\right)=\frac{1}{N}\sum_{i=1}^{R}N_i\varepsilon(\bar{X}_i)=\frac{1}{N}\sum_{i=1}^{R}N_i\bar{X}=\bar{X}.$$

6 Chi-square test for goodness of fit

We have treated in [6] the chi-square test for goodness of fit, and obtained the following result.

We divide the universe of size N into R strata by certain controls, the sizes of which are $N_1, N_2, \ldots, N_R$, respectively, and take samples of size $n_1, n_2, \ldots, n_R$ from these strata by the simple random sampling, respectively. Put

N_{ij}: the number of universe elements with the j-th category of a certain characteristic in the i-th stratum

$N_{(j)}$: the number of universe elements with the j-th category of the characteristic

$n_{ij}, n_{(j)}$: the corresponding sample sizes.

Further, put

$$\chi^2=\frac{n}{N}\sum_j \frac{\left(\sum_j N_i \frac{n_{ij}}{n_i}-N_{(j)}\right)^2}{N_{(j)}}=\frac{n}{N}\sum_j \frac{\left(\sum_i N_i \frac{n_{ij}}{n_i}\right)^2}{N_{(j)}}-n \tag{6.1}$$

Then, neglecting the finite population correction, we have

$$E(\chi^2)=\frac{n}{N}\sum_i\sum_j \frac{N_{ij}(N_i-N_{ij})}{n_i N_{(j)}} \tag{6.2}$$

The variance of χ^2 is represented by the equation (3) in [6].

As to the proportionate sampling we obtain

$$E(\chi^2)_{\text{prop.}}=M-\sum_i\sum_j \frac{N_{ij}^2}{N_i N_{(j)}} \tag{6.3}$$

If we can assume that the stratification into R strata is carried out at random, that is, the stratification of $N_{(j)}$ elements into classes of size N_{ij} $(i=1, 2, \ldots, R)$ is carried out at random, the expectation ε of χ^2 with respect to this randomization is

$$\varepsilon E(\chi^2)_{\text{prop.}} \doteqdot M-1-\frac{(M-1)(R-1)}{N} \tag{6.4}$$

For the expectation ε we apply the formulas with respect to factorial moments in the $R\times M$ contingency table [7]

$$\left.\begin{aligned}
\varepsilon(N_{ij}^{(l)}) &= \frac{N_i^{(l)}N_{(j)}^{(l)}}{N^{(l)}} \\
\varepsilon(N_{ij}^{(l)}N_{i'j'}^{(m)}) &= \frac{N_i^{(l+m)}N_{(j)}^{(l)}N_{(j')}^{(m)}}{N^{(l+m)}}, && \text{for } i=i',\ j\neq j' \\
&= \frac{N_i^{(l)}N_{i'}^{(m)}N_{(j)}^{(l+m)}}{N^{(l+m)}}, && \text{for } i\neq i',\ j=j' \\
&= \frac{N_i^{(l)}N_{i'}^{(m)}N_{(j)}^{(l)}N_{(j')}^{(m)}}{N^{(l+m)}}, && \text{for } i\neq i',\ j\neq j'
\end{aligned}\right\} \tag{6.5}$$

In the general case with the assumption $N_i=N/R$ we have

$$\mathcal{E}E(\chi^2) \doteqdot \frac{n(M-1)}{R^2}\sum_i \frac{1}{n_i} \tag{6.6}$$

$$\mathcal{E}D^2(\chi^2)_{\text{prop.}}=2(M-1)+O\left(\frac{1}{N}\right)+O\left(\frac{1}{n}\right) \tag{6.7}$$

We can generally obtain an approximate test by a modified Tchebycheff's inequality mentioned in section 5

$$\Pr\{|\chi^2-\mathcal{E}E(\chi^2)| \geqq k\sqrt{\mathcal{E}^*D^2(\chi^2)}\} \leqq \frac{1}{k^2} \tag{6.8}$$

As to the proportionate sampling we obtain

$$\vartheta^2E(\chi^2)_{\text{prop.}}=\mathcal{E}\{E(\chi^2)_{\text{prop.}}-\mathcal{E}E(\chi^2)_{\text{prop.}}\}^2=O\left(\frac{1}{N^2}\right)$$

Hence, assuming $N_i=N/R$, we have

$$\mathcal{E}^*D^2(\chi^2)_{\text{prop.}}=\mathcal{E}D^2(\chi^2)_{\text{prop.}}+\vartheta^2E(\chi^2)_{\text{prop.}}\sim 2(M-1)$$

and approximately

$$\Pr\{|\chi^2-(M-1)| \geqq k\sqrt{2(M-1)}\} \leqq \frac{1}{k^2} \tag{6.9}$$

7 Chi-square test for $s\times t$ contingency table

In this section we treat the chi-square test for $s\times t$ contingency table from the same point of view as in the preceding section. As before we divide the universe of size N into R strata whose sizes are $N_1, N_2 \ldots, N_R$, respectively, and take samples of $n_1, \ldots, n_R$ from these strata by the simple random sampling, respectively.

$A \backslash B$	B_1	$B_2 \cdots B_j \cdots B_t$	Total
A_1	n_{k11}	$n_{k12} \cdots n_{k1j} \cdots n_{k1t}$	$n_{k1\cdot}$
$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \vdots$	$\vdots$
A_i	n_{ki1}	$n_{ki2} \cdots n_{kij} \cdots n_{kit}$	$n_{ki\cdot}$
$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \vdots$	$\vdots$
A_s	n_{ks1}	$n_{ks2} \cdots n_{ksj} \cdots n_{kst}$	$n_{ks\cdot}$
Total	$n_{k\cdot 1}$	$n_{k\cdot 2} \cdots n_{k\cdot j} \cdots n_{k\cdot t}$	n_k

As for certain two characteristics A and B, let n_{kij} be the size of the sample with the i-th category in regard to characteristic A and the j-th category in regard to characteristic B in the k-th stratum $(i=1, 2, \ldots, s;\ j=1, 2, \ldots, t;\ k=1, 2, \ldots, R)$. Especially for the k-th stratum they are illustrated in the above table.

If we have information about the size N_{kij} we can test, as usual, whether two characteristics A and B are independent of each other by means of the χ^2-test with $(s-1)(t-1)$ degrees of freedom for each stratum and $R(s-1)(t-1)$ degrees of freedom for the whole universe. But, in this case the difference among strata is not considered. So, we want to test the independence of A and B from the point of view of weighted samples. Put

$$\chi^2=\frac{n}{N}\sum_i\sum_j\frac{\left\{\sum_k w_k n_{kij}-\frac{(\sum_k w_k n_{ki\cdot})(\sum_k w_k n_{k\cdot j})}{\sum_k w_k n_k}\right\}^2}{\frac{(\sum_k w_k n_{ki\cdot})(\sum_k w_k n_{k\cdot j})}{\sum_k w_k n_k}} \tag{7.1}$$

where $w_k=N_k/n_k$. In the proportionate sampling, which we consider in this section especially, we have

$$\chi^2=n\left\{\sum_i\sum_j\frac{(\sum_k n_{kij})^2}{(\sum_k n_{ki\cdot})(\sum_k n_{k\cdot j})}-1\right\} \tag{7.2}$$

Using the relations with respect to factorial moment:

$$E(n_{ij}^{(l)})=\frac{n_{i\cdot}^{(l)}n_{\cdot j}^{(l)}}{n^{(l)}} \tag{7.3}$$

$$\left.\begin{aligned} E(n_{ij}^{(l)}n_{i'j'}^{(m)})&=\frac{n_{i\cdot}^{(l+m)}n_{\cdot j}^{(l)}n_{\cdot j'}^{(m)}}{n^{(l+m)}}, &&\text{for } i=i',\ j\neq j' \\ &=\frac{n_{i\cdot}^{(l)}n_{i'\cdot}^{(m)}n_{\cdot j}^{(l+m)}}{n^{(l+m)}}, &&\text{for } i\neq i',\ j=j' \\ &=\frac{n_{i\cdot}^{(l)}n_{i'\cdot}^{(m)}n_{\cdot j}^{(l)}n_{\cdot j'}^{(m)}}{n^{(l+m)}}, &&\text{for } i\neq i',\ j\neq j' \end{aligned}\right\} \tag{7.4}$$

we have

$$E(\chi^2)=n\left\{\sum_i\sum_j\frac{\sum_k\frac{n_{ki\cdot}n_{k\cdot j}(n_{ki\cdot}-1)(n_{k\cdot j}-1)}{n_k(n_k-1)}}{\sum_k n_{ki\cdot}\sum_k n_{k\cdot j}}+\sum_i\sum_j\frac{\sum\sum_{k\neq k'}\frac{n_{ki\cdot}n_{k\cdot j}n_{k'i\cdot}n_{k'\cdot j}}{n_k n_{k'}}}{\sum_k n_{ki\cdot}\sum_k {}_{k\cdot j}}\right.$$

$$\left.+\sum_i\sum_j\frac{\sum_k\frac{n_{ki\cdot}n_{k\cdot j}}{n_k}}{\sum_k n_{ki\cdot}\sum_k n_{k\cdot j}}-1\right\} \tag{7.5}$$

If we divide the universe elements into R strata at random, we have in regard to this randomization

$$\varepsilon E(\chi^2)=\frac{n(s-1)(t-1)}{N}\sum_k\frac{N_k^2}{nN_k-N}+O\left(\frac{1}{n_i}\right) \tag{7.6}$$

