Janusz Wywiał

SOME CONTRIBUTIONS TO MULTIVARIATE METHODS IN SURVEY SAMPLING

SOME CONTRIBUTIONS TO MULTIVARIATE METHODS IN SURVEY SAMPLING

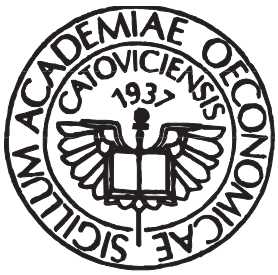
PRACE NAUKOWE

AKADEMII EKONOMICZNEJ
IM. KAROLA ADAMIECKIEGO
W KATOWICACH

University of Economics in Katowice

Janusz Wywiał

SOME CONTRIBUTIONS TO MULTIVARIATE METHODS IN SURVEY SAMPLING



KATOWICE 2003

This publication has been supported by the grant number 2 H02B 004 22 from the State Committee for Scientific Research (Komitet Badań Naukowych)

ISBN 83-7246-273-9

Printed in Poland

Wydawnictwo Akademii Ekonomicznej
im. Karola Adamieckiego w Katowicach

ul. 1 Maja 50, 40-287 Katowice, tel.fax (032) 25 77 643
e-mail:wydawucz@sulu.ae.katowice.pl

CONTENTS

INTRODUCTION

The book can be treated as a set of contributions to the estimation of a vector of the averages of variables in a finite population. The methods presented are not only a simple generalisation of the well known problems on a multidimensional case but a lot of them can be treated as original ones.

Particularly, several sampling strategies dependent on auxiliary variables are proposed. The problems of optimising a sample size are considered in detail for stratified and two-stage sampling designs in the case when more than one average in a population is estimated. The well known discrimination and clustering methods and their modifications are used for optimal stratification or clustering of a fixed population. Solutions obtained here can be useful in optimisation of estimation on the basis of a double sample.

The book presents some contributions to interpretations of the following measures of accuracy of vector estimators: the generalised variance, the mean radius and spectral radius defined as a determinant, the trace and the maximal eigenvalue of the variance-covariance matrix, respectively. Some definitions and theorems, known in a one-dimensional case are extended to the vector estimation case. They let us compare the accuracy of vector estimators. The properties of sampling designs and sampling schemes depend on the parameters of auxiliary variables like the sample generalised variance, the squared difference between the sample mean and the population mean are considered. The approximate expressions of the variance of the Horovitz-Thompson estimator of the mean value are derived for these sampling designs. The unbiased estimators of the generalised variance are found in the cases when the simple sample is drawn with as well as without replacement.

The basic properties of the vector of the regression estimators are derived. It is proven that the vector of regression estimators is efficient in the class of the vector of the difference estimators in the case of a simple sample.

Let the double sample consist of the following two samples: the first one is a simple sample drawn without replacement from a population, the other one is also a simple sample but selected from the first sample. Several problems concerning the optimisation of determining the size of the above two samples are formulated and solved. The square risk function (or the generalised variance of the vector of the regression estimators) is minimised under the fixed total cost of observation of variables in the double sample.

Next, the cost function is minimised under the fixed variances of the regression estimators of the means of particular variables either under the fixed value of the generalised variance of the vector estimators or under the fixed square risk function.

Similar optimisation problems are formulated and solved to determine the sizes of a stratified sample or a two-stage one. Moreover, the sizes of the samples drawn from given strata are determined through minimisation of the spectral radius of the vector of the sample means under the fixed total cost of observations of variables in the sample. It is shown that in the case of the proportional allocation of the sample in the strata, the mean vector of the stratified sample means is not less accurate than the mean vector of the simple sample means.

The multidimensional auxiliary variables are used to stratify a population. Strata can be obtained on the basis of the well known cluster method by Ward or the k-mean method or their modifications. For instance, in the case of the regressive superpopulation, the strata are selected through minimisation of the spectral radius of the intra-strata variance-covariance matrix of auxiliary variables.

A new class of classification estimators is proposed. They are constructed as linear combinations of sample means from sub-samples determined after a sample selection. The weights of this combination are functions of sizes of subsets of a population determined by some classification procedures. This method can be explained as a procedure of stratifying a population after a sample selection on the basis of the auxiliary variable observed in the population and in the sample.

The properties of the mean vector from the cluster sample are studied. Its variance-covariance matrix is expressed as a function of the introduced matrix of the coefficients of the intra-cluster correlation. It is proven that the vector of the cluster means is not a less accurate estimator of the vector of population averages than the vector of the simple sample means when the matrix of the coefficients of the intra-cluster correlation is defined as non-positive. A new method of dividing a fixed and finite population into groups of the same size on the basis of a multidimensional auxiliary variable is proposed. This method maximises the intra-cluster scatter of the observations of a multidimensional auxiliary variable.

The book can be treated as lecture notes. The readership consist of undergraduate statistics students, statisticians interested in survey sampling methods.

I. FOUNDATION OF SAMPLING STRATEGIES

1.1. Fixed population approach

Let us introduce the following notation: A finite population is a collection of N unit $\Omega=\{\omega_1,....,\omega_N\}$, where $N < \infty$, and N is called the size of the population. The units of a finite population are said to be identifiable if they can be labeled from 1 to N and the label of each unit is known, see e.g. Cassel, Sarndal and Wretman (1977). The label k represents the unit ω_k of a population and that is why a population can be denoted as the set of the natural numbers: $\Omega=\{1,2,...N\}$.

An m-dimensional variable is denoted by $y=[y_1...y_m]$ and its observations are columns of the matrix $\mathbf{y}=[y_{ij}]$, i=1,...,N, j=1,...,m, where y_{ij} is the value of the j-th variable attached to the i-th unit.

Definition 1.1: The matrix $\mathbf{y}$ is called a parameter of the population if to each unit $i\in\Omega$ is attached the i-th row of the matrix $\mathbf{y}$.

The space of the population parameter $\mathbf{y}$ is denoted by $\mathbf{Y}\subseteq R^{Nm}$. The vector of auxiliary variables will be denoted by $x=\{x_1,....,x_p\}$ and the matrix of their observations by $\mathbf{x}=[x_{ij}]$ of dimensions N×p. The matrix $\mathbf{y}$ as well as $\mathbf{x}$ are the parameters of a population. Let $\mathbf{y}_{*j}$ be the j-th column of the matrix $\mathbf{y}$ and $\mathbf{y}_{i*}$ be the i-th row of this matrix.

Definition 1.2: Any parametric function $\boldsymbol{\theta}: \mathbf{Y} \rightarrow \Theta\subseteq R^m$, such that $\theta_i=\theta_i(\mathbf{y})$, i=1,...,m is called parametric function or the vector of described parameters.

The set of possible values of the vector is denoted by $\boldsymbol{\Theta}$ and called the space of described parameter vector. The elements of the vector $\boldsymbol{\theta}$ characterize properties of variables. The following vectors of described parameters are frequently used in statistical research:

The vector of mean values is denoted by $\bar{\mathbf{y}}=[\bar{y}_1...\bar{y}_m]^T$, where:

$$\bar{\mathbf{y}} = N^{-1}\mathbf{y}^T\mathbf{J}_N \tag{1.1}$$

where $\mathbf{J}_N$ is the unit column vector of dimensions N×1.

The vector of the totals of variables is as follows:

$$\tilde{\mathbf{y}} = N\bar{\mathbf{y}} \tag{1.2}$$

The matrix of variances and covariances of variables will be denoted by $\mathbf{C}_*(y)=[c_{*tj}]$ (t,j=1,...,m), where:

$$c_{*tj} = c_*(y_t, y_j) = \frac{1}{N-1}\sum_{i=1}^{N}(y_{it} - \bar{y}_t)(y_{ij} - \bar{y}_j) \tag{1.3}$$

Particularly, the variance of a variable y_j is defined by the expression $v_*(y_j)=c_*(y_j, y_j)$.

The correlation matrix will be denoted by: $\mathbf{R} = \mathbf{R}(y) = [r_{tj}]$ (t,j=1,...,m), where:

$$r_{tj} = \frac{c(y_t, y_j)}{\sqrt{v_*(y_t)v_*(y_j)}} \tag{1.4}$$

Definition 1.3: Let us define the mean radius of an m-dimensional variable as the square root of the trace of variance-covariance matrix:

$$q_*(y) = \sqrt{\mathrm{tr}\mathbf{C}_*(y)} \tag{1.5}$$

The parameter $q_*(y)$ is the square root of the mean of the squared distances among the points whose co-ordinates are equal to appropriate observations of variables from the point whose co-ordinates are equal to the averages of those variables

Definition 1.4: [Wilks (1932)]: The generalized variance of an m-dimensional variable is equal to the determinant of the variance-covariance matrix. Hence:

$$g(y) = \det(\mathbf{C}_*(y)) \tag{1.6}$$

From the geometrical point of view the generalized variance can be interpreted in several ways. Let $\mathbf{e}_{*j} = \mathbf{y}_{*j} - \mathbf{J}_N \bar{y}_j$ be the vector of the scores.

Theorem 1.1: [Anderson (1958), p. 167]**:** The generalized variance g(*y*) is proportional to the squared volume of the parallelogram spanned on vectors, all attached to the origin point $\mathbf{o}_N$. The ends of these vectors are in the appropriate points $\mathbf{e}_{*i},...,\mathbf{e}_{*m}$ in an N-dimensional space.

Let $m(\mathbf{y}_{i_1*},...,\mathbf{y}_{i_m*},\mathbf{y}_{i_{m+1}*})$ be the volume (measure) of the m-dimensional parallelogram spanned on the vectors which are all attached to the point $\mathbf{y}_{i_{m+1}*}$, and the ends of this vectors are in the appropriate points $\mathbf{y}_{i_1*},...,\mathbf{y}_{i_m*}$. This volume is determined by the equation[1]:

$$m(\mathbf{y}_{i_1*},..,\mathbf{y}_{i_m*},\mathbf{y}_{i_{m+1}*}) = \left|\det\begin{bmatrix} y_{i_1*} - y_{i_{m+1}*} \\ ..... \\ y_{i_m*} - y_{i_{m+1}*} \end{bmatrix}\right| \tag{1.7}$$

[1] See, e.g. Borsuk (1969).

Let $m(\mathbf{y}_{i_1*},...,\mathbf{y}_{i_m*},\bar{\mathbf{y}})$ be the volume (measure) of the m-dimensional parallelogram spanned on the vectors which are all attached to the point $\bar{\mathbf{y}}$ and their ends are in the points: $\mathbf{y}_{i_1*},...,\mathbf{y}_{i_m*}$. Anderson (1958), p. 168-170, proved the following property.

Theorem 1.2: The generalized variance g(*y*) is proportional to the following sum of squared volumes spanned on the vectors which are all attached to the point $\bar{\mathbf{y}}$ and whose ends have co-ordinates which are appropriate m-element combinations of rows of the matrix **y**:

$$g(y) = N^{-m} \sum_{\{i_1,..,i_m\}} m^2(\mathbf{y}_{i_1*},...\mathbf{y}_{i_m*},\bar{\mathbf{y}}) \tag{1.8}$$

Theorem 1.3: [Wywiał (1992a)]: The generalized variance g(*y*) is proportional to the following sum of squared volumes spanned on the (m+1) points whose co-ordinates are (m+1) - element combinations of rows of the matrix **y**:

$$g(y) = N^{-m-1} \sum_{\{i_1,..,i_{m+1}\}} m^2(\mathbf{y}_{*i_1},...,\mathbf{y}_{*i_m},\mathbf{y}_{*i_{m+1}}) \tag{1.9}$$

Hence, the generalized variance can be used as the coefficient measuring the scatter of observations of a multidimensional variable. When g(*y*) = 0, all observations of an m-dimensional variable are on not more than (m-1)-dimensional hyperplane, see, e.g. Anderson (1958).

Let λ_j (j=1,...,m) be the eigenvalues of the variance-covariance matrix $\mathbf{C}_*(y)$ and $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m$. On the basis of the well known properties of principal components, we know that $\det\mathbf{C}_*(y) = \prod_{i=1}^{m} \lambda_i \geq 0$. Then, for each value λ_j there exists such an (m-1)-dimensional hyperplane $H_{m-1}^{(j)} : \sum_{k=1}^{m} a_k^{(j)} \mathbf{x}_k = 0$ that the mean of squared distances of the points $\mathbf{e}_{i*}=\mathbf{y}_{i*}-\bar{\mathbf{y}}$ (i=1,...N) from $H_{m-1}^{(j)}$ is equal to λ_j. In the case when $\lambda_j>0$, for j=1,..., m_0`-1 and $\lambda_j=0$ for j=$m_0$`, m_0`+1,...,m, points $\mathbf{e}_{i*}$ (i=1,...,N) are on the ($m_0$`-1)-dimensional hyperplane. Particularly, if $\lambda_m = 0$ and $\lambda_{m-1}>0$, then all points $\mathbf{e}_i$ are on the hyperplane $H_{m-1}^{(m)}$.

The parameter λ_j is equal to the variance of the j-th principal component. The values of this principal components are determined by the equation: $u_{ij} = \mathbf{a}^{(j)}\mathbf{e}_{i*}^T$ (i=1,...,N), where: $\mathbf{a}^{(j)} = [a_1^{(j)}...a_m^{(j)}]$ and $\mathbf{a}^{(j)}(\mathbf{C}_*(y)-\mathbf{I}_m\lambda_j)=\mathbf{0}_m$. Hence, $\mathbf{a}^{(j)}$ is the eigenvector of the $\mathbf{C}_*$. The j-th principal component will be denoted by $\mathbf{u}_j=[u_{1j}...u_{Nj}]^T$. From the geometrical point of view, the principal components are obtained through such a rotation of coordinates' system that entire variables are transformed into orthogonal ones called principal components. The vector $\overrightarrow{\mathbf{o}_N\mathbf{u}_j}$ is perpendicular to $\overrightarrow{\mathbf{o}_N\mathbf{u}_i}$ '(j≠t=1,...,m). This leads

us to the equation: $\det \mathbf{C}_*(y) = \prod_{j=1}^{m} \lambda_j$. Then we can infer that the generalized variance is equal to the squared m-dimensional volume of the parallelogram spanned on the vectors: $\overrightarrow{\mathbf{o}_N\mathbf{u}_j}$ (j=1,...,m). The length of the j-th principal component is equal to $\sqrt{\lambda_j}$.

The square root of the maximal eigenvalue λ_1 of the covariance matrix $\mathbf{C}_*(y)$ is called the spectral radius[2], so $\rho(y) = \sqrt{\lambda_1}$.

Let $\mathbf{z} = [z_1...z_N]^T$ be the vector of the observations of the variable *z*. The vector **z** is the linear combination of the columns of the matrix **y** and let $\mathbf{b} = [b_1...b_m]^T$ be the vector of coefficients of this combination and $\mathbf{b}^T\mathbf{b} = 1$. Hence, $\mathbf{z} = \mathbf{yb}$. Hence, the variance of the variable *z* is as follows:

$$v(z) = \mathbf{b}^T\mathbf{C}_*(y)\mathbf{b} \tag{1.10}$$

The well known properties of square forms allows us to conclude that:

$$v(z_1) = \mathbf{b}_1^T\mathbf{C}_*(y)\mathbf{b}_1 = \max_{b^Tb=1} imum \left\{\mathbf{b}^T\mathbf{C}_*(y)\mathbf{b}\right\} = \lambda_1 \tag{1.11}$$

where z_1 is the variable whose observations are expressed by the transformation: $\mathbf{z}_1=\mathbf{yb}_1$. Hence, the spectral radius of the matrix $\mathbf{C}_*$ is equal to a standard deviation $\sqrt{v(z_1)}$ of the variable whose values are the linear combinations of the data matrix **y**. The coefficients of this combination are the elements of such a vector $\mathbf{b}_1$ that the variance $v(z_1)$ takes the maximal value.