Hence only for $R=1$ we have

$$\varepsilon E(\chi^2) \doteqdot \frac{n}{n-1}(s-1)(t-1) \tag{7.7}$$

If we can assume $N_k=N/R$ $(k=1, 2, \ldots, R)$, we have

$$\varepsilon E(\chi^2) \doteqdot \frac{n}{n-R}(s-1)(t-1) \tag{7.8}$$

Next, let us consider the variance of χ^2. After some complicated calculation we have

$$\begin{aligned}
D^2(\chi^2)=n^2\Bigg[&\sum_i\sum_j \frac{1}{n_{i\cdot}^2 n_{\cdot j}^2}\Bigg\{\sum_k\Bigg(\frac{n_{ki\cdot}^{(4)} n_{k\cdot j}^{(4)}}{n_k^{(4)}}+6\frac{n_{ki\cdot}^{(3)} n_{k\cdot j}^{(3)}}{n_k^{(3)}}+7\frac{n_{ki\cdot}^{(2)} n_{k\cdot j}^{(2)}}{n_k^{(2)}} \\
&+\frac{n_{ki\cdot} n_{k\cdot j}}{n_k}\Bigg)+4\sum_{k\neq k'}\sum\Bigg(\frac{n_{ki\cdot}^{(3)} n_{k\cdot j}^{(3)}}{n_k^{(3)}}+3\frac{n_{ki\cdot}^{(2)} n_{k\cdot j}^{(2)}}{n_k^{(2)}}+\frac{n_{ki\cdot} n_{k\cdot j}}{n_k}\Bigg)\frac{n_{k'i\cdot} n_{k'\cdot j}}{n_{k'}} \\
&+3\sum_{k\neq k'}\sum\Bigg(\frac{n_{ki\cdot}^{(2)} n_{k\cdot j}^{(2)}}{n_k^{(2)}}+\frac{n_{ki\cdot} n_{k\cdot j}}{n_k}\Bigg)\Bigg(\frac{n_{k'i}^{(2)} n_{k'\cdot j}^{(2)}}{n_{k'}^{(2)}}+\frac{n_{ki'\cdot} n_{k'\cdot j}}{n_{k'}}\Bigg) \\
&+6\sum_{k\neq k'\neq k''}\sum\sum\Bigg(\frac{n_{ki\cdot}^{(2)} n_{k\cdot j}^{(2)}}{n_k^{(2)}}+\frac{n_{ki\cdot} n_{k\cdot j}}{n_k}\Bigg)\frac{n_{k'i\cdot} n_{k'\cdot j} n_{k''i\cdot} n_{k''\cdot j}}{n_{k'} n_{k''}} \\
&+\sum_{k\neq k'\neq k''\neq k'''}\sum\sum\sum \frac{n_{ki\cdot} n_{k\cdot j} n_{k'i\cdot} n_{k'\cdot j} n_{k''i\cdot} n_{k''\cdot j} n_{k'''i\cdot} n_{k'''\cdot j}}{n_k n_{k'} n_{k''} n_{k'''}}\Bigg\} \\
&+\sum_i\sum_{j\neq j'}\sum \frac{1}{n_{i\cdot}^2 n_{\cdot j} n_{\cdot j'}}\Bigg\{\sum_k\Bigg(\frac{n_{ki\cdot}^{(4)} n_{k\cdot j}^{(2)} n_{k\cdot j'}^{(2)}}{n_k^{(4)}}+\frac{n_{ki\cdot}^{(3)}(n_{k\cdot j}^{(2)} n_{k\cdot j'}+n_{k\cdot j'}^{(2)} n_{k\cdot j})}{n_k^{(3)}} \\
&+\frac{n_{ki\cdot}^{(2)} n_{k\cdot j} n_{k\cdot j'}}{n_k^{(2)}}\Bigg)+\sum_{k\neq k'}\sum\Bigg(\frac{n_{ki\cdot}^{(2)} n_{k\cdot j}^{(2)}}{n_k^{(2)}}+\frac{n_{ki\cdot} n_{k\cdot j}}{n_k}\Bigg)\Bigg(\frac{n_{k'i\cdot}^{(2)} n_{k'\cdot j'}^{(2)}}{n_{k'}^{(2)}}+\frac{n_{k'i\cdot} n_{k'\cdot j'}}{n_{k'}}\Bigg) \\
&+2\sum_{k\neq k'}\sum\Bigg(\frac{n_{ki\cdot}^{(3)} n_{k\cdot j} n_{k\cdot j'}^{(2)}}{n_k^{(3)}}+\frac{n_{ki\cdot}^{(2)} n_{k\cdot j} n_{k\cdot j'}}{n_k^{(2)}}\Bigg)\frac{n_{k'i\cdot} n_{k'\cdot j}}{n_{k'}} \\
&+\sum_{k\neq k'\neq k''}\sum\sum \frac{n_{ki\cdot} n_{k\cdot j} n_{k'i\cdot} n_{k'\cdot j}}{n_k n_{k'}}\Bigg(\frac{n_{k''i\cdot}^{(2)} n_{k''\cdot j'}^{(2)}}{n_{k''}^{(2)}}+\frac{n_{k''i\cdot} n_{k''\cdot j'}}{n_{k''}}\Bigg)+2\sum_{k\neq k'}\sum\Bigg(\frac{n_{ki\cdot}^{(3)} n_{k\cdot j}^{(2)} n_{k\cdot j'}}{n_k^{(3)}} \\
&+\frac{n_{ki\cdot}^{(2)} n_{k\cdot j} n_{k\cdot j'}}{n_k^{(2)}}\Bigg)\frac{n_{k'i\cdot} n_{k'\cdot j'}}{n_{k'}}+\sum_{k\neq k'\neq k''}\sum\sum\Bigg(\frac{n_{ki\cdot}^{(2)} n_{k\cdot j}^{(2)}}{n_k^{(2)}}+\frac{n_{ki\cdot} n_{k\cdot j}}{n_k}\Bigg)\frac{n_{k'i\cdot} n_{k'\cdot j'} n_{k''i\cdot} n_{k''\cdot j'}}{n_{k'} n_{k''}} \\
&+2\sum_{k\neq k'}\sum \frac{n_{ki\cdot}^{(2)} n_{k\cdot j} n_{k\cdot j'} n_{k'i\cdot}^{(2)} n_{k'\cdot j} n_{k'\cdot j'}}{n_k^{(2)} n_{k'}^{(2)}}+4\sum_{k\neq k'\neq k''}\sum\sum \frac{n_{ki\cdot}^{(2)} n_{k\cdot j} n_{k\cdot j'} n_{k'i\cdot} n_{k'\cdot j} n_{k''i\cdot} n_{k''\cdot j'}}{n_k^{(2)} n_{k'} n_{k''}} \\
&+\sum_{k\neq k'\neq k''\neq k'''}\sum\sum\sum \frac{n_{ki\cdot} n_{k\cdot j} n_{k'i\cdot} n_{k'\cdot j} n_{k''i\cdot} n_{k''\cdot j'} n_{k'''i\cdot} n_{k'''\cdot j}}{n_k n_{k'} n_{k''} n_{k'''}}\Bigg\} \\
&+\sum_{i\neq i'}\sum_j\sum \frac{1}{n_{i\cdot} n_{i'\cdot} n_{\cdot j}^2}\Bigg\{\sum_k\Bigg(\frac{n_{k\cdot j}^{(4)} n_{ki\cdot}^{(2)} n_{ki'\cdot}^{(2)}}{n_k^{(4)}}+\frac{n_{k\cdot j}^{(3)}(n_{ki\cdot}^{(2)} n_{ki'\cdot}+n_{ki\cdot} n_{ki'\cdot}^{(2)})}{n_k^{(3)}}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{n_{k.j}^{(2)} n_{ki.} n_{ki'.}}{n_k^{(2)}}\Big) + \sum\sum_{k\neq k'} \Big(\frac{n_{ki.}^{(2)} n_{k.j}^{(2)}}{n_k^{(2)}} + \frac{n_{ki.} n_{k.j}}{n_k}\Big)\Big(\frac{n_{k'i'.}^{(2)} n_{k'.j}^{(2)}}{n_{k'}^{(2)}} + \frac{n_{k'i.} n_{k'.j}}{n_{k'}}\Big) \\
&+\sum\sum\sum_{k\neq k'\neq k''} \frac{n_{ki.}n_{k.j}n_{k'i.}n_{k'.j}}{n_k n_{k'}}\Big(\frac{n_{k''i'.}^{(2)} n_{k''.j}^{(2)}}{n_{k''}^{(2)}} + \frac{n_{k''i'.}n_{k''.j}}{n_{k''}}\Big) \\
&+2\sum\sum_{k\neq k'}\Big(\frac{n_{k.j}^{(3)} n_{ki'.}^{(2)} n_{ki.}}{n_k^{(3)}} + \frac{n_{k.j}^{(2)} n_{ki'.} n_{ki.}}{n_k^{(2)}}\Big)\frac{n_{k'i.}n_{k'.j}}{n_{k'}} + 2\sum\sum_{k\neq k'}\Big(\frac{n_{k.j}^{(3)} n_{ki.}^{(2)} n_{ki'.}}{n_k^{(3)}} \\
&+\frac{n_{k.j}^{(2)} n_{ki.} n_{ki'.}}{n_k^{(2)}}\Big)\frac{n_{k'i'.}n_{k'.j}}{n_{k'}} + \sum\sum\sum_{k\neq k'\neq k''}\Big(\frac{n_{ki.}^{(2)} n_{k.j}^{(2)}}{n_k^{(2)}}+\frac{n_{ki.}n_{k.j}}{n_k}\Big)\frac{n_{k'i'.}n_{k'.j}n_{k''i'.}n_{k''.j}}{n_{k'}n_{k''}} \\
&+\sum\sum_{k\neq k'}\frac{n_{k.j}^{(2)} n_{ki.}n_{ki'.} n_{k'.j}^{(2)} n_{k'i.} n_{k'i'.}}{n_k^{(2)} n_{k'}^{(2)}} + 2\sum\sum\sum_{k\neq k'\neq k''}\frac{n_{k.j}^{(2)} n_{ki.} n_{ki'.} n_{k'i.} n_{k'.j} n_{k''i'.} n_{k''.j}}{n_k^{(2)} n_{k'} n_{k''}} \\
&+\sum\sum\sum\sum_{k\neq k'\neq k''\neq k'''}\frac{n_{ki.}n_{k.j}n_{k'i.}n_{k'.j}n_{k''i'.}n_{k''.j}n_{k'''i'.}n_{k'''.j}}{n_k n_{k'} n_{k''} n_{k'''}}\Big\} \\
&+\sum\sum_{i\neq i'}\sum\sum_{j\neq j'}\frac{1}{n_{i.}n_{i'.}n_{.j}n_{.j'}}\Big\{\sum_k\Big(\frac{n_{ki.}^{(2)}n_{k.j}^{(2)}n_{ki'.}^{(2)}n_{k.j'}^{(2)}}{n_k^{(4)}} + \frac{n_{ki.}^{(2)} n_{k.j}^{(2)} n_{ki'.}n_{k.j'}}{n_k^{(3)}} \\
&+\frac{n_{ki.}n_{k.j}n_{ki'.}^{(2)}n_{k.j'}^{(2)}}{n_k^{(3)}} + \frac{n_{ki.}n_{k.j}n_{ki'.}n_{k.j'}}{n_k^{(2)}}\Big) \\
&+\sum\sum_{k\neq k'}\Big(\frac{n_{ki.}^{(2)}n_{k.j}^{(2)}}{n_k^{(2)}}+\frac{n_{ki.}n_{k.j}}{n_k}\Big)\Big(\frac{n_{k'i'.}^{(2)}n_{k'.j'}^{(2)}}{n_{k'}^{(2)}}+\frac{n_{k'i'.}n_{k'.j'}}{n_{k'}}\Big) \\
&+2\sum\sum_{k\neq k'}\Big(\frac{n_{ki'.}^{(2)}n_{k.j'}^{(2)}n_{ki.}n_{k.j}}{n_k^{(3)}}+\frac{n_{ki'.}n_{k.j'}n_{ki.}n_{k.j}}{n_k^{(2)}}\Big)\frac{n_{k'i.}n_{k'.j}}{n_{k'}} \\
&+\sum\sum\sum_{k\neq k'\neq k''}\frac{n_{ki.}n_{k.j}n_{k'i.}n_{k'.j}}{n_k n_{k'}}\Big(\frac{n_{k''i'.}^{(2)}n_{k''.j'}^{(2)}}{n_{k''}^{(2)}}+\frac{n_{k''i'.}n_{k''.j'}}{n_{k''}}\Big) \\
&+2\sum\sum_{k\neq k'}\Big(\frac{n_{ki.}^{(2)}n_{k.j}^{(2)}n_{ki'.}n_{k.j'}}{n_k^{(3)}}+\frac{n_{ki.}n_{k.j}n_{ki'.}n_{k.j'}}{n_k^{(2)}}\Big)\frac{n_{k'i'.}n_{k'.j'}}{n_{k'}} \\
&+\sum\sum\sum_{k\neq k'\neq k''}\Big(\frac{n_{ki.}^{(2)}n_{k.j}^{(2)}}{n_k^{(2)}}+\frac{n_{ki.}n_{k.j}}{n_k}\Big)\frac{n_{k'i'.}n_{k'.j'}n_{k''i'.}n_{k''.j'}}{n_{k'}n_{k''}} \\
&+2\sum\sum_{k\neq k'}\frac{n_{ki.}n_{k.j}n_{ki'.}n_{k.j'}n_{k'i.}n_{k'.j}n_{k'i'.}n_{k'.j'}}{n_k^{(2)}n_{k'}^{(2)}} \\
&+4\sum\sum\sum_{k\neq k'\neq k''}\frac{n_{ki.}n_{k.j}n_{ki'.}n_{k.j'}n_{k'i.}n_{k'.j}n_{k''i'.}n_{k''.j'}}{n_k^{(2)}n_{k'}n_{k''}} \\
&+\sum\sum\sum\sum_{k\neq k'\neq k''\neq k'''}\frac{n_{ki.}n_{k.j}n_{k'i.}n_{k'.j}n_{k''i'.}n_{k''.j'}n_{k'''i'.}n_{k'''.j'}}{n_kn_{k'}n_{k''}n_{k'''}}\Big\}\Big] - (n+E(\chi^2))^2 \qquad (7.9)
\end{aligned}
$$