Let us consider the following example. The population consists of N households. Let y_{ij} (i=1,...,N;j=1,...,m) be the quantity of the j-th good which is bought by the i-th household. The prices of these goods are elements of the vector: $\mathbf{p}=[p_1...p_m]^T$. The vector of the standardized prices **b** is determined by the equation: $\mathbf{b} = \alpha^{-1}\mathbf{p}$, where $\alpha^2 = \mathbf{p}^T\mathbf{p}$. Let us suppose that $\mathbf{b}_1>\mathbf{0}$. We say that $\mathbf{b}_1$ is the worst standardized price vector if the variance of the household expenditures $z_1 = y\mathbf{b}_1$ takes the value which is not less than variance obtained for another standardized price **b**. Moreover,

$$v(z_1) = v(y\mathbf{b}_1) = \alpha_1^{-2}v(y\mathbf{p}_1) = \alpha_1^{-2}v(w_1)$$

where: $\alpha_1^2 = \mathbf{p}_1^T\mathbf{p}_1$, and w_1 is the variable whose values are elements of the vector $w_1=y\mathbf{p}_1$. The elements of w_1 determine the households expenditure under the worst price vector $\mathbf{p}_1 = \alpha\mathbf{b}_1$. Hence, the standard deviation $\sqrt{v(z_1)}$ can be treated as a specific coefficient of variation of the expenditure distribution because it is the ratio of the standard deviation $\sqrt{v(w_1)}$ and the length α_1 of the price vector $\mathbf{p}_1$.

[2] See e.g. Ralstona (1975).

1.2. Superpopulation model

Let us assume that the population parameter **y** is an observation of the random matrix $\boldsymbol{Y} = [Y_{ij}]$ (i=1,...,N; j=1,...,m). Hence, the space of the parameter **y** becomes the sample space of the random matrix ***Y***. Let F(**y**) be the distribution function of the random matrix ***Y***.

Definition 1.5: [Cassel et al. (1977)]**:** By a "superpopulation model", or simply a "model", we shall mean a specified set of conditions that define a class of distribution functions to which F(**y**) is assumed to belong.

The expected value, variance and covariance of the random matrix **Y** are denoted by $E(.)$, $D^2(.)$ and $Cov(.)$, respectively and particularly:

$$E(Y_{ij}) = \mu_{ij}, \quad D^2(Y_{ij}) = E(Y_{ij} - \mu_{ij})^2 = \sigma_{ij}^2$$

$$Cov(Y_{ij}, Y_{kt}) = E(Y_{ij} - \mu_{ij})(Y_{kt} - \mu_{kt}) = \sigma_{ik,\,jt}$$

Let $\mathbf{Y} = [\mathbf{Y}_{*1}...\mathbf{Y}_{*m}]$, where $\mathbf{Y}_{*j}^T = [Y_{1j}...Y_{Nj}]$ (j=1,...,m). An outcome $\mathbf{y}_{*j}$ of the random vector $\boldsymbol{Y}_{*j}$ can be treated as an observation of the j-th variable. In the one-dimensional case the transformation superpopulation is as follows, see e.g. Cassel et al. (1977). In the case of a j-th variable, the superpopulation model is determined by the distribution function: $F_{*j}(\mathbf{y}_{*j}|\mathbf{a}_{*j},\mathbf{b}_{*j},\mu,\sigma,\rho)$ (j=1,...,m). The elements of the vectors $\mathbf{a}_{*j}=[a_{1j}...a_{Nj}]^T$ ($a_{ij}\neq 0$ for each pair i,j) and $\mathbf{b}_{*j}=[b_{1j}...b_{Nj}]^T$ are parameters of the following transformation:

$$U_{ij} = \frac{Y_{ij} - b_{ij}}{a_{ij}}, \; i = 1,...,N \tag{1.12}$$

The parameters of the elements of the vector $\mathbf{U}_{*j} = [U_{1j}...U_{Nj}]^T$ are given by the expressions:

$$E(U_{ij}) = \mu, \; D^2(U_{ij}) = \sigma^2, \; Cov(U_{ij}, U_{kj}) = \rho\sigma^2 \tag{1.13}$$

where:

$$-(N-1)^{-1} \leq \rho \leq 1 \tag{1.14}$$

This and the expression (1.12) lead to the following one:

$$\begin{cases} E(Y_{ij}) = a_{ij}\mu + b_{ij} \\ D^2(Y_{ij}) = a_{ij}^2\sigma^2 \\ Cov(Y_{ij}, Y_{kj}) = a_{ij}a_{kj}\rho\sigma^2 \end{cases} \tag{1.15}$$

The generalization of this model on the m-dimensional superpopulation model denoted by G_T is as follows: let $F(\mathbf{y} \mid \mathbf{a},\mathbf{b},\mu,\sigma,\rho)$ be the distribution function of the random matrix $\boldsymbol{Y}$, where: $\mathbf{a}=[\mathbf{a}_{*1}...\mathbf{a}_{*m}]$ and $\mathbf{b} = [\mathbf{b}_{*1}...\mathbf{b}_{*m}]$. The elements of the random matrix $\boldsymbol{U} = [U_{ij}]$ (i=1,...,N; j=1,...,m) are determined by the equation (1.12). The parameters of the elements of the matrix $\boldsymbol{U}$ are determined by the expression (1.13) and the following one for j, t=1,...,m, i≠k=1,...,N:

$$\begin{cases} Cov(U_{ij}, U_{it}) = \rho_{jt}\sigma^2 \\ Cov(U_{ij}, U_{kt}) = \rho\sigma^2 \quad \text{dla} \quad i \neq k \quad \text{oraz} \quad j \neq t \end{cases} \tag{1.16}$$

The parameters of the random matrix $\boldsymbol{Y}$ are shown by the expression (1.15) and for j,t=1,...,m, i≠k=1,...,N, by the following one:

$$\begin{cases} Cov(Y_{ij}, Y_{it}) = a_{ij}a_{it}\rho_{jt}\sigma^2 \\ Cov(Y_{ij}, Y_{kt}) = a_{ij}a_{kt}\rho\sigma^2 \quad \text{dla} \quad i \neq k, \quad j \neq t \end{cases} \tag{1.17}$$

Let $\mathbf{R}=[\rho_{ij}]$ be the correlation matrix for the random vectors $\mathbf{U}_{i*}=[U_{i1}...U_{im}]$, where: i=1,...,N. Let $\mathbf{P} = \rho\mathbf{J}_m\mathbf{J}_m^T$ and $\mathbf{U}_*=[\mathbf{U}_{1*} \ ... \ \mathbf{U}_{N*}]^T$. The expressions (1.13) and (1.16) allows us to infer that the correlation matrix $\mathbf{R}_*$ of the vector $\mathbf{U}_*$ is of and degree Nm and is determined by the following equation:

$$\mathbf{R}_* = \begin{bmatrix} \mathbf{R} & \mathbf{P} & ... & \mathbf{P} \\ \mathbf{P} & \mathbf{R} & ... & \mathbf{P} \\ . & . & ... & . \\ \mathbf{P} & \mathbf{P} & ... & \mathbf{R} \end{bmatrix}$$

Theorem 1.4 [Wywiał (1992)]: If $\det(\mathbf{R}_*)>0$, the coefficient of correlation ρ fulfils the following inequalities:

$$-\frac{1}{(N-1)\mathbf{J}_m^T\mathbf{R}\mathbf{J}_m} \leq \rho \leq \frac{1}{\mathbf{J}_m^T\mathbf{R}\mathbf{J}_m} \tag{1.18}$$

and particularly for m=1 these inequality are reduced to this given by the expression (1.14).

Proof: The determinant of the matrix can be written in the following way (see e.g. Rao (1982)):

$$\det(\mathbf{R}_*) = \det^{N-1}(\mathbf{R}-\mathbf{P})\det(\mathbf{R}+(N-1)\mathbf{P})$$

The characteristic equation $\det(\mathbf{R}_* - \kappa\mathbf{I}_m) = 0$ can be rewritten as follows:

$$\det^{N-1}(\mathbf{R}-\mathbf{P}-\kappa\mathbf{I}_m)\det(\mathbf{R}+(N-1)\mathbf{P}-\kappa\mathbf{I}_m) = 0$$

On the basis of the equation $\det\left(\mathbf{A}+\mathbf{b}\mathbf{b}^{\mathrm{T}}\right)=\left(1+\mathbf{b}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{b}\right)\det(\mathbf{A})$ we have:

$$\det(\mathbf{R}+(\mathrm{N}-1)\mathbf{P})=\det\left(\mathbf{R}+(\mathrm{N}-1)\rho\mathbf{J}_{\mathrm{m}}\mathbf{J}_{\mathrm{m}}^{\mathrm{T}}\right)>0$$

$$\left(1+(\mathrm{N}-1)\rho\mathbf{J}_{\mathrm{m}}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{J}_{\mathrm{m}}\right)\det(\mathrm{R})\geq 0$$

This leads to the first inequality of the expression (1.18). Under the assumption that $\rho\geq 0$ we have:

$$\det(\mathbf{R}-\mathbf{P})=\det\left(\mathbf{R}+(i\sqrt{\rho}\mathbf{J}_{\mathrm{m}})(i\sqrt{\rho}\mathbf{J}_{\mathrm{m}}^{\mathrm{T}})\right)>0$$

where $i^2=-1$,

$$\left(1-\rho\mathbf{J}_{\mathrm{m}}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{J}_{\mathrm{m}}\right)\det(\mathbf{R})>0$$

This result leads to the second inequality in the expression (1.18). The assumption that $\rho<0$ leads to the same result.

Let us notice that the particular case of the transformation model allows us to consider the stratified superpopulation. Moreover, if we assume that for each i=1,...,N and j=1,...,m $b_{ij}=0$, and $a_{ij}=1$, and $\rho=0$, the superpopulation model becomes the simple statistical sample which is well known in classical statistical inference.

Let us assume that the matrix $\mathbf{X}=[x_{it}]$ (i=1,...N: t=1,...p) which consists of observations of a p-dimensional auxiliary variable: $x=[x_1...x_p]$ in a population, is available. We introduce the following notation: $\mathbf{X}^{\mathrm{T}}=\left[\mathbf{x}_{1*}^{\mathrm{T}}\ldots\mathbf{x}_{\mathrm{N}*}^{\mathrm{T}}\right]$ where $\mathbf{x}_{i*}=[x_{i1}...x_{ip}]$ and $\mathbf{X}=[\mathbf{x}_{*1}...\mathbf{x}_{*p}]$, where $\mathbf{x}_{*t}^{\mathrm{T}}=[x_{1t}\ldots x_{\mathrm{N}t}]$. If we assume that $\mu=0$, then, on the basis of the expressions 1.12-1.15, we have:

$$E(\mathrm{Y}_{ij})=\mathrm{b}_{ij}=\mathbf{x}_{i*}\boldsymbol{\beta}_{\mathrm{j}} \tag{1.19}$$

for each i=1,...N; j=1,...,m, where $\boldsymbol{\beta}_{\mathrm{j}}=[\beta_{1\mathrm{j}}...\beta_{\mathrm{pj}}]^{\mathrm{T}}$ is the vector of parameters of the linear regression. This regression explains the scatter of the j-the variable $\mathbf{Y}_{*\mathrm{j}}$ by means of the auxiliary variables. This definition leads to the following results

$$E(\mathbf{Y})=\mathbf{b}=\mathbf{X}\boldsymbol{\beta} \tag{1.20}$$

where: $\boldsymbol{\beta}=[\boldsymbol{\beta}_1...\boldsymbol{\beta}_{\mathrm{m}}]$.

Let $\mathbf{U}=\mathbf{Y}-E(\mathbf{Y})$ be the residual matrix. Hence:

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{U} \tag{1.21}$$

This model is called regression (superpopulation) model, see e.g. Royall (1970). The central moments of the second order of the elements of the matrix **Y** are determined by the expressions (1.15) and (1.17). Let us note that they depend on factors a_{ij} (i=1,...,m; j=1,...,N) which are usually determined by functions of auxiliary variables.

1.3. Sampling designs and schemes

Let us introduce the following definition, see e.g. Cassel et al (1977).

Definition 1.6: A sequence $\underline{s}=\{k_1,...,k_n\}$, such that $k_i \in \Omega$, for i=1,...,n is called an ordered sample of fixed size n. The set $\underline{\mathbf{S}}=\underline{\mathbf{S}}(\Omega)$ of all sequence $\underline{s}$ is denoted by $\underline{\mathbf{S}}$.

Definition 1.7: The number of a distinct elements of a sequence $\underline{s}$ is denoted by $v \le n$ and called the effective sample size.

Omitting the repetitions in the sequence $\underline{s}$, we can define a set s corresponding to $\underline{s}$:

$$s = \{k : k \in \underline{s}\} \tag{1.22}$$

Definition 1.8: A nonempty set s such that $s \subseteq \Omega$ is called an unordered sample. The number of elements of s is the effective sample size. The set of all sets s is denoted by **S**.

Definition 1.9: The function $P(\underline{s})$ on $\underline{\mathbf{S}}$**,** satisfying the following properties

$$P(\underline{s}) \ge 0 \text{ for all } s \in \underline{\mathbf{S}} \text{ and } \sum_{\underline{s} \in \underline{\mathbf{S}}} P(\underline{s}) = 1, \tag{1.23}$$

is called an ordered sampling design.

Definition 1.10: The function P(s) on **S,** satisfying the following properties:

$$P(s) \ge 0 \text{ for all } s \in \mathbf{S} \text{ and } \sum_{s \in \mathbf{S}} P(s) = 1, \tag{1.24}$$

is called an unordered sampling design.

Let us introduce the following set:

$$A(k_1,...,k_r) = \{s: k_i \in s, \text{ dla } i=1,...,r\}$$

Definition 1.11: The probability of selecting the fixed units $k_1,...,k_r$ to a sample s is called inclusion probability of order r and denoted by $\pi_{k_1...k_r}$. It is determined by the expression:

$$\pi_{k_1...k_r} = \sum_{s \in A(k_1,...,k_r)} P(s) \tag{1.25}$$

Similarly, we can define the inclusion probability in a sequence case.

The sampling design of the simple and ordered sample P_1 is as follows:

$$\bigwedge_{\underline{s}\in\underline{\mathbf{S}}} \quad P_1(\underline{s}) = N^{-n} \tag{1.26}$$

$P_1(s)$ is the sampling design of the simple sample drawn with replacement.

Let $\underline{\mathbf{S}}_* \subset \underline{\mathbf{S}}$. Let $\underline{s} \in \underline{\mathbf{S}}_*$, if and only if the size of a sample $\underline{s}$ is the effective sample size. Hence, $\underline{\mathbf{S}}_*$ consists of all ordered samples without any repetitions.

The sampling design of the simple and ordered sample with a fixed effective size P_2 is as follows:

$$\bigwedge_{\underline{s}\in\underline{\mathbf{S}}_*} \quad P_2(\underline{s}) = \frac{(N-n)!}{N!} \tag{1.27}$$

The sampling design of the simple and unordered sample is determined by the expression:

$$\bigwedge_{s\in\mathbf{S}} P_3(s) = \frac{1}{\binom{N}{n}} \tag{1.28}$$

The inclusion probabilities of the first order for the sampling designs P_2 and P_3 are as follows:

$$\pi_i^{(2)} = \pi_i^{(3)} = \frac{n}{N}, \quad \text{for } i=1,...,N \tag{1.29}$$

and in the case of the sampling design P_1:

$$\pi_i^{(1)} = 1-(1-N^{-1})^n, \text{ for } i=1,...,N \tag{1.30}$$

The inclusion probabilities of the second order for the sampling designs P_2 and P_3 are determined by the expression:

$$\pi_{ij}^{(2)} = \pi_{ij}^{(3)} = \frac{n(n-1)}{N(N-1)}, \text{ for } i\neq j=1,...,N \tag{1.31}$$

Let us note that Czerniak (1971) derived the inclusion probabilities of the second order for the sampling design P_1. The inclusion probabilities of the order $r<n$ for the sampling designs P_2 and P_3 were derived by Herzel (1986).

The sampling design is a draw-by draw mechanism for selecting units such that there is a predetermined set of selection probabilities for each unit in each draw. A sampling scheme is said to implement a given design $P(\underline{s})$ or $P(s)$ if the draw-by-draw mechanism reproduces the probabilities $P(\underline{s})$ or $P(s)$.

The set of probabilities implementing a sampling design is defined in the following way. Let $p(k_1)$ be the probability of selecting a fixed population element k_1 to a sample. The conditional probability of selecting the fixed population element k_1 to a sample, provided that the elements: $k_{i-1},...,k_1$ have just been selected to this sample, is denoted by

$$p(k_i|k_{i-1},...,k_1)=\frac{p(k_i,k_{i-1},...,k_1)}{p(k_{i-1},...,k_1)} \tag{1.32}$$

where $k_i=1,...,N$ for $i=1,...,n$. The defined probabilities fulfill the equation:

$$p(k_1)\prod_{i=2}^{n}p(k_i|k_{i-1},...,k_1)=P(\underline{s})$$

where: $k_i=1,...,N$ and $i=1,...,n$.

The inclusion probability can be determined on the basis of the following formula:

$$\pi_{k_1,...,k_r}=\frac{n!}{(n-r)!}p(k_1,...,k_r) \tag{1.33}$$

Particularly, in the case of the sampling designs P_1:

$$p(k_i|k_{i-1},...,k_1)=p(k_i)=N^{-1} \text{ for } i=1,...,n \text{ and } k_i=1,...,N$$

In the case of the designs P_2 or P_3:

$$p(k_i)=N^{-1} \text{ for } k_1=1,...,N$$

$$p(k_i|k_{i-1},...,k_1)=\frac{1}{N-i+1} \text{ for } i=2,...,n \text{ and } k_i=1,...,N$$

Theorem 1.5 [T.V.H. Rao (1962)]: For any given design $P(\underline{s})$ there exists at least one sampling design that implements $P(\underline{s})$.

1.4. Data and statistics

Let us remind you that a population parameter is the matrix **y** of dimensions N×m. Each row of this matrix is a function of an element of a population. Each column of the matrix **y** is treated as a set of observations of a character (variable) of the population. All variables can be treated as an m-dimensional one.

After observation of values attached to elements $k \in \underline{s}$, we obtain the outcome of the m-dimensional variable denoted as the following vector: $\mathbf{y}_{k*}=[y_{k1}...y_{km}]$. Let $\mathbf{y}_{\underline{S}}$ be the matrix of dimensions n×m. It consists of the rows: $\mathbf{y}_{k_i*}=[y_{k_i i}...y_{k_i m}]$, where i=1,...,n and $\underline{s}=\{k_1,...,k_n\}$. Hence:

$$\mathbf{y}_{\underline{s}} = \begin{bmatrix} y_{k_1*} \\ .... \\ y_{k_m*} \end{bmatrix}$$

The following definitions are introduced as generalizing the appropriate definitions by Cassel et al (1977) on a multidimensional case

Definition 1.12: The pair: $\underline{\mathbf{D}} = (\underline{S}, \mathbf{y}_{\underline{s}})$ or $\mathbf{D} = (S, \mathbf{y}_S)$ are the data of a multidimensional variable obtained through the observations of its outcomes in a sequence $\underline{s}$ or s, respectively.

The unlabeled data are denoted by $\mathbf{y}_{\underline{S}}$ and $\mathbf{y}_S$ in the sequence and set cases, respectively. They can be obtained from labeled data $\underline{\mathbf{D}}$ and **D**.

The data can be treated as outcomes of a multidimensional random variable. The sample spaces of the random variable $\underline{\boldsymbol{D}}$ and ***D,*** taking values $\underline{\mathbf{D}}$ and **D,** are as follows:

$$\underline{\boldsymbol{D}}=\{\underline{\mathbf{D}}\mathbf{:}\ \underline{s}\in\underline{\mathbf{S}},\ \mathbf{Y}\in\boldsymbol{\Upsilon}\},\qquad \boldsymbol{D}=\{\mathbf{D:}\ s\in\mathbf{S},\ \mathbf{Y}\in\boldsymbol{\Upsilon}\}$$

where $\boldsymbol{\Upsilon}\subseteq R^{nm}$ is the space of a population parameter.

Definition 1.13: Statistic $Z=u(\underline{\mathbf{D}})$ is a function on such $\underline{\mathbf{D}}$ that, for any given $\underline{s}\in\underline{\mathbf{S}}$, u(.), depends on **Y** only through those $\mathbf{y}_{k*}$ for which $k\in\underline{s}$.

The statistic in the set case is defined similarly.

1.5. Strategies

1.5.1. The fixed population approach

The estimation problem of a vector of parametric function: $\boldsymbol{\theta}=[\theta_1...\theta_m]\in\boldsymbol{\Theta}$ is being considered. It is estimated on the basis of the data $\mathbf{d} = (s, \mathbf{y}_s)$ which is an outcome of the random matrix $\mathbf{D} = (S, \mathbf{y}_S)$ determined by the definition 1.12. The probability distribution of the matrix **D** depends on the population parameter **y** and on the sampling design.

A vector of statistics $\mathbf{t}_D=[t_{D1}...t_{Dm}]$ will be called an estimator of describing parameters $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, if $\mathbf{t}_D$ takes values from the space $\boldsymbol{\Theta}$. Similarly, the estimator $\mathbf{t}_{\underline{\mathbf{D}}}$ as a function of the data $\underline{\mathbf{D}}$ from an ordered sample $\underline{S}$ can be defined. The symbols $\mathbf{t_D}$, $\mathbf{t}_{\underline{\mathbf{D}}}$ can be replaced by $\mathbf{t}_S$, $\mathbf{t}_{\underline{S}}$. The estimator $\mathbf{t}_S$ is a random variable because it depends on the random sample S while $\mathbf{t}_s$ is a value of the statistic $\mathbf{t}_S$. The value $\mathbf{t}_s$ is determined by the data observed in an outcome s of the random sample S.

For example the vector of population averages $\bar{\mathbf{y}}=[\bar{y}_1...\bar{y}_m]$ can be estimated by means of the vector of sample averages: $\bar{\mathbf{y}}_S=[\bar{y}_{S1}...\bar{y}_{Sm}]$, where:

$$\bar{y}_{iS}=\frac{1}{n}\sum_{k\in S}y_{ki}$$

Hence, an estimate of a parameter vector $\boldsymbol{\theta}$ depends on the estimator and sampling design. These two elements determine a sampling strategy.

Definition 1.14: [Cassel et al. (1977)]**:** An ordered pair of an estimator and a sampling design ($\mathbf{t}_S$, P(s)) or $\left(\mathbf{t}_{\underline{S}}, P(\underline{s})\right)$ is called the sampling strategy of a parameter vector $\boldsymbol{\theta}$.

For example let $P(s)=P_3(s)$, where the sampling design $P_3(s)$ is determined by the equation (1.28). In this case the sampling strategy $(\bar{\mathbf{y}}_S, P_3(s))$ is called the vector of simple sample means.

The vector of estimation errors is as follows:

$$\mathbf{B}=\mathbf{B}(\mathbf{t}_s,\boldsymbol{\theta})=\mathbf{t}_S-\boldsymbol{\theta} \tag{1.34}$$

Definition 1.15: The pair $\{\mathbf{t}_S,P(s)\}$ is an unbiased strategy of the parameter vector $\boldsymbol{\theta}=\boldsymbol{\theta}(\mathbf{Y})$ if and only if:

$$\bigwedge_{\mathbf{Y}\in\boldsymbol{\Upsilon}} E(\mathbf{t}_S)=\sum_{s\in\mathbf{S}}\mathbf{t}_s P(s)=\boldsymbol{\theta} \tag{1.35}$$

Definition 1.16: The pair $(\mathbf{t}_S,P(s))$ is a consistent strategy of the parameter vector $\boldsymbol{\theta}=\boldsymbol{\theta}(\mathbf{Y})$ if and only if $\mathbf{t}_U=\boldsymbol{\theta}(\mathbf{y})$ for each $\mathbf{y}\in\boldsymbol{\Upsilon}$.

The vector of simple sample means $(\bar{\mathbf{y}}_S, P_3(s))$ is an unbiased and consistent sampling strategy for a vector of population means $\bar{\mathbf{y}}=[\bar{y}_1...\bar{y}_m]$.

Let $\mathbf{I}_m$ be the unit matrix of degree m and let $\mathbf{o}_m$ be the column vector of all its m elements equal to zero. Let $\mathbf{A}$ be a matrix of dimensions m×m, let $\mathbf{b}$ be a row vector of dimensions 1×m and let $\mathbf{J}_N$ be a column vector of dimensions N×1, each element of the vector $\mathbf{J}_N$ being equal to one. Let $\mathbf{D}$ and $\mathbf{D}'$ be data dependent on population parameters $\mathbf{y}$ and $\mathbf{y}'$, respectively. Moreover, let $\mathbf{y}'=\mathbf{yA}+\mathbf{J}_N\mathbf{b}$.

Definition 1.17: A vector estimator $\mathbf{t}_D$ is linear invariant if and only if $\mathbf{y'} = \mathbf{yA} + \mathbf{J}_n\mathbf{b}$ and for each outcomes **d'** and **d** of data **D'** and **D**, respectively, it is[3]:

$$\mathbf{t}_{\mathbf{d}'} = \mathbf{b} + \mathbf{t}_\mathbf{d}\mathbf{A} \tag{1.36}$$

The vector estimator $\mathbf{t}_D$ is origin invariant if $\mathbf{A}=\mathbf{I}_m$ and $\mathbf{b}\neq\mathbf{o}_m$. The vector estimator $\mathbf{t}_D$ is scale invariant if the matrix **A** is diagonal with positive diagonal elements and $\mathbf{b}=\mathbf{o}_m$.

The matrix of the second mixed moments of estimation errors is denoted by $\mathbf{V}_{SR}(\mathbf{t}_S)=\mathbf{V}_{SR}(\mathbf{t}_S)=E(\mathbf{B}^T\mathbf{B})=[E(B_iB_j)]$, where:

$$E(B_iB_k) = \sum_{s\in S}(t_{is} - \theta_i)(t_{ks} - \theta_k)P(s) \tag{1.37}$$

We say that $\mathbf{V}_{SR}(\mathbf{t}_S)$ is the matrix of mean square errors.

The covariance matrix of an estimator $\mathbf{t}_S$ is:

$$\mathbf{V}(\mathbf{t}_S)=E(\mathbf{t}_S-E(\mathbf{t}_S))^T(t_S-E(\mathbf{t}_S)).$$

This matrix can be decomposed in the following way:

$$\mathbf{V}_{SR}(\mathbf{t}_S) = \mathbf{V}(\mathbf{t}_S) + E(\mathbf{B}^T)E(\mathbf{B}) \tag{1.38}$$

When a vector $\mathbf{t}_S$ is an unbiased estimator of a vector $\boldsymbol{\theta}$, $\mathbf{V}_{SR}(\mathbf{t}_S) = \mathbf{V}(\mathbf{t}_S)$.

For example the covariance matrix of the sampling strategy $(\bar{\mathbf{y}}_S, P_3(s))$ is:

$$\mathbf{V}(\bar{\mathbf{y}}_S, P_3) = \frac{N-n}{Nn}\mathbf{C}_* \tag{1.39}$$

where: $\mathbf{C}_*=[c_*(y_i,y_j)]$ and:

$$c_*(y_i, y_j) = \frac{1}{N-1}\sum_{k=1}^{N}(y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j)$$

$$v_*(y_i) = c_*(y_i, y_i) = \frac{1}{N-1}\sum_{k=1}^{N}(y_{ik} - \bar{y}_i)^2 .$$

An unbiased estimator of the variance $\mathbf{V}(\bar{\mathbf{y}}_S, P_3)$ is determined by the equation:

$$\mathbf{V}_S(\bar{\mathbf{y}}_S, P_3) = \frac{N-n}{Nn}\mathbf{C}_{*S} \tag{1.40}$$

[3] See, e.g. Cassel et al. (1977), p. 78, where one-dimensional case is considered.

where: $\mathbf{C}_{*S}=[c_{*S}(y_i,y_j)]$ and

$$c_{*S}(y_i, y_j) = \frac{1}{n-1}\sum_{k\in S}(y_{ik} - \bar{y}_{iS})(y_{jk} - \bar{y}_{jS}),\ \bar{y}_{iS} = \frac{1}{n}\sum_{k\in S} y_{ik} \tag{1.41}$$

$$v_{*S}(y_i) = c_{*S}(y_i, y_i) = \frac{1}{n-1}\sum_{k\in S}(y_{ik} - \bar{y}_{iS})^2$$

Usually, accuracy of the vector estimator $\mathbf{t}_S$ is characterized by means of the vector $[E(B_1)^2...E(B_m)^2]$. Its elements are the mean square errors of the appropriate elements of the vector $\mathbf{t}_S=[t_{1S}...t_{mS}]$. The variances of these estimators are represented by the vector $[D^2(t_{1S})...D^2(t_{mS})]$.

The next method of assessing the accuracy of vector estimation is based on the trace of the matrix of mean square errors which is defined by the equation:

$$q^2_{SR}(\mathbf{t}_S) = \mathrm{tr}\mathbf{V}_{SR}(\mathbf{t}_S) \tag{1.42}$$

The parameter $q_{SR}(\mathbf{t}_S)$ can be interpreted as a mean distance between the point whose coordinates are determined by elements of the vector $\boldsymbol{\theta}$ and the points whose coordinates are assigned by outcomes of a vector estimator $\mathbf{t}_S$.

The coefficient

$$q(\mathbf{t}_S)=\sqrt{\mathrm{tr}\mathbf{V}(\mathbf{t}_S)} \tag{1.43}$$

can be treated as a mean distance between the vector $E(\mathbf{t}_S)$ and the outcomes of the estimator $\mathbf{t}_S$. It can be called the mean radius of the estimator $\mathbf{t}_S$.