It is so much complicated that we cannot use it practically. But when $N_k=N/R$, $N_{i.}=N/s$ and $N_{.j}=N/t$, using the notation ε^* in section 5, we have

$$
\varepsilon^* D^2(\chi^2)=2\frac{n}{n-R}(s-1)(t-1)+O\Big(\frac{1}{n}\Big) \qquad (7.10)
$$

Hence we can apply the usual χ^2 distribution of $(s-1)(t-1)$ degrees of freedom only for this special case.

8 Correlation coefficient in the stratified random sampling

In this section we treat the correlation coefficient in the stratified rondom sampling. Let N be the size of the universe, R the number of strata, N_i the size of the i-th stratum $(i=1, 2, \ldots, R)$, n the total sample size, and n_i the sample size in the i-th stratum $(i=1, 2, \ldots, R)$. When there are two characteristics X and Y with which we are concerned, let f_{ijk} be the number of combinations of the characteristics (x_j, y_k) in the i-th stratum, and put

$$\left.\begin{aligned} &\bar{x}_i=\frac{1}{n_i}\sum_{j=1}^{s} f_{ij\cdot}x_j, && \bar{y}_i=\frac{1}{n_i}\sum_{k=1}^{t} f_{i\cdot k}y_k \\ &f_{ij\cdot}=\sum_{k=1}^{t} f_{ijk}, && f_{i\cdot k}=\sum_{j=1}^{s} f_{ijk} \\ &n_i=\sum_{j=1}^{s} f_{ij\cdot}=\sum_{k=1}^{t} f_{i\cdot k} && \\ &\bar{x}=\frac{1}{N}\sum_{i=1}^{R} N_i\bar{x}_i, && \bar{y}=\frac{1}{N}\sum_{i=1}^{R} N_i\bar{y}_i \end{aligned}\right\} \tag{8.1}$$

In order to calculate the sample correlation coefficient r we give the weight $w_i=N_i/n_i$ to the sample of the i-th stratum and put

$$r=\frac{\frac{1}{N}\sum_i\sum_j\sum_k w_i f_{ijk}(x_j-\bar{x})(y_k-\bar{y})}{\sqrt{\frac{1}{N}\sum_i\sum_j w_i f_{ij\cdot}(x_j-\bar{x})^2}\sqrt{\frac{1}{N}\sum_i\sum_k w_i f_{i\cdot k}(y_k-\bar{y})^2}} \tag{8.2}$$

LEMMA 1 *Let $f(X, Y, Z)$ be the function of three random variables X, Y, and Z in a certain region Ω and have the continuous derivatives up to the second order in regard to X, Y and Z. Then under the assumption that the higher order terms in Taylor's expansions such as variances, covariances and central moments are infinitesimal, we have*

$$Ef(X, Y, Z)\doteqdot f(E(X), E(Y), E(Z)) \tag{8.3}$$

$$\begin{aligned} D^2f(X, Y, Z)\doteqdot &\left(\frac{\partial f}{\partial X}\right)_0^2 D^2(X)+\left(\frac{\partial f}{\partial Y}\right)_0^2 D^2(Y)+\left(\frac{\partial f}{\partial Z}\right)_0^2 D^2(Z) \\ &+2\left(\frac{\partial f}{\partial X}\right)_0\left(\frac{\partial f}{\partial Y}\right)_0 \operatorname{cov}(X, Y)+2\left(\frac{\partial f}{\partial Y}\right)_0\left(\frac{\partial f}{\partial Z}\right)_0 \operatorname{cov}(Y, Z) \\ &+2\left(\frac{\partial f}{\partial Z}\right)_0\left(\frac{\partial f}{\partial X}\right)_0 \operatorname{cov}(Z, X) \end{aligned} \tag{8.4}$$

PROOF: Applying Taylor's theorem to $f(X, Y, Z)$ in a neighbourhood of the point $(E(X), E(Y), E(Z))$, we have

$$f(X, Y, Z)=f(E(X), E(Y), E(Z))+(X-E(X))\left(\frac{\partial f}{\partial X}\right)_0$$

$$+(Y-E(Y))\left(\frac{\partial f}{\partial Y}\right)_0+(Z-E(Z))\left(\frac{\partial f}{\partial Z}\right)_0+R_1 \tag{8.5}$$

where the suffix 0 denotes the value at the point $(E(X), E(Y), E(Z))$ and

$$R_1=\frac{1}{2}\Big[\left(\frac{\partial^2 f}{\partial X^2}\right)_1(X-E(X))^2+\left(\frac{\partial^2 f}{\partial Y^2}\right)_1(Y-E(Y))^2+\left(\frac{\partial^2 f}{\partial Z^2}\right)_1(Z-E(Z))^2$$

$$+2\left(\frac{\partial^2 f}{\partial X\partial Y}\right)_1(X-E(X))(Y-E(Y))+2\left(\frac{\partial^2 f}{\partial Y\partial Z}\right)_1(Y-E(Y))(Z-E(Z))$$

$$+2\left(\frac{\partial^2 f}{\partial Z\partial X}\right)_1(Z-E(Z))(X-E(X))\Big]$$

Taking the expectation of (8.5), we have

$$Ef(X, Y, Z)=f(E(X), E(Y), E(Z))+E(R_1)$$

Here, we can neglect the second term of the right hand side, because the higher order terms of the variances σ_X^2, σ_Y^2, σ_Z^2, and convariances σ_{XY}, σ_{YZ}, σ_{ZX} are infinitesimal compared with the first term.

On the other hand we get

$$D^2 f(X, Y, Z) \doteqdot E\{f(X, Y, Z)-f(E(X), E(Y), E(Z))\}^2$$

$$=E\Big\{\left(\frac{\partial f}{\partial X}\right)_0^2(X-E(X))^2+\left(\frac{\partial f}{\partial Y}\right)_0^2(Y-E(Y))^2+\left(\frac{\partial f}{\partial Z}\right)_0^2(Z-E(Z))^2$$

$$+2(X-E(X))(Y-E(Y))\left(\frac{\partial f}{\partial X}\right)_0\left(\frac{\partial f}{\partial Y}\right)_0+2(Y-E(Y))(Z-E(Z))\left(\frac{\partial f}{\partial Y}\right)_0\left(\frac{\partial f}{\partial Z}\right)_0$$

$$+2(Z-E(Z))(X-E(X))\left(\frac{\partial f}{\partial Z}\right)_0\left(\frac{\partial f}{\partial X}\right)+R_2\Big\}$$

where R_2 denotes the term of higher order. From this relation we obtain

$$D^2 f(X, Y, Z) \doteqdot \left(\frac{\partial f}{\partial X}\right)_0^2 D^2(X)+\left(\frac{\partial f}{\partial Y}\right)_0^2 D^2(Y)+\left(\frac{\partial f}{\partial Z}\right)_0^2 D^2(Z)$$

$$+2\left(\frac{\partial f}{\partial X}\right)_0\left(\frac{\partial f}{\partial Y}\right)_0 \operatorname{cov}(X, Y)+2\left(\frac{\partial f}{\partial Y}\right)_0\left(\frac{\partial f}{\partial Z}\right)_0 \operatorname{cov}(Y, Z)$$

$$+2\left(\frac{\partial f}{\partial Z}\right)_0\left(\frac{\partial f}{\partial X}\right) \operatorname{cov}(Z, X)$$

LEMMA 2 : *Let* $f(X, Y, Z)=\frac{Z}{\sqrt{XY}}$ *under the same conditions as in Lemma 1. Then we have*

$$E\left(\frac{Z}{\sqrt{XY}}\right)\doteqdot\frac{E(Z)}{\sqrt{E(X)E(Y)}} \tag{8.6}$$

$$D^2\left(\frac{Z}{\sqrt{XZ}}\right)\doteqdot\frac{1}{E(X)E(Y)}\left\{\frac{E^2(Z)}{4E^2(X)}D^2(X)+\frac{E^2(Z)}{4E^2(Y)}D^2(Y)+D^2(Z)\right.$$
$$\left.-\frac{E(Z)}{E(X)}\operatorname{cov}(X, Z)-\frac{E(Z)}{E(Y)}\operatorname{cov}(Y, Z)+\frac{E^2(Z)}{2E(X)E(Y)}\operatorname{cov}(X, Y)\right\} \tag{8.7}$$

PROOF : Evident from lemma 1.