The expression (1.38) leads to the conclusion: if $E(\mathbf{t}_S)=\boldsymbol{\theta}$, $q_{SR}(\mathbf{t}_S)=q(\mathbf{t}_S)$.

For example the expressions (1.39) and (1.41) lead to the following one:

$$q(\bar{\mathbf{y}}_S) = \sqrt{\frac{N-n}{Nn}\sum_{i=1}^{m} v_*(y_i)} \tag{1.44}$$

Wilks (1932) introduced the generalized variance as the measure of a multivariate scatter[4]. It is the following determinant of a variance-covariance matrix:

$$g(\mathbf{t}_S) = \det\mathbf{V}(\mathbf{t}_S) \tag{1.45}$$

[4] Wywial (1996a, 1997) defined a more general coefficient of multivariate scatter than the generalized variance.

Similarly, we define the generalized mean square error as the determinant:

$$g_{SR}(\mathbf{t}_S) = \det\mathbf{V}_{SR}(\mathbf{t}_S). \tag{1.46}$$

On the basis of Wilks' (1962, p. 546) result we have:

$$g_{SR}(\mathbf{t}_S) = g(\mathbf{t}_S)\,(1 + E(\mathbf{B})\mathbf{V}^{-1}(\mathbf{t}_S)E^T(\mathbf{B})) \geq g(\mathbf{t}_S) \tag{1.47}$$

If $\mathbf{t}_S$ is an unbiased estimator of a vector $\boldsymbol{\theta}$, $g_{SR}(\mathbf{t}_S)=g(\mathbf{t}_S)$.

The spectral radius[5] of a matrix is equal to its maximal eigenvalue. Let $\rho_{SR}(\mathbf{t}_S)$ and $\rho(\mathbf{t}_S)$ be the spectral radius of the matrices $\mathbf{V}_{SR}(\mathbf{t}_S)$ and $\mathbf{V}(\mathbf{t}_S)$, respectively. Let us define the following statistic: $t_{oS}=\mathbf{t}_S\mathbf{w}^T$, where $\mathbf{w}=[w_1...w_m]$ and $\mathbf{w}\mathbf{w}^T=1$ and $\mathbf{w}\in R^z-\{\mathbf{o}_m\}$. The statistic t_{oS} estimates the linear combination: $\theta_o=\boldsymbol{\theta}\mathbf{w}^T$. The mean square error of the statistic t_{oS} is: $v_{SR}(t_{oS})=E(t_{oS}-\theta_o)^2=\mathbf{w}\mathbf{V}_{SR}(\mathbf{t}_S)\mathbf{w}^T$. The variance of the estimator t_{oS} is: $D^2(t_{oS}) = \mathbf{w}\mathbf{V}(\mathbf{t}_S)\mathbf{w}^T$. The well known properties of the maximal eigenvalue of a non-negative definite matrix lead to the following result:

$$\rho_{SR}(\mathbf{t}_S) = \max_{\mathbf{w}\mathbf{w}^T=1} \operatorname{imum}\left\{\mathbf{w}\mathbf{V}_{SR}(\mathbf{t}_S)\mathbf{w}^T\right\} \tag{1.48}$$

$$\rho(\mathbf{t}_S) = \max_{\mathbf{w}\mathbf{w}^T=1} \operatorname{imum}\left\{\mathbf{w}\mathbf{V}(\mathbf{t}_S)\mathbf{w}^T\right\} \tag{1.49}$$

Then, $\rho_{SR}(\mathbf{t}_S)$ is a mean square error of estimation of a linear combination $\theta_o=\boldsymbol{\theta}\mathbf{w}^T$ by means of the statistic $t_{oS}=\mathbf{t}_S\mathbf{w}^T$ under the worst vector $\mathbf{w}_*$ of coefficients of the linear combination $\boldsymbol{\theta}\mathbf{w}^T$ in this sense that $\rho_{SR}(\mathbf{t}_S)=E(\mathbf{t}_S\mathbf{w}_*^T-\boldsymbol{\theta}\mathbf{w}_*^T)^2 \geq E(\mathbf{t}_S\mathbf{w}^T-\boldsymbol{\theta}\mathbf{w}^T)^2$, for such $\mathbf{w}$ that: $\mathbf{w}\mathbf{w}^T=1$ and $\mathbf{w}_*\mathbf{w}_*^T=1$, where: $\mathbf{w},\mathbf{w}_*\in R^m-\{\mathbf{o}_m\}$. The parameter $\rho(\mathbf{t}_S)$ is similarly explained. Moreover, $\rho(\mathbf{t}_S)=\rho_{SR}(\mathbf{t}_S)$, when $\mathbf{t}_S$ is an unbiased estimator of the vector $\boldsymbol{\theta}$.

Let us consider the linear combination $\boldsymbol{\theta}_\#=\boldsymbol{\theta}\boldsymbol{\alpha}^T$, where $\boldsymbol{\alpha}\in R^m-\{0\}$. Then, the linear combination $\boldsymbol{\theta}\boldsymbol{\alpha}^T$ is unrestricted in comparison to the linear combination $\boldsymbol{\theta}\mathbf{w}^T$. Let $t_{\#S}=\mathbf{t}_S\boldsymbol{\alpha}^T$ be an estimator of the parameter $\boldsymbol{\theta}_\#$. Its mean square error and variance are as follows: $v_{SR}(t_{\#S})=E(t_{\#S}-\boldsymbol{\theta}_\#)^2=\boldsymbol{\alpha}^T\mathbf{V}_{SR}(\mathbf{t}_S)\boldsymbol{\alpha}$, $D^2(t_{\#S})=\boldsymbol{\alpha}^T\mathbf{V}(\mathbf{t}_S)\boldsymbol{\alpha}$, respectively. Let us assume that the vector $\boldsymbol{\alpha}$ is not known during the estimation. In this case the accuracy of the estimation can be assessed on the basis of the maximal values of the following coefficients:

$$u_{SR}(t_{\#S},\boldsymbol{\alpha}) = \frac{v_{SR}(t_{\#S})}{\boldsymbol{\alpha}\boldsymbol{\alpha}^T} = \mathbf{w}\mathbf{V}_{SR}(\mathbf{t}_S)\mathbf{w}^T \tag{1.50}$$

[5] See e.g. Ralston (1975).

$$u(t_{\#S}, \boldsymbol{\alpha}) = \frac{D^2(t_{\#S})}{\boldsymbol{\alpha}\boldsymbol{\alpha}^T} = \mathbf{w}\mathbf{V}(\mathbf{t}_S)\mathbf{w}^T \tag{1.51}$$

where:

$$\mathbf{w} = \frac{\boldsymbol{\alpha}}{\sqrt{\boldsymbol{\alpha}\boldsymbol{\alpha}^T}}$$

Then, $\mathbf{w}\mathbf{w}^T=1$ and

$$\max_{\boldsymbol{\alpha}\in R^m-\{0\}} \operatorname{imum}\{u_{SR}(t_{\#S}, \boldsymbol{\alpha})\} = \rho_{SR}(\mathbf{t}_S) \tag{1.59}$$

$$\max_{\boldsymbol{\alpha}\in R^m-\{0\}} \operatorname{imum}\{u(t_{\#S}, \boldsymbol{\alpha})\} = \rho(\mathbf{t}_S) \tag{1.60}$$

We should find such a strategy of estimation of a vector $\boldsymbol{\theta}$ which is unbiased and whose variance (or mean square error) takes the minimal value in some set of strategies.

Let $(\mathbf{t}_S^{(1)}, P^{(1)}(s)), (\mathbf{t}_S^{(2)}, P^{(2)}(s))$ be two sampling strategies of estimation of a vector $\boldsymbol{\theta}=[\theta_1...\theta_m]$.

Definition[6] **1.18:** A sampling strategy $(\mathbf{t}_S^{(1)}, P^{(1)}(s))$ for estimation of a vector $\boldsymbol{\theta}=\boldsymbol{\theta}(\mathbf{y})$ is not worse than the strategy $(\mathbf{t}_S^{(2)}, P^{(2)}(s))$ if and only if

$$\bigwedge_{\boldsymbol{\alpha}\in R^m-\{0\}} \bigwedge_{\mathbf{Y}\in\mathbf{Y}} v_{SR}(\mathbf{t}_S^{(1)}\boldsymbol{\alpha}^T, P^{(1)}(s)) \le v_{SR}(\mathbf{t}_S^{(2)}\boldsymbol{\alpha}^T, P^{(2)}(s)) \tag{1.52}$$

where:

$$v_{SR}(\mathbf{t}_S^{(k)}, P^{(k)}(s)) = E(\mathbf{t}_S^{(k)}\boldsymbol{\alpha}^T - \boldsymbol{\theta}\boldsymbol{\alpha}^T), \text{ for } k=1,2$$

The sampling strategy $(\mathbf{t}_S^{(1)}, P^{(1)}(s))$ for estimation of a vector $\boldsymbol{\theta}=\boldsymbol{\theta}(\mathbf{y})$ is better than the strategy $(\mathbf{t}_S^{(2)}, P^{(2)}(s))$ if and only if the inequality (1.52) is fulfilled and for at least one fixed parameter $\mathbf{y}$ the inequality (1.52) is sharp.

Hence, for a fixed $\mathbf{y}$, we can say that the mean square scatter of the random vector $\mathbf{t}_S^{(1)}$ around the point whose coordinates are determined by the vector $\boldsymbol{\theta}(\mathbf{y})$ is not bigger than the scatter of the random vector $\mathbf{t}_S^{(2)}$ around the same point, if the inequality (1.52) is fulfilled.

[6] It is the adaptation of the definitions of Borovkov (1984) and Cassel et al. (1977).

Theorem[7] **1.6:** A sampling strategy $(\mathbf{t}_S^{(1)}, P^{(1)}(s))$ for the estimation of a vector $\boldsymbol{\theta}(\mathbf{y})$ is not worse (not better) than a sampling strategy $(\mathbf{t}_S^{(2)}, P^{(2)}(s))$ if and only if for each $\mathbf{y}\in\mathbf{Y}$ the matrix: $\mathbf{R}=\mathbf{V}_{SR}(\mathbf{t}_S^{(2)}, P^{(2)}(s))-\mathbf{V}_{SR}(\mathbf{t}_S^{(1)}, P^{(1)}(s))$ is non-negative (non-positive) definite. When for at least one parameter $\mathbf{y}\in\mathbf{Y}$ the matrix $\mathbf{R}$ is positive (negative) definite, then the sampling strategy $(\mathbf{t}_S^{(1)}, P^{(1)}(s))$ is better (worse) than the strategy $(\mathbf{t}_S^{(2)}, P^{(2)}(s))$.

For example, let us compare the strategies: $(\bar{\mathbf{y}}_S, P_1)$ and $(\bar{\mathbf{y}}_S, P_3)$, where the first (second) strategy is called the vector of the mean value from a simple sample drawn with (without) replacement. It is well known that: $\mathbf{V}(\bar{\mathbf{y}}_S, P_1)=\frac{N-1}{Nn}\mathbf{C}_*$. This and the expression (1.39) lead to the following one:

$$\mathbf{R}_0=\mathbf{V}(\bar{\mathbf{y}}_S, P_1)-\mathbf{V}(\bar{\mathbf{y}}_S, P_3)=\frac{n-1}{Nn}\mathbf{C}_*$$

The matrix $\mathbf{R}_0$ is positive definite if $\det\mathbf{C}_*>0$. Hence, the strategy $(\bar{\mathbf{y}}_S, P_3)$ is better than $(\bar{\mathbf{y}}_S, P_1)$.

Theorem 1.7: [Rao (1982)]: Let the matrices: $\mathbf{V}_{SR}(\mathbf{t}_S^{(2)}, P^{(2)}(s))$ and $\mathbf{V}_{SR}(\mathbf{t}_S^{(1)}, P^{(1)}(s))$ be positive definite. If a sampling strategy $(\mathbf{t}_S^{(1)}, P^{(1)}(s))$ for the estimation of a vector $\boldsymbol{\theta}(\mathbf{y})$ is not worse than the strategy $(\mathbf{t}_S^{(2)}, P^{(2)}(s))$ in the sense of definition 1.6, then:

$$q_{SR}\left(\mathbf{t}_S^{(1)}, P^{(1)}(s)\right)\le q_{SR}\left(\mathbf{t}_S^{(2)}, P^{(2)}(s)\right) \tag{1.52}$$

$$g_{SR}\left(\mathbf{t}_S^{(1)}, P^{(1)}(s)\right)\le g_{SR}\left(\mathbf{t}_S^{(2)}, P^{(2)}(s)\right) \tag{1.53}$$

$$\boldsymbol{\rho}_{SR}\left(\mathbf{t}_S^{(1)}, P^{(1)}(s)\right)\le \boldsymbol{\rho}_{SR}\left(\mathbf{t}_S^{(2)}, P^{(2)}(s)\right) \tag{1.54}$$

$$\bigwedge_{i=1,\ldots,m} v_{SR}\left(t_{Si}^{(1)}, P^{(1)}(s)\right)\le v_{SR}\left(t_{Si}^{(2)}, P^{(2)}(s)\right) \tag{1.55}$$

Proof. Rao (1982), p. 89, showed: if $\mathbf{B}$ is positive definite and $(\mathbf{A}-\mathbf{B})$ is non-negative definite then $\det(\mathbf{A})\ge\det(\mathbf{B})$. This and the theorem 1.6 lead to the expression (1.53). The well known properties of the trace of a sum of matrices lead to the expression (1.52). Let $\mathbf{A}=\mathbf{V}_{SR}(\mathbf{t}_S^{(2)}, P^{(2)}(s))$ and $\mathbf{B}=\mathbf{V}_{SR}(\mathbf{t}_S^{(1)}, P^{(1)}(s))$. The maximal eigenvalue of a matrices $\mathbf{A}$ and $\mathbf{B}$ are denoted by $\lambda_1(\mathbf{A})$ and $\lambda_1(\mathbf{B})$, respectively. Hence:

[7] See Borovkov (1984), Rao (1982).

$$\lambda_1(\mathbf{A}) = \max_{\boldsymbol{\alpha}^T\boldsymbol{\alpha}=1}\{\boldsymbol{\alpha}^T\mathbf{A}\boldsymbol{\alpha}\}$$

$$\lambda_1(\mathbf{B}) = \max_{\boldsymbol{\beta}^T\boldsymbol{\beta}=1}\{\boldsymbol{\beta}^T\mathbf{B}\boldsymbol{\beta}\}$$

If $(\mathbf{A}-\mathbf{B})$ is non-negative definite then for all non-zero vectors $\boldsymbol{\gamma}$:

$$\gamma^T\mathbf{A}\gamma - \gamma^T\mathbf{B}\gamma \geq 0 \tag{1.56}$$

Hence:

$$\lambda_1(\mathbf{A}) \geq \boldsymbol{\gamma}^T\mathbf{A}\boldsymbol{\gamma} \geq \boldsymbol{\gamma}^T\mathbf{B}\boldsymbol{\gamma} \geq 0\,,$$

so

$$\lambda_1(\mathbf{A}) \geq \lambda_1(\mathbf{B})$$

This leads to to the expression (1.54). The inequality (1.56) lets us derive the expression (1.55).