From lemma 2, neglecting the finite population correction, we have

$$E(r)\doteqdot\frac{\frac{1}{N}\sum_i N_i\mu_{11}(i)+\frac{1}{N}\sum_i N_i(\overline{X}_i\overline{Y}_i-\overline{X}\overline{Y})-\frac{1}{N^2}\sum_i\frac{N_i^2}{n_i}\mu_{11}(i)}{\left[\begin{array}{c}\sqrt{\sum_i\frac{N_i}{N}\mu_{20}(i)+\sum_i\frac{N_i}{N}(\overline{X}_i-\overline{X})^2-\frac{1}{N^2}\sum_i\frac{N_i^2}{n_i}\mu_{20}(i)}\\ \times\sqrt{\sum_i\frac{N_i}{N}\mu_{02}(i)+\sum_i\frac{N_i}{N}(\overline{Y}_i-\overline{Y})^2-\frac{1}{N^2}\sum_i\frac{N_i^2}{n_i}\mu_{02}(i)}\end{array}\right]}$$
$$=\rho+O\left(\frac{1}{n_i}\right) \tag{8.8}$$

where μ_{hk} denotes the h, k-th central moment in the i-th stratum and ρ denotes the population correlation coefficient.

The variance in case $R=2$ is

$$D^2(r)=\frac{1}{\mu_{20}\mu_{02}}\left[\frac{\mu_{11}^2}{4\mu_{20}^2}\left\{\frac{1}{n_1}\left(\frac{N_1^2}{N^2}(\mu_{40}(1)-\mu_{20}^2(1))+\frac{4N_1^2N_2^2}{N^4}\mu_{20}(1)(\overline{X}_1-\overline{X}_2)^2\right.\right.\right.$$
$$\left.+4\frac{N_1^2N_2}{N^3}\mu_{30}(1)(\overline{X}_1-\overline{X}_2)\right)+\frac{1}{n_2}\left(\frac{N_2^2}{N^2}(\mu_{40}(2)-\mu_{20}^2(2))+\frac{4N_1^2N_2^2}{N^4}\mu_{20}(2)(\overline{X}_1-\overline{X}_2)^2\right.$$
$$\left.\left.+\frac{4N_1N_2^2}{N^3}\mu_{30}(2)(\overline{X}_2-\overline{X}_1)\right)\right\}+\frac{\mu_{11}^2}{4\mu_{02}^2}\left\{\frac{1}{n_1}\left(\frac{N_1^2}{N^2}(\mu_{04}(1)-\mu_{02}^2(1))\right.\right.$$
$$\left.+\frac{4N_1^2N_2^2}{N^4}\mu_{02}(1)(\overline{Y}_1-\overline{Y}_2)^2+\frac{4N_1^2N_2}{N^3}\mu_{03}(1)(\overline{Y}_1-\overline{Y}_2)\right)+\frac{1}{n_2}\left(\frac{N_2^2}{N^2}(\mu_{04}(2)-\mu_{02}^2(2))\right.$$
$$\left.\left.+\frac{4N_1^2N_2^2}{N^4}\mu_{02}(2)(\overline{Y}_1-\overline{Y}_2)^2+\frac{4N_1N_2^2}{N^3}\mu_{03}(2)(\overline{Y}_2-\overline{Y}_1)\right)\right\}+\frac{1}{n_1}\left\{\frac{N_1^2}{N^2}(\mu_{22}(1)-\mu_{11}^2(1))\right.$$
$$+\frac{N_1^2N_2^2}{N^4}(\mu_{20}(1)(\overline{Y}_1-\overline{Y}_2)^2+\mu_{02}(1)(\overline{X}_1-\overline{X}_2)^2+2\mu_{11}(1)(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2))$$

$$+\frac{2N_1^2N_2}{N^3}(\mu_{12}(1)(\overline{X}_1-\overline{X}_2)+\mu_{21}(1)(\overline{Y}_1-\overline{Y}_2))\Big\}+\frac{1}{n_2}\Big\{\frac{N_2^2}{N^2}(\mu_{22}(2)-\mu_{11}^2(2))$$

$$+\frac{N_1^2N_2^2}{N^4}(\mu_{20}(2)(\overline{Y}_1-\overline{Y}_2)^2+\mu_{02}(2)(\overline{X}_1-\overline{X}_2)^2+2\mu_{11}(2)(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2))$$

$$+\frac{2N_1N_2^2}{N^3}(\mu_{21}(2)(\overline{X}_2-\overline{X}_1)+\mu_{12}(2)(\overline{Y}_2-\overline{Y}_1))\Big\}-\frac{\mu_{11}}{\mu_{20}}\Big\{\frac{1}{n_1}\Big(\frac{N_1^2}{N^2}(\mu_{31}(1)$$

$$-\mu_{11}(1)\mu_{20}(1)\Big)+\frac{N_1^2N_2}{N^3}((\overline{Y}_1-\overline{Y}_2)\mu_{30}(1)+3(\overline{X}_1-\overline{X}_2)\mu_{21}(1)$$

$$-\mu_{20}(1)(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2)-\mu_{11}(\overline{X}_1-\overline{X}_2)^2)+\frac{3N_1^2N_2^2}{N^4}((\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2)\mu_{20}(1)$$

$$+(\overline{X}_1-\overline{X}_2)^2\mu_{11}(1))+\frac{N_1^3N_2}{N^4}((\overline{X}_1-\overline{X}_2)\mu_{11}(1)+(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2)\mu_{20}(1))\Big)$$

$$+\frac{1}{n_2}\Big(\frac{N_2^2}{N^2}(\mu_{31}(2)-\mu_{11}(1)\mu_{20}(2))+\frac{N_1N_2^2}{N^3}((\overline{Y}_2-\overline{Y}_1)\mu_{30}(2)+3(\overline{X}_2-\overline{X}_1)\mu_{21}(2)$$

$$-(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2)\mu_{20}(2)-(\overline{X}_1-\overline{X}_2)^2\mu_{11}(2))+\frac{3N_1^2N_2}{N^4}((\overline{X}_2-\overline{X}_1)(\overline{Y}_2-\overline{Y}_1)\mu_{20}(2)$$

$$+(\overline{X}_1-\overline{X}_2)^2\mu_{11}(2))+\frac{N_1N_2^3}{N^4}((\overline{X}_1-\overline{X}_1)\mu_{11}(2)+(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2)\mu_{20}(2))\Big)\Big\}$$

$$-\frac{\mu_{11}}{\mu_{02}}\Big\{\frac{1}{n_1}\Big(\frac{N_1^2}{N^2}(\mu_{13}(1)-\mu_{11}(1)\mu_{20}(1))+\frac{N_1^2N_2}{N^3}(\overline{X}_1-\overline{X}_2)\mu_{03}(1)+3(\overline{Y}_1-\overline{Y}_2)\mu_{12}(1)$$

$$-(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2)\mu_{02}(1)-\mu_{11}(1)(\overline{Y}_2-\overline{Y}_1)^2)+\frac{3N_1^2N_2^2}{N^4}((\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2)\mu_{02}(1)$$

$$+(\overline{Y}_1-\overline{Y}_2)^2\mu_{11}(1))+\frac{N_1^3N_2}{N^4}((\overline{Y}_1-\overline{Y}_2)\mu_{11}(1)+(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2)\mu_{02}(1))\Big)$$

$$+\frac{1}{n_2}\Big(\frac{N_2^2}{N^2}(\mu_{13}(2)-\mu_{11}(2)\mu_{02}(2))+\frac{N_1N_2^2}{N^3}((\overline{X}_2-\overline{X}_1)\mu_{03}(2)+3(\overline{Y}_2-\overline{Y}_1)\mu_{12}(2)$$

$$-(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2)\mu_{02}(2)-(\overline{Y}_1-\overline{Y}_2)^2\mu_{11}(2))+\frac{3N_1^2N_2^2}{N^4}((\overline{X}_2-\overline{X}_1)(\overline{Y}_2-\overline{Y}_1)\mu_{02}(2)$$

$$+(\overline{Y}_1-\overline{Y}_2)^2\mu_{11}(2))+\frac{N_1N_2^3}{N^4}((\overline{Y}_2-\overline{Y}_1)\mu_{11}(2)+(\overline{X}_2-\overline{X}_1)(\overline{Y}_2-\overline{Y}_1)\mu_{02}(2))\Big)\Big\}\Big]$$

$$+\frac{\mu_{11}^2}{2\mu_{20}\mu_{02}}\Big[\frac{1}{n_1}\Big\{\frac{N_1^2}{N^2}(\mu_{22}(1)-\mu_{20}(1)\mu_{02}(1))-((\overline{X}_1-\overline{X}_2)^2\mu_{02}(1)$$

$$-2(\overline{X}_1-\overline{X}_2)\mu_{12}(1))\frac{N_1^2N_2}{N^3}-\frac{N_1^2N_2}{N^3}((\overline{Y}_1-\overline{Y}_2)^2\mu_{20}(1)-2(\overline{Y}_1-\overline{Y}_2)\mu_{21}(1))$$

$$+\frac{N_1^2N_2^2}{N^4}((\overline{Y}_1-\overline{Y}_2)^2\mu_{20}(1)+(\overline{X}_1-\overline{X}_2)^2\mu_{02}(1)+4\mu_{11}(1)(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2))$$

$$+\frac{N_1^3N_2}{N^4}((\overline{X}_1-\overline{X}_2)^2\mu_{02}(1)+(\overline{Y}_1-\overline{Y}_2)^2\mu_{20}(1))\Big\}+\frac{1}{n_2}\Big\{\frac{N_2^2}{N^2}(\mu_{22}(2)-\mu_{20}(2)\mu_{02}(2))$$

$$-\frac{N_1N_2^2}{N^3}((\overline{X}_1-\overline{X}_2)^2\mu_{02}(2)-2(\overline{X}_1-\overline{X}_2)\mu_{12}(2)+(\overline{Y}_1-\overline{Y}_2)^2\mu_{20}(2)$$

$$-2(\overline{Y}_1-\overline{Y}_2)\mu_{21}(2))+\frac{N_1^2N_2^2}{N^4}((\overline{Y}_1-\overline{Y}_2)^2\mu_{20}(2)+(\overline{X}_1-\overline{X}_2)^2\mu_{02}(2)$$

$$+4\mu_{11}(2)(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2)+\frac{N_1N_2^3}{N^4}((\overline{X}_1-\overline{X}_2)^2\mu_{02}(2)+(\overline{Y}_1-\overline{Y}_2)^2\mu_{20}(2))\Big\}\Big]$$

$$+O(n_i^{-3/2}) \qquad (8.9)$$

Especially in the proportionate sampling

$$D^2(r)_{\text{prop.}}=\frac{\rho^2}{4n}\Big[\Big\{\frac{\mu_{40}}{\mu_{20}^2}-\frac{1}{\mu_{20}^2}\Big(\frac{N_1}{N}\mu_{20}^2(1)+\frac{N_2}{N}\mu_{20}^2(2)+\frac{2N_1N_2}{N^3}(\overline{X}_1-\overline{X}_2)^2(N_2\mu_{20}(1)$$

$$+N_1\mu_{20}(2))+\frac{N_1N_2}{N^5}(N_1^3+N_2^3)(\overline{X}_1-\overline{X}_2)^4\Big)\Big\}+\Big\{\frac{\mu_{04}}{\mu_{02}^2}-\frac{1}{\mu_{02}^2}\Big(\frac{N_1}{N}\mu_{02}^2(1)+\frac{N_2}{N}\mu_{02}^2(2)$$

$$+\frac{2N_1N_2}{N^2}(\overline{Y}_1-\overline{Y}_2)^2(N_2\mu_{20}(1)+N_1\mu_{02}(2))+\frac{N_1N_2}{N^5}(N_1^2+N_2^3)(\overline{Y}_1-\overline{Y}_2)^4\Big)\Big\}$$

$$+\Big\{\frac{4\mu_{22}}{\mu_{11}^2}-\frac{4}{\mu_{11}^2}\Big(\frac{N_1}{N}\mu_{11}^2(1)+\frac{N_2}{N}\mu_{11}^2(2)+\frac{2N_1N_2}{N^3}(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2)(N_2\mu_{11}(1)$$

$$+N_2\mu_{11}(2))+\frac{N_1N_2}{N^5}(N_1^3+N_2^3)(\overline{X}_1-\overline{X}_2)^2(\overline{Y}_1-\overline{Y}_2)^2\Big)\Big\}-\Big\{\frac{4\mu_{31}}{\mu_{11}\mu_{20}}$$

$$-\frac{4}{\mu_{11}\mu_{20}}\Big(\frac{N_1}{N}\mu_{11}(1)\mu_{20}(1)+\frac{N_2}{N}\mu_{11}(2)\mu_{20}(2)+\frac{N_1N_2}{N^2}(\overline{X}_1-\overline{X}_2)$$

$$\times(\overline{Y}_1-\overline{Y}_2)\Big(\mu_{20}(1)\frac{N+N_2}{N}+\mu_{20}(2)\frac{N+N_1}{N}\Big)+\frac{N_1N_2}{N^2}(\overline{X}_1-\overline{X}_2)^2\Big(\mu_{11}(1)\frac{N+N_2}{N}$$

$$+\mu_{11}(2)\frac{N+N_1}{N}\Big)+\frac{N_1N_2}{N^5}(N_1^3+N_2^3-NN_1N_2)(\overline{X}_1-\overline{X}_2)^3(\overline{Y}_1-\overline{Y}_2)\Big)\Big\}$$

$$-\Big\{\frac{4\mu_{13}}{\mu_{11}\mu_{02}}-\frac{4}{\mu_{11}\mu_{02}}\Big(\frac{N_1}{N}\mu_{11}(1)\mu_{02}(1)+\frac{N_2}{N}\mu_{11}(2)\mu_{02}(2)$$

$$+\frac{N_1N_2}{N^2}(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2)\Big(\mu_{02}(1)\frac{N+N_2}{N}+\mu_{02}(2)\frac{N+N_1}{N}\Big)+\frac{N_1N_2}{N^2}(\overline{Y}_1-\overline{Y}_2)^2$$

$$\times\Big(\mu_{11}(1)\frac{N+N_2}{N}+\mu_{11}(2)\frac{N+N_1}{N}\Big)$$

$$+\frac{N_1N_2}{N^5}(N_1^2+N_2^3-NN_1N_2)(\overline{X}_1-\overline{X}_2)(\overline{Y}_1-\overline{Y}_2)^3\Big)\Big\}+\Big\{\frac{2\mu_{22}}{\mu_{20}\mu_{02}}$$

$$-\frac{2}{\mu_{20}\mu_{02}}\Big(\frac{N_1}{N}\mu_{20}(1)\mu_{02}(1)+\frac{N_2}{N}\mu_{20}(2)\mu_{02}(2)+\frac{N_1N_2}{N^3}(\overline{X}_1-\overline{X}_2)^2(N_2\mu_{02}(1)$$

$$+N_1\mu_{02}(2))+\frac{N_1N_2}{N^3}(\overline{Y}_1-\overline{Y}_2)^2(N_2\mu_{20}(1)+N_1\mu_{20}(2))$$

$$+\frac{N_1N_2}{N^5}(N_1^3+N_2^3)(\overline{X}_1-\overline{X}_2)^2(\overline{Y}_1-\overline{Y}_2)^2\Big)\Big\}\Big]+O(n^{-3/2}) \tag{8.10}$$

If we divide the universe elements into R strata at random, we obtain

$$\mathcal{E}D^2(r)_{\text{prop.}}=\frac{\rho^2}{4n}\Big(\frac{\mu_{40}}{\mu_{20}^2}+\frac{\mu_{04}}{\mu_{02}^2}+\frac{4\mu_{22}}{\mu_{11}^2}-\frac{4\mu_{31}}{\mu_{11}\mu_{20}}-\frac{4\mu_{13}}{\mu_{11}\mu_{02}}+\frac{2\mu_{22}}{\mu_{20}\mu_{02}}\Big)+O(n^{-3/2}) \tag{8.11}$$

In case of arbitrary $R(\geqq 3)$ the formula of $D^2(r)$ is so much complicated that we can not use it for the practical purpose. But, under certain conditions, we can show that $\mathcal{E}D^2(r)_{\text{prop.}}$ in this case is also equal to (8.11). First, we will prove the following lemma.

LEMMA 3 *Let $f(m_i, m_j)$ and $g(m_k, m_l)$ be the functions of the central moments m_i, m_j, and m_k, m_l, respectively, which are constructed from the sample of size n. Suppose that the following two conditions are satisfied:*

1) *In some neighbourhood of the point $m_i=\mu_i$, $m_j=\mu_j$, $m_k=\mu_k$, $m_l=\mu_l$, the function f and g are continuous and have continuous first and second derivatives with respect to these arguments.*
2) *For all possible values of x_i, we have $|f|<Cn^p$, and $|g|<Cn^p$, where C and p are non-negative constants.*

Denote by the suffix 0 the values at the point $m_i=\mu_i$, $m_j=\mu_j$ or $m_k=\mu_k$, $m_l=\mu_l$. Then, the mean, the variance and the convariance of the random variables f and g are given by

$$E(f)=f_0+O\Big(\frac{1}{n}\Big) \tag{8.12}$$

$$D^2(f)=D^2(m_i)\Big(\frac{\partial f}{\partial m_i}\Big)_0^2+D^2(m_j)\Big(\frac{\partial f}{\partial m_j}\Big)_0^2+2\operatorname{cov}(m_i, m_j)\Big(\frac{\partial f}{\partial m_i}\Big)_0\Big(\frac{\partial f}{\partial m_j}\Big)_0$$
$$+O(n^{-3/2}) \tag{8.13}$$

$$\operatorname{cov}(f, g)=\operatorname{cov}(m_i, m_k)\Big(\frac{\partial f}{\partial m_i}\Big)_0\Big(\frac{\partial g}{\partial m_k}\Big)_0+\operatorname{cov}(m_i, m_l)\Big(\frac{\partial f}{\partial m_i}\Big)_0\Big(\frac{\partial g}{\partial m_l}\Big)_0$$
$$+\operatorname{cov}(m_j, m_k)\Big(\frac{\partial f}{\partial m_j}\Big)_0\Big(\frac{\partial g}{\partial m_k}\Big)_0+\operatorname{cov}(m_j, m_l)\Big(\frac{\partial f}{\partial m_j}\Big)_0\Big(\frac{\partial g}{\partial m_l}\Big)_0$$
$$+O(n^{-3/2}) \tag{8.14}$$

PROOF (8.12) and (8.13) are proved in Cramér [8]. (8.14) is proved in the same manner, using the same notation as Cramér's. Denote by Z the set of all points in $\mathfrak{R}_n$ such that all inequalities $|m_i-\mu_i|<\varepsilon$, $|m_j-\mu_j|<\varepsilon$, $|m_k-\mu_k|<\varepsilon$ and $|m_l-\mu_l|<\varepsilon$ hold and by Z^* the complementary set of Z. We have then

$$P(Z^*)<\frac{4A}{\varepsilon^{2\kappa}n^{\kappa}}, \qquad P(Z)>1-\frac{4A}{\varepsilon^{2\kappa}n^{\kappa}}$$

where A is a constant independent of n and ε, and κ is a positive integer. If ε is sufficiently small, we have by condition 1) for any point in set Z

$$f(m_i, m_j)=f(\mu_i, \mu_j)+(m_i-\mu_i)\left(\frac{\partial f}{\partial m_i}\right)_0+(m_j-\mu_j)\left(\frac{\partial f}{\partial m_j}\right)_0+R_f$$

and

$$g(m_k, m_l)=g(\mu_k, \mu_l)+(m_k-\mu_k)\left(\frac{\partial g}{\partial m_k}\right)_0+(m_l-\mu_l)\left(\frac{\partial g}{\partial m_l}\right)_0+R_g$$

where

$$R_f=\frac{1}{2}\left[\left(\frac{\partial^2 f}{\partial m_i^2}\right)_1(m_i-\mu_i)^2+2\left(\frac{\partial^2 f}{\partial m_i\partial m_j}\right)_1(m_i-\mu_i)(m_j-\mu_j)+\left(\frac{\partial^2 f}{\partial m_j^2}\right)_1(m_j-\mu_j)^2\right]$$

$$R_g=\frac{1}{2}\left[\left(\frac{\partial^2 g}{\partial m_k^2}\right)_1(m_k-\mu_k)^2+2\left(\frac{\partial^2 g}{\partial m_k\partial m_l}\right)_1(m_k-\mu_k)(m_l-\mu_l)+\left(\frac{\partial^2 g}{\partial m_l^2}\right)_1(m_l-\mu_l)^2\right]$$

and the suffix 1 denotes the value at some intermediate point between (μ_i, μ_j) and (m_i, m_j) or between (μ_k, μ_l) and (m_k, m_l).