Kish (1965) defined the following coefficientt:

$$\mathrm{deff}(t_S, P(s)) = \frac{v_{SR}(t_S, P(s))}{D^2(\bar{y}_S, P_1(s))}$$

where $P_1(s)$ is given by the expression (1.26), $D^2(\bar{y}_S, P_1(s)) = \frac{v(y)}{n}$ is the variance of the mean from the simple sample of size n drawn with replacement, $v_{SR}(t_S, P(s))$ is the mean square error of a strategy $(t_S, P(s))$ determined on the basis of a sample of size n. The coefficient $\mathrm{deff}(t_S, P(s))$ is called as sampling effect[8] and it measures accuracy of a strategy $(t_S, P(s))$ in relation to the mean from the simple sample drawn with replacement. Particularly, $\mathrm{deff}(\bar{y}_S, P_2(s)) = \frac{N-n}{N-1}$, where $P_2(s)$ is defined by the expression (1.27) and $(\bar{y}_S, P_2(s))$ is the mean from the simple sample drawn without replacement.

Rao and Scott (1981) generalized the deff coefficient into the multivariate case in the following way:

$$\mathrm{deff}(\mathbf{t}_S, P(s)) = \rho\left(\mathbf{V}_{SR}(\mathbf{t}_S, P(s))\mathbf{V}^{-1}(\bar{\mathbf{y}}_S, P_1(s))\right) \tag{1.57}$$

where $\rho(.)$ is the maximal eigenvalue of a matrix. The sizes of the samples of the sampling designs are fixed and the same level. The equation (1.57) is equivalent to the following:

[8] Bracha (1996) sugested that $\mathrm{deff}(t_S, P(s))$ should be rather called as strategy effect.

$$\operatorname{deff}(\mathbf{t}_S, P(s)) = \sup_{\mathbf{c}\in R^m-\{0\}}\left(\frac{v_{SR}\left(\mathbf{c}^T\mathbf{t}_S, P(s)\right)}{D^2\left(\mathbf{c}^T\bar{\mathbf{y}}_S, P_1(s)\right)}\right) = \sup_{\mathbf{c}\in R^m-\{0\}}\left(\frac{\mathbf{c}^T\mathbf{V}_{SR}\left(\mathbf{t}_S, P(s)\right)\mathbf{c}}{\mathbf{c}^T\mathbf{V}\left(\bar{\mathbf{y}}_S, P_1(s)\right)\mathbf{c}}\right)$$

where $\mathbf{c} \in R^m-\{0\}$ is a column vector. Hence, there exists such $\mathbf{c}_0$ that $\operatorname{deff}(\mathbf{t}_S, P(s))$ takes a maximal value of the ratio of the mean square error of the strategy $\left(\mathbf{c}_0{}^T\mathbf{t}_S, P(s)\right)$ and the variance of the strategy $\left(\mathbf{c}_0{}^T\bar{\mathbf{y}}_S, P_1(s)\right)$. Particularly, $\operatorname{deff}(\bar{y}_S, P_2(s)) = \dfrac{N-n}{N-1}$.

Cramèr (1958) proposed to compare the scatter of multidimensional random variables on the basis of the concentration ellipsoid. In our case the concentration ellipsoid of a vector $\mathbf{t}_S$ is defined by the following expression:

$$K(\mathbf{t}_S) = \left\{\mathbf{z}: \ [\mathbf{z} - E(\mathbf{t}_S)]\mathbf{V}^{-1}[\mathbf{z} - E(\mathbf{t}_S)]^T \le m+2, \ \ \mathbf{z}\in R^m\right\} \tag{1.58}$$

Let us assume that strategies $(\mathbf{t}_S^{(i)}, P^{(i)}(s))$ (i=1,2) are unbiased and that $\det\mathbf{V}\left(\mathbf{t}_S^{(i)}\right) > 0$. It is easy to prove that if the strategy $(\mathbf{t}_S^{(1)}, P^{(1)}(s))$ is not worse than the strategy $(\mathbf{t}_S^{(2)}, P^{(2)}(s))$, $K(\mathbf{t}_S^{(1)}) \subseteq K(\mathbf{t}_S^{(2)})$.

Let us note the problem of construction of the confidences' sets for estimation of the parameter vector $\boldsymbol{\theta}$. We limit our consideration to the confidence set in the shape of an ellipsoid. Let $\mathbf{t}_S$ be an unbiased estimator of the vector $\boldsymbol{\theta}$. An unbiased estimator of covariance matrix $\mathbf{V}(\mathbf{t}_S)$ will be denoted by $\tilde{\mathbf{V}}(\mathbf{t}_S)$. Hence, the ellipsoid confidence set for the vector $\boldsymbol{\theta}$ is determined by the expression:

$$P\{Q_S < q_\gamma \mid \boldsymbol{\theta}\} \ge \gamma$$

where:

$$Q_S = (\mathbf{t}_S - \boldsymbol{\theta})\ \tilde{\mathbf{V}}^{-1}(\mathbf{t}_S)(\mathbf{t}_S - \boldsymbol{\theta})^T \tag{1.59}$$

Under some assumptions connected with a limit distribution of the estimator vector $\mathbf{t}_S$, the limit distribution of the statistic Q_S is of the χ^2_m with m degree of freedom.

The precision of the estimation can be assessed by means of the volume H of the ellipsoid[9]:

$$H = cq_\gamma^{m/2}g^{m/2}(\mathbf{t}_S) \tag{1.60}$$

[9] See, e.g. Cramer (1958).

where:

$$c = \frac{\pi^{m/2}}{\Gamma(0.5m+1)}$$

Hence, the volume H depends on the generalized variance $g(\mathbf{t}_S)=\det\mathbf{V}(\mathbf{t}_S)$.

For example, let us consider the well known unbiased strategy $(\bar{\mathbf{y}}_S, P_3(s))$, called the mean from a simple sample drawn without replacement. The unbiased estimator of its variance-covariance matrix is determined by the expression (1.40). In this case the statistic Q_S takes the following form:

$$Q_S = (\bar{\mathbf{y}}_S - \bar{\mathbf{y}})\,\mathbf{V}_s^{-1}(\bar{\mathbf{y}}_S, P_3)(\bar{\mathbf{y}}_S - \bar{\mathbf{y}})^T \tag{1.61}$$

here the matrix $\mathbf{V}_S(\bar{\mathbf{y}}_S, P_3)$ is defined by the expression (1.40).

The precision of the estimation can be assessed by means of the volume of the ellipsoid $Q_S<q_\gamma$. This volume is determined by the expression (1.60), where $g(\mathbf{t}_S)=g(\bar{\mathbf{y}}_S, P_3)=\det \mathbf{V}_S(\bar{\mathbf{y}}_S, P_3)$. Hence, on the basis of the expression (1.39) we have:

$$g(\bar{\mathbf{y}}_S, P_3)= \left(\frac{N-n}{(N-1)n}\right)^m g \tag{1.62}$$

where:

$$g = \left(\frac{N-1}{N}\right)^m \det \mathbf{C}_*$$

Mikhail and Mir (1981) proposed the following unbiased estimator of the generalized variance[10] g.

$$\tilde{g}_S = \frac{\binom{N}{n}}{\binom{N-m-1}{n-m-1}} \frac{n(n-1)^m}{N^{m+1}} g_S \tag{1.63}$$

where:

$$g_S = \det \mathbf{C}_{*S}$$

[10] See Wywial (1996, 1997).

is a biased estimator of a population generalized variance g. Hence, when in the expression (1.62) we substitute the generalized variance g for its estimator $\tilde{g}_S$, we obtain the unbiased estimator of the generalized variance of the strategy $(\bar{y}_S, P_3)$.

1.5.2. The superpopulation approach

According to the definition 1.5 of the superpopulation model the population parameter **y** can be treated as an outcome of a random matrix **Y** of dimensions N×m. Properties of probability distribution of the matrix **Y** determine the superpopulation model. The parametric function $\boldsymbol{\theta}=\boldsymbol{\theta}(\mathbf{y})$, determined by the definition 1.2, should be treated as an outcome of the random vector $\boldsymbol{\Theta}=\boldsymbol{\theta}(\boldsymbol{Y})$ of dimensions 1×m.

Let us assume that the size of a sample is fixed and equal to n. Let us remember that the matrix $[\mathbf{y}_{k_1}...\mathbf{y}_{k_n}]$ is a sub-matrix of the matrix **y**. Moreover, outcomes $[\mathbf{y}_{k_1}...\mathbf{y}_{k_n}]$ are treated as outcomes of random variables $[\mathbf{Y}_{k_1}...\mathbf{Y}_{k_n}]$ observed in a sample $\underline{s}=\{k_1,...,k_n\}$. Let $\mathbf{D}=((k,\mathbf{y}_k), k\in s)$ be the data from unordered sample s. The data **D** can be treated as an outcome of the random variable $\mathbf{D}=((k,\mathbf{Y}_k), k\in s)$.

Let us assume that noninformative sampling designs will be considered. It means that a draw of a sample s does not depend on the distribution of random variables $\mathbf{Y}_{k_1}...\mathbf{Y}_{k_n}$, where $k_i\in s$, see, e.g.: Cassel et al (1977). Hence:

$$P(\mathbf{D}=\mathbf{D}) = P(s)P(\mathbf{Y}_k=\mathbf{y}_k, \text{ for } k\in s) \tag{1.64}$$

The value $\boldsymbol{\theta}(\mathbf{y})$ of a random variable $\boldsymbol{\Theta}=\boldsymbol{\theta}(\boldsymbol{Y})$ is predicted by means of a statistic $\mathbf{t}(\mathbf{D})$. This statistic will be denoted by $\mathbf{T}_S$ and called the predictor of $\boldsymbol{\Theta}$. An outcome $\mathbf{T}_s$ of the predictor $\mathbf{T}_S$ is obtained for a fixed sample s. If $[\mathbf{Y}_{k_1}=\mathbf{y}_{k_1}...\mathbf{Y}_{k_n}=\mathbf{y}_{k_n}]$, an outcome of $\mathbf{T}_S$ is denoted by $\mathbf{t}_S$. Finally, if a sample s is fixed and $[\mathbf{Y}_{k_1}=\mathbf{y}_{k_1}...\mathbf{Y}_{k_n}=\mathbf{y}_{k_n}]$ then an outcome of $\mathbf{T}_S$ will be denoted by $\mathbf{t}_s$.

Definition 1.19[11]: The statistic $\mathbf{T}_S$ is called a p-unbiased predictor of $\boldsymbol{\Theta}$ if and only if for a given design p and for $\boldsymbol{Y}\in\boldsymbol{\mathcal{Y}}$

$$E(\mathbf{T}_S)=\boldsymbol{\theta}(\mathbf{Y}) \tag{1.65}$$

The strategy $(\mathbf{T}_S, p)$ will be called p-unbiased if $\mathbf{T}_S$ is a p-unbiased predictor of $\boldsymbol{\theta}(\mathbf{Y})$.

[11] See Cassel et al. (1977), p. 92.

The statistic $\mathbf{T}_S$ is called a ξ-unbiased predictor of $\boldsymbol{\Theta}$ if and only if for a given probability distribution ξ

$$E(\mathbf{T}_S - \boldsymbol{\Theta}) = \mathbf{0} \tag{1.66}$$

$\mathbf{T}_S$ is called a p-ξ-unbiased predictor of $\boldsymbol{\Theta}$ if and only if for a given p and ξ

$$E\mathrm{E}(\mathbf{T}_S - \boldsymbol{\Theta}) = \mathbf{0} \tag{1.67}$$

Precision of the strategy ($\mathbf{T}_S$, p) will be determined by the expected value of the matrix of mixed second moments of prediction errors:

$$E[\mathbf{V}_{SR}(\mathbf{T}_S)] = E\mathrm{E}(\mathbf{T}_S - \boldsymbol{\Theta})^T(\mathbf{T}_S - \boldsymbol{\Theta}) \tag{1.68}$$

Let $\boldsymbol{V}(\mathbf{T}_s)=E(\mathbf{T}_s-E(\mathbf{T}_s)]^T[(\mathbf{T}_s-E(\mathbf{T}_s)]$ be the ξ-covariance matrix of prediction errors and let $\boldsymbol{B}(\mathbf{T}_s)= E(\mathbf{T}_S -\boldsymbol{\Theta})$ be the ξ-bias of a strategy $\mathbf{T}_s$. Moreover, let $\boldsymbol{V}(\boldsymbol{\Theta})=E(\boldsymbol{\Theta}-E(\boldsymbol{\Theta})]^T[(\boldsymbol{\Theta}-E(\boldsymbol{\Theta})]$.

Theorem[12] **1.8**: If a sampling design p of a strategy ($\mathbf{T}_S$, p) is noninformative, then

$$E\mathbf{V}_{SR}(\mathbf{T}_S) = \mathrm{E}\boldsymbol{V}(\mathbf{T}_S) + \mathrm{E}\boldsymbol{B}^2(\mathbf{T}_S) + \boldsymbol{V}(\boldsymbol{\Theta}) - 2E\{\boldsymbol{\Theta} - E(\boldsymbol{\Theta})\}\mathrm{E}\{\mathrm{T}_S - E(\boldsymbol{\Theta})\} \tag{1.69}$$

Particularly:

a) If $\mathbf{T}_S$ is p-unbiased:

$$E\mathbf{V}(\mathbf{T}_S) = E\boldsymbol{\mathbf{V}}(\mathbf{T}_S) + \mathrm{E}\boldsymbol{B}^2(\mathbf{T}_S) - \boldsymbol{V}(\boldsymbol{\Theta}) \tag{1.70}$$

b) If $\mathbf{T}_S$ is p-ξ-unbiased:

$$E\mathbf{V}(\mathbf{T}_S) = E\mathbf{V}(\mathbf{T}_S) - \boldsymbol{V}(\boldsymbol{\Theta}) \tag{1.71}$$

The synthetic coefficients of accuracy of a vector prediction are as follows: the mean square radius of a strategy ($\mathbf{T}_S$,p) is determined by the equation:

$$q_{SR}(\mathbf{T}_S) = \sqrt{\mathrm{tr}E\mathbf{V}_{SR}(\mathbf{T}_S)} \tag{1.72}$$

The generalized mean square error of prediction is defined by the expression:

$$g_{SR}(\mathbf{T}_S)=\det E\mathbf{V}_{SR}(\mathbf{T}_S) \tag{1.73}$$

[12] This theorem is the generalization of the theorem by Cassel et all (1977), pp.94, on multidimensional case.

The mean square spectral radius of a strategy $(\mathbf{T}_S,p)$ is denoted by $\rho_{SR}(\mathbf{T}_S)$ and it is equal to the maximal eigenvalue value of the mean square errors matrix $E\mathbf{V}_{SR}(\mathbf{T}_S)$.

Let M be a class of superpopulation models. Let $\mathbf{B}_{S1}=\mathbf{T}_{S1}-\Theta$ and $\mathbf{B}_{S2}=\mathbf{T}_{S2}-\Theta$ be prediction errors. The following definition is similar to the definition 1.18.