From (8.12) and the similar formula for g, we have

$$\operatorname{cov}(f, g)=E(fg-f_0g_0)=\int_Z (fg-f_0g_0)\,dP+\int_{Z^*}(fg-f_0g_0)\,dP$$

and

$$\left|\int_{Z^*}(fg-f_0g_0)\,dP\right|\leqq 2C^2n^{2p}\frac{4A}{\varepsilon^{2\kappa}n^{\kappa}}=O(n^{2p-\kappa}),$$

Taking κ such that $\kappa>2p+\frac{3}{2}$, we obtain

$$\operatorname{cov}(f, g)=\int_Z (fg-f_0g_0)\,dP+O(n^{-3/2}) \tag{8.15}$$

while

$$\int_Z (fg-f_0g_0)\,dP=\left(\frac{\partial f}{\partial m_i}\right)_0 g_0\int_Z (m_i-\mu_i)\,dP+\left(\frac{\partial f}{\partial m_j}\right)_0 g_0\int_Z (m_j-\mu_j)\,dP$$
$$+\left(\frac{\partial g}{\partial m_k}\right)_0 f_0\int_Z (m_k-\mu_k)\,dP+\left(\frac{\partial g}{\partial m_l}\right)_0 f_0\int_Z (m_l-\mu_l)\,dP$$

$$+\left(\frac{\partial f}{\partial m_i}\right)_0\left(\frac{\partial g}{\partial m_k}\right)_0\int_Z (m_i-\mu_i)(m_k-\mu_k)\,dP$$
$$+\left(\frac{\partial f}{\partial m_i}\right)_0\left(\frac{\partial g}{\partial m_l}\right)_0\int_Z (m_i-\mu_i)(m_l-\mu_l)\,dP$$
$$+\left(\frac{\partial f}{\partial m_j}\right)_0\left(\frac{\partial g}{\partial m_k}\right)_0\int_Z (m_j-\mu_j)(m_k-\mu_k)\,dP$$
$$+\left(\frac{\partial f}{\partial m_j}\right)_0\left(\frac{\partial g}{\partial m_l}\right)_0\int_Z (m_j-\mu_j)(m_l-\mu_l)\,dP$$
$$+\int_Z R_{fg}\,dP, \tag{8.16}$$

where R_{fg} consists from R_f, R_g, $(m_i-\mu_i)$, $(m_j-\mu_j)$, $(m_k-\mu_k)$ and $(m_l-\mu_l)$.
For example

$$\int_Z (m_i-\mu_i)\,R_g\,dP<\frac{1}{2}M\{E(m_i-\mu_i)(m_k-\mu_k)^2+2E(m_i-\mu_i)(m_k-\mu_k)(m_l-\mu_l)$$
$$+E(m_i-\mu_i)(m_l-\mu_l)^2\}$$

where M denotes a certain constant defined from the condition 1.) By Schwarz' inequality we have

$$|E(m_i-\mu_i)(m_k-\mu_k)^2|\leqq\left[\int_{\mathfrak{R}_n}(m_i-\mu_i)^2\,dP\cdot\int_{\mathfrak{R}_n}(m_k-\mu_k)^4\,dP\right]^{\frac{1}{2}}$$
$$=O(n^{-3/2})$$
$$|E(m_i-\mu_i)(m_k-\mu_k)(m_l-\mu_l)|\leqq\left[\int_{\mathfrak{R}_n}(m_i-\mu_i)^4\,dP\right.$$
$$\left.\times\int_{\mathfrak{R}_n}(m_k-\mu_k)^4\,dP\right]^{\frac{1}{4}}\left[\int_{\mathfrak{R}_n}(m_l-\mu_l)^2\,dP\right]^{\frac{1}{2}}=O(n^{-3/2})$$

Therefore

$$\int_Z (m_i-\mu_i)\,R_g\,dP\leqq O(n^{-3/2})$$

and

$$\int_Z R_{fg}\,dP\leqq O(n^{-3/2}).$$

Besides this relation it holds

$$\int_Z (m_i-\mu_i)\,dP=E(m_i-\mu_i)+O(^{-3/2})=O(n^{-3/2})$$
$$\int_Z (m_i-\mu_i)(m_j-\mu_j)\,dP=\text{cov}\,(m_i,m_j)+O(n^{-3/2}).$$

From (8.15) and (8.16) we have (8.14). This completes the proof of the lemma.

This lemma holds also for the case of several arguments. In the case of an arbitrary number of strata, we have $D^2(r)$ in a similar manner, but for the expectation of $D^2(r)$ we obtain only the terms except the ones containing $(\overline{X}_i-\overline{X}_j)$, $(\overline{Y}_i-\overline{Y}_j)$ or its power neglecting the infinitesimal higher order terms. Using lemma 3 where $m_i=\bar{x}_i$ etc., we have

$$D^2\left[\sum_{i=1}^{R}\frac{N_i}{N}s_{ix}^2+\sum_{i=1}^{R}\frac{N_i}{N}(\bar{x}_i-\bar{x})^2\right]=\sum_{i=1}^{R}\frac{N_i^2}{N^2n_i}(\mu_{40}(i)-\mu_{20}^2(i))$$
$$+I_1\text{ (terms that contain }\overline{X}_i-\overline{X}_j\text{ or its powers)}+O(n_i^{-3/2}) \quad (8.17)$$

$$D^2\left[\sum_{i=1}^{R}\frac{N_i}{N}r_is_{ix}s_{iy}+\sum_{i=1}^{R}\frac{N_i}{N}(\bar{x}_i-\bar{x})(\bar{y}_i-\bar{y})\right]=\sum_{i=1}^{R}\frac{N_i^2}{N^2n_i}(\mu_{22}(i)-\mu_{11}^2(i))$$
$$+I_2\text{ (terms that contain }\overline{X}_i-\overline{X}_j\text{ and }\overline{Y}_i-\overline{Y}_j\text{ or their powers)}$$
$$+O(n_i^{-3/2}) \quad (8.18)$$

$$\operatorname{cov}\left[\sum_{i=1}^{R}\frac{N_i}{N}s_{ix}^2+\sum_{i=1}^{R}\frac{N_i}{N}(\bar{x}_i-\bar{x})^2,\quad \sum_{i=1}^{R}\frac{N_i}{N}r_is_{ix}s_{iy}+\sum_{i=1}^{R}\frac{N_i}{N}(\bar{x}_i-\bar{x})(\bar{y}_i-\bar{y})\right]$$
$$=\sum_{i=1}^{R}\frac{N_i^2}{N^2n_i}(\mu_{31}(i)-\mu_{11}(i)\mu_{20}(i))+I_3\text{ (terms that contain }\overline{X}_i-\overline{X}_j\text{ and}$$
$$\overline{Y}_i-\overline{Y}_j\text{ or their powers)}+O(n_i^{-3/2}) \quad (8.19)$$

$$\operatorname{cov}\left[\sum_{i=1}^{R}\frac{N_i}{N}s_{ix}^2+\sum_{i=1}^{R}\frac{N_i}{N}(\bar{x}_i-\bar{x})^2,\quad \sum_{i=1}^{R}\frac{N_i}{N}s_{iy}^2+\sum_{i=1}^{R}\frac{N_i}{N}(\bar{y}_i-\bar{y})^2\right]$$
$$=\sum_{i=1}^{R}\frac{N_i^2}{N^2n_i}(\mu_{22}(i)-\mu_{20}(i)\mu_{02}(i))+I_4\text{ (terms that contain }\overline{X}_i-\overline{X}_j\text{ and}$$
$$\overline{Y}_i-\overline{Y}_j\text{ or their powers)}+O(n_i^{-3/2}) \quad (8.20)$$

hence

$$D^2(r)=\frac{1}{\mu_{20}\mu_{02}}\left[\frac{\mu_{11}^2}{4\mu_{20}^2}\sum_{i=1}^{R}\frac{N_i^2}{N^2n_i}(\mu_{40}(i)-\mu_{20}^2(i))+\frac{\mu_{11}^2}{4\mu_{02}^2}\sum_{i=1}^{R}\frac{N_i^2}{N^2n_i}(\mu_{04}(i)-\mu_{02}^2(i))\right.$$
$$+\sum_{i=1}^{R}\frac{N_i^2}{N^2n_i}(\mu_{22}(i)-\mu_{11}^2(i))-\frac{\mu_{11}}{\mu_{20}}\sum_{i=1}^{R}\frac{N_i^2}{N^2n_i}(\mu_{31}(i)-\mu_{11}(i)\mu_{20}(i))$$
$$-\frac{\mu_{11}}{\mu_{02}}\sum_{i=1}^{R}\frac{N_i^2}{N^2n_i}(\mu_{13}(i)-\mu_{11}(i)\mu_{02}(i))+\frac{\mu_{11}^2}{2\mu_{20}\mu_{02}}\sum_{i=1}^{R}\frac{N_i^2}{N^2n_i}(\mu_{22}(i)-\mu_{20}(i)\mu_{02}(i))$$
$$+I\text{ (terms that contain }X_i-X_j\text{ and }Y_i-Y_j\text{ or their powers)}$$
$$+O(n_i^{-3/2}) \quad (8.21)$$

In the case of the proportionate sampling, we have

$$D^2(r)_{\text{prop.}}=\frac{\rho^2}{4n}\Big[\frac{1}{\mu_{20}^2}\sum_i\frac{N_i}{N}(\mu_{40}(i)-\mu_{20}^2(i))+\frac{1}{\mu_{02}^2}\sum_i\frac{N_i}{N}(\mu_{04}(i)-\mu_{02}^2(i))$$

$$+\frac{4}{\mu_{11}^2}\sum_i\frac{N_i}{N}(\mu_{22}(i)-\mu_{11}^2(i))-\frac{4}{\mu_{11}\mu_{20}}\sum_i\frac{N_i}{N}(\mu_{31}(i)-\mu_{11}(i)\mu_{20}(i))$$

$$-\frac{4}{\mu_{11}\mu_{02}}\sum_i\frac{N_i}{N}(\mu_{13}(i)-\mu_{11}(i)\mu_{02}(i))+\frac{2}{\mu_{20}\mu_{02}}\sum_i\frac{N_i}{N}(\mu_{22}(i)-\mu_{20}(i)\mu_{02}(i))$$

$$+I_1+O(n^{-3/2})$$

$$=\frac{\rho^2}{4n}\Big[\frac{\mu_{40}}{\mu_{20}^2}+\frac{\mu_{04}}{\mu_{02}^2}+\frac{4\mu_{22}}{\mu_{11}^2}-\frac{4\mu_{31}}{\mu_{11}\mu_{20}}-\frac{4\mu_{13}}{\mu_{11}\mu_{02}}+\frac{2\mu_{22}}{\mu_{20}\mu_{02}}$$

$$-\frac{1}{\mu_{20}^2}\sum_i\frac{N_i}{N}\mu_{02}^2(i)-\frac{1}{\mu_{02}^2}\sum_i\frac{N_i}{N}\mu_{02}^2(i)-\frac{4}{\mu_{11}^2}\sum_i\frac{N_i}{N}\mu_{11}^2(i)$$

$$+\frac{4}{\mu_{11}\mu_{20}}\sum_i\frac{N_i}{N}\mu_{11}(i)\mu_{20}(i)+\frac{4}{\mu_{11}\mu_{20}}\sum_i\frac{N_i}{N}\mu_{11}(i)\mu_{02}(i)$$

$$-\frac{2}{\mu_{20}\mu_{02}}\sum_i\frac{N_i}{N}\mu_{20}(i)\mu_{02}(i)\Big]+I_2+O(n^{-3/2}) \tag{8.22}$$

so that taking expectation $\mathcal{E}$ in regard to the stratification we obtain

$$\mathcal{E}D^2(r)_{\text{prop.}}=\frac{\rho^2}{4n}\Big(\frac{\mu_{40}}{\mu_{20}^2}+\frac{\mu_{04}}{\mu_{02}^2}+\frac{4\mu_{22}}{\mu_{11}^2}-\frac{4\mu_{31}}{\mu_{11}\mu_{21}}-\frac{4\mu_{13}}{\mu_{11}\mu_{12}}+\frac{2\mu_{22}}{\mu_{20}\mu_{02}}\Big)$$

$$+O(n^{-3/2}) \tag{8.23}$$

because, in the same manner as taking the ordinary expectation of lemma 3, we have

$$\mathcal{E}\mu_{20}^2(i)=\mu_{20}^2+O(N_i^{-1})$$
$$\mathcal{E}\mu_{11}^2(i)=\mu_{11}^2+O(N_i^{-1})$$
$$\mathcal{E}\mu_{11}(i)\mu_{20}(i)=\mu_{11}\mu_{20}+O(N_i^{-1})$$
$$\mathcal{E}\mu_{20}(i)\mu_{02}(i)=\mu_{20}\mu_{02}+O(N_j^{-1})$$
$$\mathcal{E}(\overline{X}_i-\overline{X}_j)^l\mu_{hk}(i)=O(N_i^{-1})$$

From this result we can estimate the correlation coefficient in some cases of the stratified proportionate samplings with the better accuracy than the simple random sampling.