Definition[13] **1.20:** $\{\mathbf{T}_1, P_1\}$ is not a worse strategy than $(\mathbf{T}_2,P_2)$ if and only if, for each probability distribution $\xi\in M$

$$\bigwedge_{\mathbf{e}\in R^z-\{\mathbf{o}_z\}} E\mathrm{E}(\mathbf{B}_{S1}\mathbf{e}^T)^2 \le E\mathrm{E}(\mathbf{B}_{S2}\mathbf{e}^T)^2 \tag{1.74}$$

or

$$\bigwedge_{\mathbf{e}\in R^z-\{\mathbf{o}_z\}} \mathbf{e}E\mathbf{V}_{SR}(\mathbf{T}_{S1},P_1)\mathbf{e}^T \le \mathbf{e}E\mathbf{V}_{SR}(\mathbf{T}_{S2}P_2)\mathbf{e}^T \tag{1.75}$$

If there exist such $\mathbf{e}\in R^m$ and such $\xi\in M$ that the inequalities (1.74) and (1.75) are sharp then $(\mathbf{T}_{S1}, P_1)$ is a better strategy than $(\mathbf{T}_{S2}, P_2)$.

Theorem[14] **1.9:** If for each $\xi\in M$ the matrix $\mathbf{L}=E\mathbf{V}_{SR}(\mathbf{T}_{S2},P_2)-E\mathbf{V}_{SR}(\mathbf{T}_{S1},P_1)$ is not negative definite, $(\mathbf{T}_{S1},P_1)$ is not worse than the strategy $(\mathbf{T}_{S2},P_2)$.

Theorem[15] **1.10:** If $\det E\mathbf{V}_{SR}(\mathbf{T}_S,P_1)>0$, $(\mathbf{T}_{S1},P_1)$ is not a worse predictor of a vector Θ than a strategy $(\mathbf{T}_{S2},P_2)$ in the sense of the definition 1.20 then for each $\xi\in M$

$$q_{SR}(\mathbf{T}_{S1},P_1) \le q_{SR}(\mathbf{T}_{S2},P_2) \tag{1.76}$$

$$g_{SR}(\mathbf{T}_{S1},P_1) \le g_{SR}(\mathbf{T}_{S2},P_2) \tag{1.77}$$

$$\rho_{SR}(\mathbf{T}_{S1},P_1) \le \rho_{SR}(\mathbf{T}_{S2},P_2) \tag{1.78}$$

[13] This definition is an adaptation of ones considered by Borovkov (1984) and Cssel et al. (1977), p. 93.

[14] See Borovkov (1984), Rao (1982).

[15] See Rao (1982).

II. VECTOR OF HORVITZ-THOMPSON ESTIMATOR

2.1. Basic definitions

Let us assume that a sample s of a fixed size n is drawn without replacement from a finite population. If $k \in s$, $a_k=1$ and if $k \notin s$, $a_k=0$. This leads to the following properties:

$$\pi_k = E(a_k), \qquad \pi_{ki} = E(a_k a_i), \qquad \sum_{k=1}^{N} \pi_k = n,$$

$$D^2(a_k) = \pi_k(1-\pi_k), \qquad Cov(a_k, a_i) = \pi_{ki} - \pi_k \pi_i.$$

Horvitz and Thompson (1952) proposed the following estimator of population average $\bar{y}$:

$$t_{HTS} = \frac{1}{N} \sum_{k=1}^{N} \frac{a_k y_k}{\pi_k} \tag{2.1}$$

The statistic t_{HTS} is the unbiased estimator of the population mean $\bar{y}$. Its variance is determined by the expression:

$$D^2(t_{HTS}) = \frac{1}{N^2} \sum_{k=1}^{N} \left(\frac{y_k}{\pi_k} \right)^2 \pi_k(1-\pi_k) + \frac{1}{N^2} \sum_{k \neq i}^{N} \sum_{=1}^{N} \frac{y_k y_i}{\pi_k \pi_i} (\pi_{ki} - \pi_i \pi_k) \tag{2.2}$$

For the effective sample size:

$$D^2(t_{HTS}) = \frac{1}{N^2} \sum_{k<i}^{N} \sum_{=1}^{N} (\pi_k \pi_i - \pi_{ki}) \left(\frac{y_k}{\pi_k} - \frac{y_i}{\pi_i} \right)^2 \tag{2.3}$$

The variance $D^2(t_{HTS})$ is estimated by means of the following statistic:

$$\overline{D}^2(t_{HTS}) = \frac{1}{N^2} \sum_{k=1}^{N} a_k \left(\frac{y_k}{\pi_k} \right)^2 (1-\pi_k) + \frac{1}{N^2} \sum_{k \neq i=1}^{N} \sum^{N} a_k a_i \frac{y_k y_i}{\pi_k \pi_i} \frac{\pi_{ki} - \pi_k \pi_i}{\pi_{ki}} \tag{2.4}$$

This statistic is the unbiased estimator of the variance $D^2(t_{HTS})$ but it can take negative values.

Sen (1953), Yates and Grundy (1953) derived the following non-negative valued estimator, in the case when $\pi_k \pi_i - \pi_{ki} > 0$ for each $k \neq i = 1,...,N$:

$$\tilde{D}^2(t_{HTS}) = \frac{1}{N^2} \sum_{k>i}^{N} \sum_{i=1}^{N} a_k a_i \frac{y_k y_i}{\pi_k \pi_i} \frac{\pi_k \pi_i - \pi_{ki}}{\pi_{ki}} \left(\frac{y_k}{\pi_k} - \frac{y_i}{\pi_i} \right)^2 \tag{2.5}$$

Let us note that Horvitz and Thompson proposed the estimator of a population average in the case when a sample is selected with replacement. Properties of the Horvitz-Thompson estimator were studied e.g. by: Godambe and Joshi (1965), Joshi (1965, 1966), Ramakrishnan (1975), Sarndal (1976), Hartley and Rao (1968, 1969).

Let us assume that a sample is selected without replacement from a fixed and finite population. The vector of Horvitz-Thompson statistics will be denoted by $\mathbf{t}_{HTS}=[t_{HT1S}...t_{HTmS}]$, where:

$$t_{HTiS} = \frac{1}{N} \sum_{k=1}^{m} \frac{a_k y_{ik}}{\pi_k}, \qquad i = 1,\ldots,m.$$

It is the estimator of the vector of population averages denoted by $\bar{\mathbf{y}} = [\bar{y}_1 \cdots \bar{y}_m]$. The vector $\mathbf{t}_{HTiS}$ is the unbiased estimator of the vector $\bar{\mathbf{y}}$.

The covariances of elements of the vector $\mathbf{t}_{HTiS}$ are as follows (see Thompson (1997)):

$$\mathrm{Cov}(t_{HTiS}, t_{HTjS}) = N^{-2} \sum_{k=1}^{N} \frac{y_{ik} y_{jk}}{\pi_k} (1 - \pi_k) +$$

$$+ 2N^{-2} \sum_{k>h=1}^{N} \sum^{N} \frac{y_{ik} y_{jh}}{\pi_k \pi_h} (\pi_{kh} - \pi_k \pi_h) \tag{2.6}$$

or:

$$\mathrm{Cov}(t_{HTiS}, t_{HTjS}) = N^{-2} \sum_{k>h=1}^{N} \sum^{N} \left(\frac{y_{ik}}{\pi_k} - \frac{y_{ih}}{\pi_h} \right) \left(\frac{y_{jk}}{\pi_k} - \frac{y_{jh}}{\pi_h} \right) (\pi_k \pi_h - \pi_{kh}) \tag{2.7}$$

Let $\mathbf{G}_\pi$ be a diagonal matrix of degree N with the following diagonal elements: π_k, k=1,..,N. Moreover, let $\mathbf{a}=[a_1...a_N]$ and $\boldsymbol{\pi}= [\pi_1...\pi_N]$. Hence:

$$t_{HTS} = \frac{1}{N} \mathbf{a} \mathbf{G}_\pi^{-1} \mathbf{y} \tag{2.8}$$

This leads to the following expressions:

$$\mathbf{t}_{HTS} - \bar{\mathbf{y}} = \left[\frac{1}{N} \sum_{k=1}^{N} \frac{y_{ik}}{\pi_k} (a_k - \pi_k) \right] = \frac{1}{N} \boldsymbol{\pi} (\mathbf{a}\text{-}\boldsymbol{\pi}) \mathbf{G}_{\pi}^{-1} \mathbf{y}$$

$$\mathbf{V}(\mathbf{t}_{HTS}) = N^{-2} \mathbf{y}^{T} \mathbf{G}_{\pi}^{-1} \mathbf{V}(a) \mathbf{G}_{\pi}^{-1} \mathbf{y} \tag{2.9}$$

where: $\mathbf{V(a)}=[\mathrm{Cov}(a_i a_j)]$ is a variance-covariance matrix of the vector **a**, $\mathrm{Cov}(a_i a_j)=\pi_{ij}-\pi_i\pi_j$, $D^2(a_i)= \mathrm{Cov}(a_i a_i)\pi_i(1-\pi_i)$, $i \neq j=1,...,N$.

If for each $k \neq h=1,..,N$, the inequality $\pi_{kh}>0$ is fulfilled, the unbiased estimator of an element of the matrix $\mathbf{V}(\mathbf{t}_{HTS})$ is the following statistic [see e.g. Konijn (1973), p. 235]:

$$\overline{\mathrm{C}}\mathrm{ov}(t_{HTiS}, t_{HTjS}) = N^{-2} \sum_{k=1}^{N} \frac{y_{ik} y_{jk}}{\pi_k^2} (1 - \pi_k) a_k +$$

$$+ 2N^{-2} \sum_{k>h=1}^{N} \sum^{N} \frac{y_{ik} y_{jk}}{\pi_k \pi_h} \frac{(\pi_{kh} - \pi_k \pi_h)}{\pi_{kh}} a_k a_h \tag{2.10}$$

for i,j=1,...,m.

Next estimator is as follows (see the estimator of the variance by Yates and Grundy):

$$\overline{\mathrm{C}}\mathrm{ov}(t_{HTiS}, t_{HTjS}) = \sum_{k>h=1}^{N} \sum^{N} \frac{(\pi_{kh} - \pi_k \pi_h)}{\pi_{kh}} \left(\frac{y_{ik}}{\pi_k} - \frac{y_{ih}}{\pi_h} \right) \left(\frac{y_{jk}}{\pi_k} - \frac{y_{jh}}{\pi_h} \right) a_k a_h \tag{2.11}$$

2.2. Taylor's approximation of variance

A variance of the estimator $\bar{y}_{HT,\,S}$ can be expressed in the following way:

$$D^2(\bar{y}_{HT,S}) = \frac{1}{N^2} \left(\sum_{i=1}^{N} \frac{y_i^2}{\pi_i} + \sum_{i=1}^{N} \sum_{i \neq j=1}^{N} \frac{\pi_{ij} y_i y_j}{\pi_i \pi_j} \right) - \bar{y}^2 \tag{2.12}$$

Let us treat this variance as a function of the vector of inclusion probabilities denoted by $\boldsymbol{\pi}=[\pi_1...\pi_N\ \pi_{12}...\pi_{1N}\ \pi_{23}...\pi_{2N}...\pi_{N-1,N}]$. So $u(\boldsymbol{\pi}) = D^2(\bar{y}_{HT,S})$. A derivative of this function is denoted as follows:

$$\frac{\partial^{r+t+z} u}{\partial \pi_i^r \partial \pi_j^t \partial \pi_{ij}^z} = H_{i,j}^{r+t+z}(\pi)$$

When $\pi_i = \frac{n}{N}$ and $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$, at the point $\boldsymbol{\pi}_o = \left[\frac{n}{N}\ \frac{n}{N} \ldots \frac{n(n-1)}{N(N-1)}\right]$, the derivatives are as follows:

$$H_i^{r+0+0}(\boldsymbol{\pi}_0) = (-1)^r \frac{N^{r-1} r!}{n^{r+1}(N-1)}\left[(N-2n+1)y_i^2 + 2N(n-1)\bar{y}y_i\right],\ r \geq 1$$

$$H_{i,j}^{r+t+0}(\boldsymbol{\pi}_0) = 2(-1)^{r+t} \frac{(n-1)N^{r+t-1} r!\, t!}{(N-1)n^{r+t+1}} y_i y_j,\ \ r, t \geq 1$$

$$H_{i,j}^{0+0+1}(\boldsymbol{\pi}_0) = \frac{1}{n^2} y_i y_j$$

$$H_{i,j}^{r+t+1}(\boldsymbol{\pi}_0) = 2(-1)^{r+t} \frac{N^{r+t} r!\, t!}{n^{r+t+2}} y_i y_j,\quad r \geq 0,\ t \geq 1$$

Particularly:

$$H_i^{1+0+0}(\boldsymbol{\pi}_0) = \frac{-1}{n^2(N-1)}\left[(N-2n+1)y_i^2 + 2N(n-1)\bar{y}\ y_i\right] \tag{2.13}$$

$$H_i^{2+0+0}(\boldsymbol{\pi}_0) = \frac{2N}{n^3(N-1)}\left[(N-2n+1)y_i^2 + 2N(n-1)\bar{y}\ y_i\right] \tag{2.14}$$

$$H_{i,j}^{1+1+0}(\boldsymbol{\pi}_0) = \frac{2(n-1)N}{n^3(N-1)} y_i y_j \tag{2.15}$$

$$H_{i,j}^{1,0,1}(\boldsymbol{\pi}_0) = \frac{-2N}{n^3} y_i y_j \tag{2.16}$$

Let

$$\Delta_i = \pi_i - \frac{n}{N},\ \ \Delta_{ij} = \pi_{ij} - \frac{n(n-1)}{N(N-1)} \tag{2.17}$$

The Taylor's series expansion of the variance $D^2\left(\bar{y}_{HT,S}\right) = u(\pi)$ is as follows.

$$u(\pi) = u(\pi_0) + \sum_{i=1}^{N} H_i^{1+0+0}(\pi_0)\Delta_i + \sum_{i=1}^{N}\sum_{i\neq j=1}^{N} H_{i,j}^{0+0+1}(\pi_0)\Delta_{ij} +$$

$$+ \frac{1}{2}\sum_{i=1}^{N} H_i^{2+0+0}(\pi_0)\Delta_i^2 + \frac{1}{2}\sum_{i=1}^{N}\sum_{i\neq j=1}^{N} H_{i,j}^{1+1+0}(\pi_0)\Delta_i\Delta_j +$$

$$+ \frac{1}{2}\sum_{i=1}^{N}\sum_{i\neq j=1}^{N} H_{i,j}^{1+0+1}(\pi_0)\Delta_i\Delta_{ij} + R(\theta, N, n) \qquad (2.18)$$

where $\theta \in (0, 1)$ and $R(\theta, N, n)$ is the remainder.

Let $p_i=f_1(x_i)$ and $q_i=f_2(x_i)$, be functions of an auxiliary variable and

$$\Delta_i = \frac{\alpha}{N} p_i = O\left(N^{-1}\right) \qquad (2.19)$$

$$\Delta_{ij} = \frac{n}{N^2}\left[a\left(p_i + p_j\right) + bq_i q_j + c\right] = O\left(nN^{-2}\right) \qquad (2.20)$$

where $\alpha,a,b,c = O\left(n^k N^h\right)$ and $k,h\leq 0$.

Let

$$v_{zge} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^z p^g q^e, \quad \bar{p} = v_{010}, \bar{q} = v_{001}, \qquad v_{*zge} = \frac{N}{N-1} v_{zge} \qquad (2.21)$$

where $z,g,e\in\{0,1,2\}$.

Wywiał (2000) derived the following Taylor's expansion of the Horvitz-Thompson estimator:

$$D^2\left(\bar{y}_{HT,S}\right) = \frac{N-n}{Nn} v_{*200} + \frac{1}{n}\left[2(a-\alpha)\bar{y}(v_{110} + \bar{y}\,\bar{p}) +\right.$$

$$\left. + b(v_{101} + \bar{y}\,\bar{q})^2 + c\bar{y}^2\right] + \frac{\alpha}{n^2}\left[-\alpha v_{210} + \bar{y}^2\bar{p} +\right.$$

$$+ (\alpha - a)(v_{110} + \overline{yp})^2 + (2\alpha - a)\bar{y}(v_{120} + \bar{y}v_{020}) +$$

$$\left. - b(v_{111} + \bar{y}v_{011})(v_{101} + \overline{yq}) - c\bar{y}(v_{110} + \overline{yp})\right] + O\left(n^{-3}\right) \qquad (2.22)$$

2.3. Approximation of variances of strategies

2.3.1. Sampling design proportionate to sample average of auxiliary variable

Let $\mathbf{x}=[x_1...x_N]$ be the vector of an auxiliary variable's observations. Moreover, let each value of the auxiliary variable be positive. The sample and population means of the auxiliary variable are denoted by $\bar{x}_{\underline{s}} = \frac{1}{n}\sum_{k\in\underline{s}} x_k$ and $\bar{x} = \frac{1}{N}\sum_{k=1}^{N} x_k$, respectively. The size n of a sample is the effective sample size. Let sampling design of an unordered sample s be proportional to the sample mean of the auxiliary variable, then: $P(\underline{s}) \propto n\bar{x}_S$ and

$$P_4(s) = \frac{n\bar{x}_s}{n\sum_{s\in\mathbf{S}} \bar{x}_s}. \tag{2.23}$$

Hence, when $\bar{x}_{s_1} > \bar{x}_{s_2}$ the probability of selecting the sample s_1 is larger than the probability of selecting the sample s_2.

Lahiri (1951) considered the following sampling design:

$$P_4(s) = \frac{1}{\binom{N}{n}} \frac{\bar{x}_S}{\bar{x}} \tag{2.24}$$

The probabilities of the first and second order are as follows (see e.g. Wywiał (1991a, 1992, 1995, 2000):

$$\pi_k = \frac{N-n}{(N-1)N} \frac{x_k - \bar{x}}{\bar{x}} + \frac{n}{N} \tag{2.25}$$

$$\pi_{kt} = \frac{n(n-1)}{N(N-1)} + \frac{(n-1)(N-n)}{(N-2)(N-1)N} \frac{x_k + x_t - 2\bar{x}}{\bar{x}} \tag{2.26}$$

where $k \neq t = 1, ..., N$.

When we state that c = 0, b = 0 and

$$\alpha = \frac{N-n}{N-1} = 1 + O\left(nN^{-1}\right),\ a = \frac{(n-1)(N-n)N}{n(N-2)(N-1)} = 1 + O\left(nN^{-1}\right),\ p_k = \frac{x_k - \bar{x}}{\bar{x}}, \tag{2.27}$$

the expression (2.22) leads to the following approximation of the variance:

$$D^2\left(\bar{y}_{HT,S} P_4\right) = \frac{N-n}{Nn} v_{*2}(y) + \frac{1}{n^2}\left[2\frac{\bar{y}}{\bar{x}} v_{11}(y,x) - \frac{1}{\bar{x}} v_{21}(y,x) + \right.$$

$$\left. - \frac{v_{11}^2(y,x)}{\bar{x}^2} + \frac{\bar{y}}{\bar{x}^2} v_{12}(y,x) + \frac{\bar{y}^2}{\bar{x}^2} v_2(x)\right] + O\left(n^{-1}N^{-1}\right) + O\left(n^{-3}\right) \tag{2.28}$$

where:

$$v_{rs} = v_{rs}(y,x) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^r \left(x_i - \bar{x}\right)^s,\quad v_{*rs} = \frac{N}{N-1} v_{rs}, \tag{2.29}$$

$$v_r(y) = v_{r0}(y,x),\ v_r(x) = v_{0r}(y,x) \tag{2.30}$$

Let

$$\eta_{rs} = \eta_{rs}(y,x) = \frac{v_{rs}}{v_2^{r/2}(y) v_2^{s/2}(x)},\ \eta_{11}(y,x) = \rho \tag{2.31}$$

$$\gamma_x = \frac{\sqrt{v_2(x)}}{\bar{x}},\ \gamma_y = \frac{\sqrt{v_2(y)}}{\bar{y}},\ \kappa = \frac{\gamma_x}{\gamma_y} \tag{2.32}$$

$$D^2\left(\bar{y}_{HT,S} \middle| P_4\right) = D^2\left(\bar{y}_s, P_3\right) + \frac{v_2(y)}{n^2}\left(2\kappa\rho - \gamma_x\eta_{21} + \gamma_x\kappa\eta_{12} + \kappa^2 - \gamma_x^2\rho^2\right) +$$
$$+ O\left(n^{-1}N^{-1}\right) + O\left(n^{-3}\right) = D^2\left(\bar{y}_s, P_o\right) + O\left(n^{-2}\right) \tag{2.33}$$

where:

$$D^2\left(\bar{y}_s, P_3\right) = \frac{N-n}{Nn} v_{*2}(y) \tag{2.34}$$

is the variance of the mean from the simple sample s drawn without replacement.

Hence, when $N \to \infty$, $n \to \infty$ in such a way that $N - n \to \infty$,

$$D^2\left(\bar{y}_{HT,S}, P_4\right) \to D^2\left(\bar{y}_S, P_3\right). \tag{2.35}$$

Let us assume that (x_i, y_i) $i = 1, ..., N$ can be treated as outcomes of a two-dimensional normal random variable. In this case the expression (2.33) is reduced to the following one:

$$D^2\left(\bar{y}_{HT,S}\middle|P_4\right)=\frac{N-n}{Nn}v_{*2}(y)+\frac{v_2(y)}{n^2}\left(2\kappa\rho+\kappa^2-\gamma_x^2\rho^2\right)+$$

$$+O\left(n^{-1}N^{-1}\right)+O\left(n^{-3}\right) \tag{2.36}$$

Hence, under the stated assumption, the strategy $\left(\bar{y}_{HT,S}, P_4\right)$ is more accurate than the simple sample mean when

$$\rho<\frac{\kappa}{\gamma_x^2}\left(1-\sqrt{1+\gamma_x^2}\right) \quad \text{or} \quad \rho>\frac{\kappa}{\gamma_x^2}\left(1+\sqrt{1+\gamma_x^2}\right) \quad \text{and}$$

$$\gamma_y>\frac{1}{\gamma_x}\left(1+\sqrt{1+\gamma_x^2}\right) \tag{2.37}$$

Moreover, the size of the sample should be rather small.

2.3.2. Sampling design proportionate to total of values of auxiliary variable which are not observed in sample

Wywiał (1992,1995, 2000) considered the sampling design $P_5(\underline{s})$ proportional to the total of the non-observed values in the sample. Hence, $P_5(\underline{s}) \propto N\bar{x}-n\bar{x}_{\underline{s}}=\sum_{i\notin s}x_i=\sum_{i\in\Omega-s}x_i$. This design is as follows:

$$P_5(s)=\frac{N\bar{x}-n\bar{x}_s}{\binom{N}{n}(N-n)\bar{x}} \tag{2.38}$$

When $\sum_{i\in\Omega-s_1}x_i>\sum_{i\in\Omega-s_2}x_i$ the probability of selecting the sample s_1 is larger than the probability of selecting the sample s_2.

The inclusion probabilities of the first and second order are as follows:

$$\pi_k=\frac{n}{N}-\frac{n}{N(N-1)}p_k \tag{2.39}$$

$$\pi_{k1}=\frac{n(n-1)}{N(N-1)}-\frac{n(n-1)}{N(N-1)(N-2)}\left(p_k+p_1\right) \tag{2.40}$$

where: $p_k = \frac{x_k - \bar{x}}{\bar{x}}$, k≠l=1,...,N (2.41)

Wywiał (1992, 1995) determined the conditional probabilities of the sampling scheme implementing the sampling design P_2. Wywiał (2000) provided the following expressions:

$$\alpha = -\frac{n}{N-1} = -\frac{n}{N} + O\left(N^{-2}\right) = O\left(nN^{-1}\right);$$

$$a = -\frac{(n-1)N}{(N-1)(N-2)} = -\frac{n}{N} + O\left(N^{-2}\right) = O\left(nN^{-1}\right); \qquad b=0; \qquad c=0$$

This and the expression (2.21) lead to the approximation of the variance of the strategy:

$$D^2(t_{HTS}, P_5) = D^2(\bar{y}_S, P_3) + O\left(n^{-3}\right) \tag{2.42}$$

Let us note that in the case of two-stage sampling, P_5 can be a sampling design of the first phase of the sampling and x_k can be treated as communication cost (e.g. by bus) from the center of research to the k-th region (cluster).

2.3.3. Sampling design proportional to sample variance of auxiliary variable

The sample variance $v_{2s}(x)$ and population variance $v_2(x)$ of an auxiliary variable are denoted as follows:

$$v_{2s}(x) = \frac{1}{n}\sum_{k\in s}\left(x_k - \bar{x}\right)^2, \qquad v_2(x) = \frac{1}{N}\sum_{k\in\Omega}\left(x_k - \bar{x}\right)^2$$

Singh and Srivastava (1980) proposed the sampling design $P_6(\underline{s})$ proportional to the sample variance $v_s(x)$:

$$P_6(s) = \frac{n(N-1)}{N(n-1)\binom{N}{n}}\frac{v_{2s}(x)}{v_2(x)} \tag{2.43}$$

Singh and Srivastava (1980) proposed the sampling scheme implementing this sampling design.

Let

$$p_k = \frac{(x_k - \bar{x})^2}{v} - 1, \qquad q_k = \frac{x_k - \bar{x}}{\sqrt{v}}, \quad \text{for } k=1,...,N \tag{2.44}$$

The probabilities of inclusion are as follows (see Wywiał (1995)):

$$\pi_k = \frac{n}{N} + \frac{N-n}{N(N-2)} p_k \tag{2.45}$$

$$\pi_{kt} = 1 - \frac{N-n}{N}\left(\frac{2(n-1)}{N-2} - \frac{N-n-1}{N-3}\right) +$$

$$+ \frac{(N-n)}{N(N-2)}\left(1 - \frac{(N-1)(N-n-1)}{N(N-3)}\right)(p_k + p_t + 2) +$$

$$- \frac{2(N-n)(N-n-1)}{N^2(N-2)(N-3)} q_k q_t$$

or

$$\pi_{kt} = \frac{n(n-1)}{N(N-1)} + \frac{(N-n)}{N(N-2)}\left(1 - \frac{(N-1)(N-n-1)}{N(N-3)}\right)(p_k + p_t) +$$

$$- \frac{2(N-n)(N-n-1)}{N^2(N-2)(N-3)} q_k q_t - \frac{2(N-n)(N-n-1)}{N^2(N-1)(N-2)(N-3)} \tag{2.46}$$

for $k \neq t = 1,....,N$. Hence:

$$\alpha = \frac{N-n}{N-2} = 1 + \frac{n}{N} + O\left(nN^{-1}\right);$$

$$a = \frac{N(N-n)}{n(N-2)}\left(1 - \frac{(N-1)(N-n-1)}{N(N-3)}\right) = 1 - \frac{1}{n} - \frac{n}{N} + \frac{5}{N} + O(n^{-1}N^{-2})$$

$$b = -\frac{2(N-n)(N-n-1)}{n(N-2)(N-3)} = -\frac{2}{n} + \frac{4}{N} - \frac{8}{nN} + O\left(nN^{-2}\right);$$

$$c = -\frac{2(N-n)(N-n-1)}{n(N-1)(N-2)(N-3)} = -\frac{2}{nN} + \frac{4}{N^2} - \frac{10}{nN^2} + O\left(n^{-1}N^{-3}\right);$$

The approximate variance of the strategy (t_{HTS},P_6) is as follows:

$$D^2(t_{HTS},P_6)=D^2(\bar{y}_S,P_3)+$$

$$+\frac{\bar{y}^2}{n^2}\left(-\gamma_y^2\left(2\rho^2+\eta_{22}+1\right)+\gamma_y\left(\eta_{14}-4\eta_{12}\right)+\eta_{04}-1\right)+$$

$$O\left(n^{-1}N^{-1}\right)+O\left(n^{-3}\right) \tag{2.47}$$

where: $\eta_{rs}=\eta_{rs}(y,x)=\frac{v_{rs}}{v_2^{r/2}(y)v_2^{s/2}(x)}$, $\eta_{11}(y,x)=\rho$

Under the assumption that, $\eta_{22}=1+2\rho^2$, $\eta_{12}=0$ and $\eta_{14}=0$ (e.g. in the case when variables (y,x) have an approximately two-dimensional normal distribution) the variance takes the following form:

$$D^2(t_{HTS},P_6)=D^2(\bar{y}_S,P_3)+\frac{2\bar{y}^2}{n^2}\left(1-\gamma_y^2\left(1+2\rho^2\right)\right)+O\left(n^{-1}N^{-1}\right)+O\left(n^{-3}\right) \tag{2.48}$$

For sufficiently large n and N the strategy $D(t_{HTS},P_6)$ is better than $(\bar{y}_S,P_3)$ if $|\gamma_y|>1$ or if $1>|\gamma|>\frac{1}{\sqrt{3}}$ and $|\rho|>\sqrt{\frac{1}{2}\left(\frac{1}{\gamma_y^2}-1\right)}$.

2.3.4. Sampling design proportional to function of sample variance of auxiliary variable

Wywiał (1992, 1995, 2000) considered the sampling design $P_7(s)$ proportional to $Nv_2(x)-nv_{2s}(x)$. Hence:

$$P_7(s)=\frac{N-1}{(N-n)\binom{N}{n}}-\frac{(N-1)nv_{2s}(x)}{N(N-n)\binom{N}{n}v_2(x)} \tag{2.49}$$

or

$$P_7(s)=\frac{N-1}{(N-n)\binom{N}{n}}-\frac{N-1}{N-n}P_6(s) \tag{2.50}$$

The probabilities of inclusion are defined by the following expressions:

$$\pi_k = \frac{n}{N} - \frac{n-1}{N(N-2)} p_k \tag{2.51}$$

$$\pi_{kt} = \frac{n(n-1)(N-1)}{N-n} - \frac{n-1}{N-1} \pi_{kt}^{(6)} \tag{2.52}$$

where: $k \neq t = 1,...,N$, and $\pi_{kt}^{(6)}$ is determined by the expression (2.46) and p_k by (2.44). This inclusion probability can be transformed into the following form:

$$\pi_{kt} = \frac{n(n-1)}{N(N-n)} + \frac{n}{N^2}\left(a\left(p_k + p_t\right) + bq_k q_t + c\right) \tag{2.53}$$

where:

$$a = -\frac{(n-1)N}{n(N-2)}\left(1 - \frac{(N-1)(N-n-1)}{N(N-3)}\right) = -\frac{n}{N} + O(N^{-1}) = O(nN^{-1}) \tag{2.54}$$

$$b = \frac{2(n-1)(N-n-1)}{n(N-2)(N-3)} = \frac{2nN - 2N - 2n^2 + 2}{n(N-2)(N-3)} = \frac{2}{N} + O(nN^{-2}) = O(N^{-1}) \tag{2.55}$$

$$c = \frac{b}{N-1} = \frac{2nN - 2N - 2n^2 + 2}{n(N-1)(N-2)(N-3)} = \frac{2}{N^2} + O(nN^{-2}) = O(N^{-2}) \tag{2.56}$$

$$\alpha = -\frac{n-1}{N-2} = -\frac{n}{N} + O\left(N^{-1}\right) = O\left(nN^{-1}\right) \tag{2.57}$$

The approximate variance of the sampling strategy is:

$$D^2(t_{HTS}, P_7) = D^2(\bar{y}_S, P_3) + O\left(n^{-3}\right) \tag{2.58}$$

Hence, the variances of both strategies $(t_{HTS}, P_7), (\bar{y}_S, P_3)$ are almost the same.