9 Application to the sampling inspection

We need often the method of sampling inspection for the sub-sampling. Let the size of a lot be N and assume that the lot is partitioned into M sublots of equal size. The plan of the sampling inspection consists of how many sublots we should sample and how many units we should sample from these sublots.

Let P be the total fraction defective, P_i that of the i-th sublot, N_i the number of defectives of the i-th sublot, m the size of samples of these sublots and n the size of sample units from each sublot.

An optimum plan under the condition of the constant cost is obtained in the following way. First, set

c_1: the cost of the inspection per sublot
c_2: the loss by the interchange of defectives
c_3: the cost of the inspection per unit
C: total cost

Then the total cost on the average is

$$C=(c_1+c_2nP+c_3n)m$$

and the variance of the estimated fraction defectives p is

$$D^2(p)=\frac{M-m}{M-1}\frac{\sigma_b^2}{m}+\frac{\frac{N}{M}-n}{\frac{N}{M}-1}\frac{\sigma_w^2}{mn}\doteqdot\frac{\sigma_b^2}{m}+\frac{\sigma_w^2}{mn} \tag{9.1}$$

where σ_b^2 is the between-variance and σ_w^2 is the within-variance:

$$\sigma_w^2=\frac{1}{M}\sum_{j=1}^{M}P_j(1-P_j) \tag{9.2}$$

and

$$\sigma_b^2=\frac{1}{M}\sum_{j=1}^{M}(P_j-P)^2 \tag{9.3}$$

Therefore, putting $f=D^2(p)+\lambda\{(c_1+c_2nP+c_3n)m-C\}$ and $\sigma_b^2=k\sigma_w^2$, we have from $\frac{\partial f}{\partial m}=0,\ \frac{\partial f}{\partial n}=0$

$$n=\sqrt{\frac{c_1}{k(c_2P+c_3)}} \tag{9.4}$$

and

$$m=\frac{C}{c_1+(c_2P+c_3)n} \tag{9.5}$$

which give an optimum plan.

In order to write the OC-function it is not sufficient to use the equation (9.1). Because we don't know the distribution of the estimated fractives p. We shall consider this problem below.

From (9.1), (9.2) and (9.3) we have

$$D^2(p)=\frac{M-m}{M-1}\frac{1}{Mm}\sum_{j=1}^{M}(P_j-P)^2+\frac{N-Mn}{N-M}\frac{1}{mnM}\sum_{j=1}^{M}P_j(1-P_j) \tag{9.6}$$

If the partition into sublot is performed at random, and if we take the expectation $\mathcal{E}$ of $D^2(p)$ as in section 4, we get

$$\mathcal{E}D^2(p)=\frac{M-m}{M-1}\frac{1}{Mm}\mathcal{E}(\sum_j(P_j-P)^2)+\frac{N-Mn}{N-M}\frac{1}{mnM}\mathcal{E}(\sum_j P_j(1-P_j))$$

When we put $\mathrm{I}=\sum_j(P_j-P)^2$, $\mathrm{II}=\sum_j P_j(1-P_j)$, this can be written as

$$=\frac{M-m}{M-1}\frac{1}{Mm}\mathcal{E}(\mathrm{I})+\frac{N-Mn}{N-M}\frac{1}{mnM}\mathcal{E}(\mathrm{II}) \tag{9.7}$$

Now, considering $P_j=MN_j/N$ and $\sum_j P_j=MP$, we obtain

$$\mathrm{I}=\sum_j P_j^2-MP^2=\frac{M^2}{N^2}\sum_j N_j^2-MP^2 \tag{9.8}$$

$$\mathrm{II}=\sum_j P_j-\sum_j P_j^2=MP-\frac{M^2}{N^2}\sum_j N_j^2 \tag{9.9}$$

so that we can calculate the expectation of $\sum_j N_j^2$.

Using the result of section 4, we have

$$\mathcal{E}(\sum_j N_j^2)=\frac{N-M}{N-1}\frac{NP(NP-1)}{M}+NP \tag{9.10}$$

Therefore, we have

$$\mathcal{E}(\mathrm{I})=M\cdot\frac{N-M}{N-1}\cdot\frac{P(NP-1)}{N}+\frac{PM^2}{N}-MP^2 \tag{9.11}$$

$$\mathcal{E}(\mathrm{II})=MP-M\cdot\frac{N-M}{N-1}\cdot\frac{P(NP-1)}{N}-\frac{PM^2}{N} \tag{9.12}$$

and from (9.7)

$$\mathcal{E}D^2(p)=\frac{M-m}{M-1}\frac{1}{m}\left\{\frac{N-M}{N-1}\cdot\frac{P(NP-1)}{N}+\frac{PM}{N}-P^2\right\}$$

$$+\frac{N-Mn}{N-M}\frac{1}{mn}\left\{P-\frac{N-M}{N-1}\cdot\frac{P(NP-1)}{N}-\frac{PM}{N}\right\} \tag{9.13}$$

Further, when we can assume that the distribution of p is approximately normal with mean P and variance $\mathcal{E}^2D(p)$, we can write the OC-curve as usual.

Example. Consider the case when

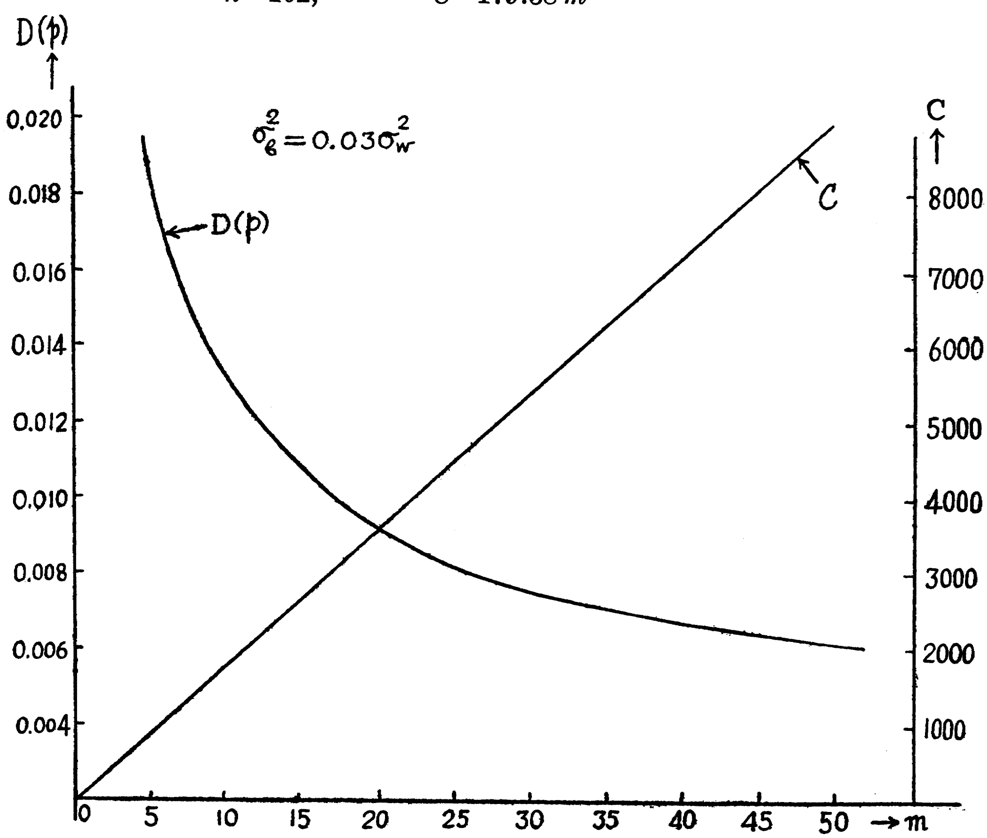
$N=120000$, $M=500$, $P=0.04$

$c_1=135$, $c_2=8.5$, $c_3=0.05$ (yen)

$k=0.03$,

In this case we have from (9.4) and (9.5)

$n=102$, $C=179.38\,m$



If we take the following plan:

$n=100$

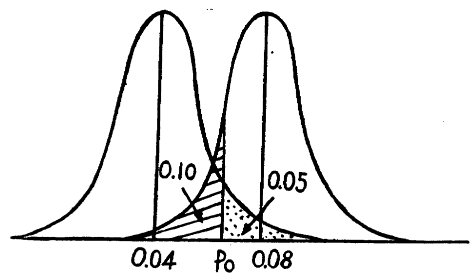
$m=\ 10$

then we obtain from (9.13)

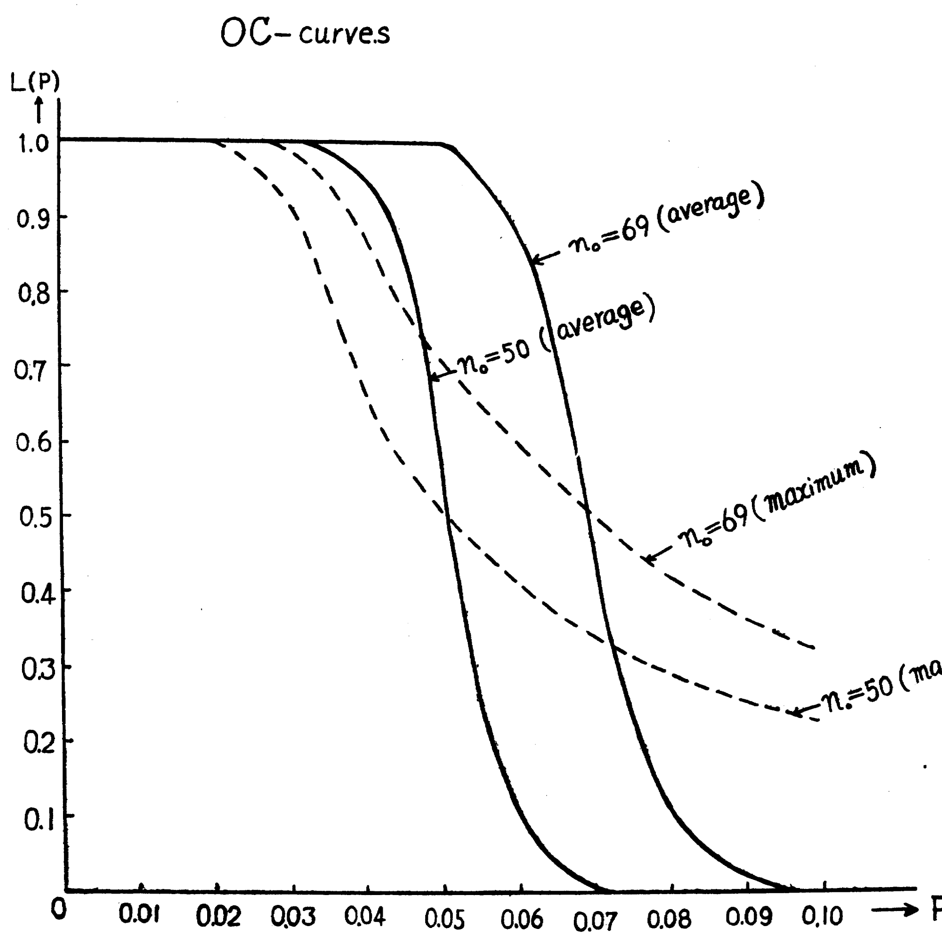
$$\varepsilon D^2(p)=0.000038086 \qquad \text{for the case } P=0.04$$

Let the lot tolerance percentage defective (L.T.P.D.) p_t be 0.08, the number of the defectives which are tolerable equal to 80.

P	$\varepsilon D^2(p)$	$\sqrt{\varepsilon D^2(p)}$
0.01	.000 009 818	.0031
0.02	.000 019 438	.0044
0.03	.000 028 861	.0054
0.04	.000 038 086	.0062
0.05	.000 047 114	.0069
0.06	.000 055 944	.0075
0.07	.000 064 576	.0080
0.08	.000 073 012	.0085
0.09	.000 081 248	.0090
0.10	.000 089 288	.0094



Then the maximum acceptable number of defective is $n_0 = mnp_0 = 1000\,p_0$ where p_0 satisfies either



$$\frac{0.08-p_0}{\sqrt{\varepsilon D^2(p) \text{ for } P=0.08}}=1.2816 \tag{9.14}$$

for the consumer's risk $\beta=0.10$, or

$$\frac{p_0-0.04}{\sqrt{\varepsilon D^2(p) \text{ for } P=0.04}}=1.6449 \tag{9.15}$$

for the producer's risk $\alpha=0.05$.

Using the above table we obtain approximately from (9.14) $p_0=0.0691$ or from (9.15) $p_0=0.0502$. If we attach importance to the consumer's risk, we put $n_0=69$ so that the producer's risk is almost 0, as is seen from the fact that

$$\frac{0.0691-0.04}{0.0062}=4.69, \quad \frac{1}{\sqrt{2\pi}}\int_{4.69}^{\infty} e^{-\frac{t^2}{2}}dt \doteqdot 0.$$

If we attach importance to the producer's risk, we put $n_0=50$ so that the consumer's risk is almost 0, as is seen from the fact that

$$\frac{0.08-0.0502}{0.0085}=3.505, \quad \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-3.505} e^{-\frac{t^2}{2}}dt \doteqdot 0.$$

Therefore we have average OC-curves for these plans. In this graph we show also the OC-curves for the maximum variance $\varepsilon D^2(p)+3\vartheta D^2(p)$ as it is practically used in place of the average variance $\varepsilon D^2(p)$.

10 Conclusion from our results

In the preceding sections we discussed about methods of stratification and analysis, where we have found that the proportionate stratified sampling plays the principal role. But we cannot often analyse our data for the arbitrary stratified sampling because of a complicated analysing procedure. Therefore, except for estimation of the mean or the total value, we must use the proportionate stratified sampling for every purpose, which satisfies the necessary conditions for analysis and takes samples of large size. In this case the sample size is determined as follows.

Let R be the number of strata, $c_i(1\leqq i\leqq R)$ the cost of survey per unit, and c_0 the cost of stratification. Then the total cost $C(n)$ is

$$C(n)=c_0+\sum_{i=1}^{R}c_i n_i \tag{10.1}$$

But in our survey where we use the proportional allocation, we have

$$C(n)=c_0+\frac{n}{N}\sum_{i=1}^{R}c_iN_i$$

where N denotes total size of our universe and N_i the size of the i-th stratum.

Suppose we want to estimate K different parameters $Z_j(1\leqq j\leqq K)$ and the loss function $l(z_j)$ is equal to $w_j(z_j-Z_j)^2$ with the weight w_j. Then the expectation of $l(z_j)$ becomes $w_jD^2(z_j)$ so that the expectation of the loss function $L(n)$ is in regard to estimation

$$\sum_{j=1}^{K}w_jD^2(z_j) \tag{10.3}$$

Hence in case when $L(n)$ can practically considered to be $\sum_{j=1}^{K}w_jD^2(z_j)$, the sample size n which minimizes the total loss function

$$f(n)=L(n)+C(n) \tag{10.4}$$

can be found from the following equation

$$\frac{df}{dn}=\sum_{j=1}^{K}w_j\frac{dD^2(z_j)}{dn}+\frac{1}{N}\sum_{i=1}^{R}c_iN_i=0 \tag{10.5}$$

In many cases, however, we have $D^2(z_j)\doteqdot\frac{A_j}{n}$ where A_j is the function of parameters not including n, then we can get from (10.5)

$$n=\sqrt{\frac{N\sum_{j=1}^{K}w_jA_j}{\sum_{i=1}^{R}c_iN_i}}. \tag{10.6}$$

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Hayashi, C., F, Maruyama, and M. Ishida: On some criteria for stratification, *Annals of the Institute of Statistical Mathematics*, vol. II, No. 2, 1951.

[2] Dalenius, T.: The problem of optimum stratification, *Skandinavisk Aktuarietidskrift*, 1950.

[3] Aoyama, H.: On the double sampling, *Memoir of the Institute of Statistical Mathematics*, vol. 6, No. 6, 1950 (In Japanese).

[4] Madow, W. G. and Madow, L. H.: On the theory of systematic sampling, I, *Annals of Mathematical Statistics*, vol. 15, 1944.

[5] Aoyama, H.: On practical systematic sampling, *Annals of the Institute of Statistical Mathematics*, vol. III, No. 2, 1952.

[6] Aoyama, H.: On the chi-square test for weighted samples, *Annals of the Institute of Statistical Mathematics*, vol. V, No. 1, 1953. Errata, *this issue of the Annals*.

[7] Wilks, S. S.: *Mathematical Statistics*, Princeton, 1943.

[8] Cramér, H.: *Mathematical Methods of Statistics*, Princeton, 1946.

ERRATA

These Annals, Vol. II, No. 1, 1950. P. 18 insert after last line of section 2

"In these cases we have not the maximum value but only the stationary value just as the minimax solution. If we want to obtain the maximum value, we must estimate the rational rate k_1 and k_2 from experiences in the past time. This fact holds also in the following sections."

Vol. V, No. 1, 1953

Page	line				
27,	9,	read	$M-1-\frac{(M-1)(R-1)}{N}$	instead of	the right hand side of (6)
27,	12,	insert under the assumption after "we have" $N_i=N/R$			
27,	14,	read	$2(M-1)+O\left(\frac{1}{N}\right)+O\left(\frac{1}{n}\right)$	instead of	the right hand side of (8)
27,	last,	read	(strike off the table)		
28,	5-6,	read	(strike off the sentence "under the condition $M=R(R-1)$ and $R \neq 1$"		

Vol. VI, No. 1, 1954

Page	line				
13,	12,	read	$\binom{M}{Mp_i}p^{Mp_i}q^{Mq_i}$	instead of	$\binom{M}{Mp_i}p^{Mp_i}p^{Mq_i}$
14,	3,	read	$0.96\,N$	instead of	$096\,N$
15,	6,	read	$\ldots k\sqrt{\varepsilon^* D^2(X)}\} \leqq \frac{1}{k^2}$	instead of	$\ldots k\sqrt{\varepsilon^* D^2(X)} \leqq \frac{1}{k^2}$
15,	23,	read	$X_{(i)}$	instead of	$X_{i)}$
24,	7,	read	$-\mu_{11}(2)\mu_{20}(2)\ldots$	instead of	$-\mu_{11}(1)\mu_{20}(2)\ldots$
24,	10,	read	$\frac{N_1^2N_2}{N^3}((\bar{X}_1-\bar{X}_2)\ldots$	instead of	$\frac{N_1^2N_2}{N^3}(\bar{X}_1-\bar{X}_2)\ldots$
25,	9,	read	$\frac{2N_1N_2}{N^3}(\bar{Y}_1-\bar{Y}_2)^2\cdots$	instead of	$\frac{2N_1N_2}{N^2}(\bar{Y}_1-\bar{Y}_2)^2\cdots$
			$+\frac{N_1N_2}{N^5}(N_1^3+N_2^3)\cdots$	instead of	$+\frac{N_1N_2}{N^5}(N_1^2+N_2^3)\cdots$
28,	2 from the bottom,		$+O(n^{-3/2})$	instead of	$+O(^{-3/2})$
30,	11,	read	$-\frac{4\mu_{31}}{\mu_{11}\mu_{20}}-\frac{4\mu_{13}}{\mu_{11}\,\mu_{02}}+\cdots$	instead of	$-\frac{4\mu_{31}}{\mu_{11}\mu_{21}}-\frac{4\mu_{13}}{\mu_{11}\mu_{12}}+\cdots$
36,	3 from the bottom,		the coming issue	instead of	this issue

Page	line				
54,	6,	read	[20], Lemma	instead of	[20 , Lemma
68,	28,	read	$e^{I_{\varepsilon}(t)}$	instead of	$eI^{\varepsilon(t)}$
97,	6,	read	(X_j, Y_j) has	instead of	$(X_j, Y_j$ has
98,	2,	read	$\lim_n \frac{D(S_{nj})}{D_n}$	instead of	$\lim_n \sum_l \frac{D(S_{nj})}{D_n}$