2.3.5. Sampling design proportional to squared estimation error of auxiliary variable mean

Wywiał (1995, 2000) considered the properties of the following design:

$$P_8(s) = \frac{Nn}{(N-n)\binom{N}{n}} \frac{(\bar{x}_s - \bar{x})^2}{v_*} \tag{2.59}$$

The probabilities of inclusion are as follows:

$$\pi_k = \frac{n}{N} + \frac{N-2n}{N(N-2)} p_k \tag{2.60}$$

$$\pi_{kt} = \frac{n(n-1)}{N(N-1)} + \frac{(n-1)[(N-2n+1)(p_k + p_t) + 2(N-n-1)q_k q_t]}{N(N-2)(N-3)} + \\ + \frac{2(n-1)(N-n-1)}{N(N-1)(N-2)(N-3)} \tag{2.61}$$

where p_k and q_k are defined by the expression (2.44) and $k \neq t = 1,...,N$.

$$\alpha = \frac{N-2n}{N-2} = 1 - \frac{2n}{N} + O(N^{-1}) = 1 + O\left(nN^{-1}\right)$$

$$a = \frac{(n-1)N(N-2n+1)}{n(N-2)(N-3)} = 1 - \frac{1}{n} - \frac{2n}{N} + O\left(N^{-1}\right) = 1 + O\left(n^{-1}\right)$$

$$b = \frac{2(n-1)N(N-n-1)}{n(N-2)(N-3)} = 2 - \frac{2}{n} - \frac{2n}{N} + O\left(N^{-1}\right) = 2 + O\left(n^{-1}\right)$$

$$c = -\frac{2(n-1)N(N-n-1)}{n(N-1)(N-2)(N-3)} = -\frac{2}{N} - \frac{2}{nN} + O\left(nN^{-2}\right) = O\left(N^{-1}\right)$$

The approximate variance is as follows:

$$D^2(t_{HTS}, P_8) = D^2(\bar{y}_S, P_3) + \frac{2}{n} v_2(y)\left(\rho^2 + \frac{\eta_{12}}{\gamma_y}\right) +$$

$$+ \frac{1}{n^2} v_2(y)\left(1 - \eta_{22} - 2\rho\eta_{13} - \frac{1}{\gamma_y}(1 + 2\rho\eta_{03} + 2\eta_{12} - \eta_{04} - \eta_{14})\right) + O\left(n^{-3}\right) \tag{2.62}$$

This expression is more simple when we assume that the variables (y,x) have an approximately normal distribution. Hence, $\eta_{22}=1+2\rho^2$, $\eta_{13}=3\rho$ and

$$D^2(t_{HTS}, P_8) = D^2(\bar{y}_S, P_3) + \frac{2}{n} v_2(y)\rho^2 + \frac{2v_2(y)}{n^2}\left(\frac{1}{\gamma_y} - 4\rho^2\right) + O\left(n^{-3}\right) \tag{2.63}$$

2.3.6. Sampling design proportionate to decreasing function of squared estimation error of auxiliary variable mean

Arnold and Groeneveld (1981) proved the following inequality: $\left(\frac{N}{n} - 1\right) v \geq (\bar{x}_s - \bar{x})^2$ for each $s \in \mathbf{S}$, where $\mathbf{S}$ is the sample space. Wywial (2000) considered a sampling design proportional to $\left(\frac{N}{n} - 1\right) v - n(\bar{x}_s - \bar{x})^2$. It is as follows:

$$P_9(s) = \frac{N-1}{(N-2)\binom{N}{n}}\left(1 - \frac{n(\bar{x}_s - \bar{x})^2}{(N-n)v}\right) \tag{2.64}$$

or $$P_9(s) = \frac{N-1}{(N-2)\binom{N}{n}} - \frac{1}{(N-2)} P_8(s)$$

where: $P_8(s)$ is defined by the equation (2.59). Hence, when estimation error $(\bar{x}_{s_2} - \bar{x})^2$ is larger than $(\bar{x}_{s_1} - \bar{x})^2$, the probability of selecting the sample s_1 is larger than the probability of selecting the sample s_2.

The probabilities of inclusion are as follows:

$$\pi_k = \frac{n}{N} - \frac{N-2n}{N(N-2)^2} p_k \tag{2.65}$$

$$\pi_{kt} = \frac{n(n-1)}{N(N-2)} - \frac{1}{(N-2)} \pi_{kt}^{(8)} \tag{2.66}$$

or

$$\pi_{kt} = \frac{n(n-1)}{N(N-2)} - \frac{(n-1)\left((N-2n+1)(p_k + p_t) + 2(N-n-1)q_k q_t\right)}{N(N-2)^2(N-3)} + \\ - \frac{2(N-n-1)}{N(N-1)(N-2)^2(N-3)}$$

where: p_k, q_k and $\pi_{kt}^{(8)}$ are defined by the right sides of the expressions (2.61) and (2.61), respectively. After appropriate transformations we have:

$$\alpha = -\frac{N-2n}{(N-2)^2} = \frac{1}{N} - \frac{2n}{N^2} + O(N^{-2}) = O(N^{-1})$$

$$a = -\frac{(n-1)N(N-2n+1)}{n(N-2)^2(N-3)} = O\left(N^{-1}\right)$$

$$b = -\frac{2(n-1)N(N-n-1)}{n(N-2)^2(N-3)} = -\frac{2}{N} + O\left(n^{-1}N^{-1}\right) = O\left(N^{-1}\right)$$

$$c = -\frac{2N(N-n-1)}{n(N-1)(N-2)^2(N-3)} = -\frac{2n}{nN^2} + \frac{2}{N^3} + O(n^{-1}N^{-3})$$

The approximate variance is as follows:

$$D^2(t_{HTS}, P_9) = D^2(\bar{y}_S, P_3) + O\left(n^{-3}\right) \tag{2.69}$$

2.3.7. SIMULATION ANALYSIS OF ESTIMATION PRECISION

Let us assume that the variables (*x*,*y*) have an approximately two-dimensional normal distribution. The variable *y* is treated as a variable under study and the variable *x* as an auxiliary variable. The pseudovalues of a normal two-dimensional random variable have been obtained by means

of Hastings' (1955) generator[1]. The simulation procedures were written down by means of the matrix language of the SPSS statistical package[2]. We considered sample sizes of 2, 5 and 10 elements. The population size was 100 elements. The relative efficiency was determined as a ratio of the variance of Horvitz-Thompson strategy to the variance of the mean from the simple sample drawn without replacement. The variance of the Horvitz-Thompson strategy was determined on the basis of 1000 replications of the set {x,y} of the fixed size.

The computer simulations lead to the conclusion that the simple sample mean is a better estimator of population averages than the strategies (t_{HTS}, P_7) and (t_{HTS}, P_8).

The figure 2.1 shows that the strategies (t_{HTS}, P_5) and (t_{HTS}, P_9) are only a little more efficient than the simple sample mean. The relative efficiencies of the strategies are expressed by non-linear functions of the population size as well as of the correlation coefficient of the auxiliary variable and variable under study and the ratio of the variation coefficients $k=\gamma_x/\gamma_y$. As we have expected, under a fixed population size, the relative efficiencies of the strategies (t_{HTS},P_4), (t_{HTS}, P_6) usually decrease when the sample size is smaller and smaller.

In conclusion, from a practical point of view, the strategies (t_{HTS}, P_4) and (t_{HTS}, P_6) can be preferred. In the case of the strategy (t_{HTS}, P_4) the correlation coefficient should take a high positive value and the variation coefficient k should belong to the appropriate interval. In the case of the strategy (t_{HTS}, P_6) the absolute value of the correlation coefficient should be close to one. Both these strategies should be especially preferred instead of the simple sample mean when a population size as well as a sample size are rather small. Hence, they can be useful in the case of a two stage sampling design.

More details on results of the simulation study are showed by Wywiał (2000).

[1] See Zieliński (1979).

[2] SPSS Advanced Statistics 7.5. SPSS Inc. 1997.

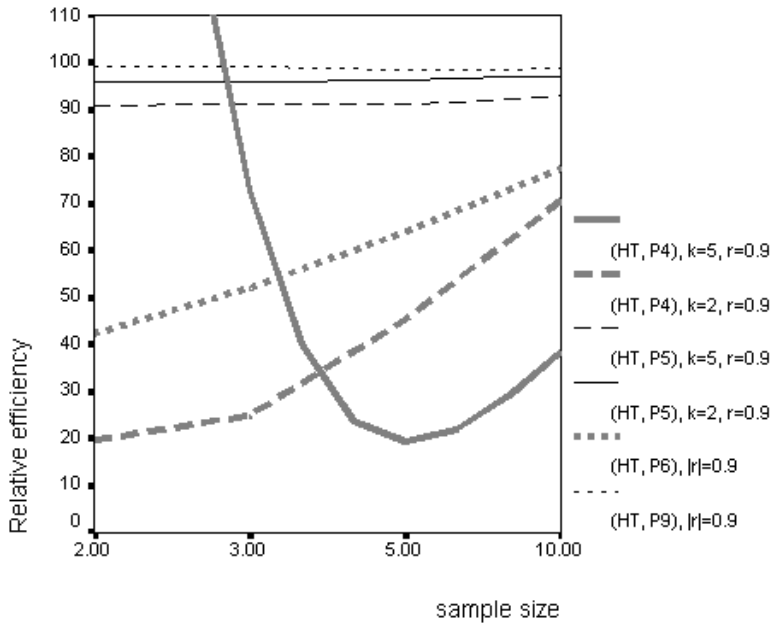


Figure 2.1.

2.3.8. Space sampling design

The present paragraph deals with sampling in a space population. It is assumed that the neighborhood of the population elements is fixed and identified by so called neighbor matrix. Four sampling designs are constructed on the basis of that matrix. Two of them prefer drawing population elements, which are neighbors. Next two ones prefer sampling elements which are not adjacent to each other. Those designs can be useful in research devoted to ecology, protecting environment, economic problems and so on.

The position of population elements can be identified by neighborhood matrix $\mathbf{A}=[a_{ij}]$. If the elements (i,j) are neighbors (are not neighbors) then $a_{ij}=1$ ($a_{ij}=0$).

Let us consider the following population:

5	2 1 4	3

The neighborhood matrix is as follows:

$$\mathbf{A}=\begin{bmatrix}1 & 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 0 & 1\\ 1 & 1 & 1 & 1 & 0\\ 1 & 0 & 1 & 1 & 1\\ 1 & 1 & 0 & 1 & 1\end{bmatrix}$$

The space of samples of size three is as follows:

S={(1,2,3); (1,2,4); (1,2,5); (1,3,4); (2,3,4); (2,3,5); (3,4,5); (1,3,5); (2,4,5); (1,4,5)}

Let $\mathbf{A}(s)=[a_{ij}(s)]$ be the neighborhood matrix of elements of a sample s. Hence, for the samples of the space **S** we have:

$$\mathbf{A}(1,2,3)=\mathbf{A}(1,2,5)=\mathbf{A}(1,3,4)=\mathbf{A}(1,4,5)=\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix};$$

$$\mathbf{A}(1,2,4)=\mathbf{A}(1,3,5)=\mathbf{A}(2,3,5)=\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 0\\ 1 & 0 & 1\end{bmatrix};$$

$$\mathbf{A}(2,3,4)=\mathbf{A}(3,4,5)=\begin{bmatrix}1 & 1 & 0\\ 1 & 1 & 1\\ 0 & 1 & 1\end{bmatrix};\qquad \mathbf{A}(2,4,5)=\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}.$$

Let a sample design prefers the neighbor elements to be drawn without replacement. Wywial (1996b) considered the following sampling design:

$$P_{10}(s)=\frac{\sum_{i=1}^{n}\sum_{j>i}a_{ij}(s)}{\sum_{s\in\mathbf{S}}\sum_{i=1}^{n}\sum_{j>i}a_{ij}(s)}. \tag{2.70}$$

Then for our population:

$$P(1,2,3)=P(1,2,5)=P(1,3,4)=P(1,4,5)=\frac{1}{8},$$

$$P(1,2,4)=P(2,3,4)=P(2,3,5)=P(3,4,5)=P(1,3,5)=P(2,4,5)=\frac{1}{12}.$$

The inclusion probabilities are as follows:

$$\pi_1=\frac{2}{3}, \qquad \pi_2=\pi_3=\pi_4=\pi_5=\frac{7}{12},$$

$$\pi_{12}=\pi_{13}=\pi_{14}=\pi_{15}=\frac{1}{3}, \qquad \pi_{23}=\pi_{25}=\pi_{34}=\pi_{45}=\frac{7}{24},$$

$$\pi_{24}=\pi_{35}=\frac{1}{4}.$$

In the case of sampling with replacement, the probability of selecting a k-th element is defined proportionally to the sum of elements which are its neighbors. Then:

$$p_{11}(k)=\frac{\sum_{j=1}^{N} a_{kj}}{\sum_{i=1}^{N}\sum_{j=1}^{N} a_{kj}}. \tag{2.71}$$

In the case of the population under consideration:

$$p_2(1)=\frac{5}{21}, \qquad p_2(2)=p_2(3)=p_2(4)=p_2(5)=\frac{4}{21}.$$

Let us define the design that prefers drawing without replacement the elements which are not neighbors. Then:

$$P_{12}(s) \propto \frac{1}{2}(n^2-n)+\alpha-\sum_{i=1}^{n}\sum_{j>i} a_{ij}.$$

$P_{12}(s)>0$ provided that $\alpha>0$.

$$P_{12}(s)=\frac{\frac{1}{2}n(n-1)+\alpha-\sum_{i=1}^{n}\sum_{j>i} a_{ij}}{\binom{N}{n}\left[\frac{1}{2}n(n-1)+\alpha\right]+\beta} \tag{2.72}$$

or

$$P_{12}(s)=\frac{\frac{1}{2}n(n-1)+\alpha-\beta P_1}{\binom{N}{n}\left[\frac{1}{2}n(n-1)+\alpha\right]+\beta} \tag{2.73}$$

where:

$$\beta = \sum_{s\in S}\sum_{i=1}^{n}\sum_{j>i} a_{ij} \ .$$

It is obvious that:

$$\lim_{\alpha\to\infty} P_{12} = \frac{1}{\binom{N}{n}} \tag{2.74}$$

Hence, if $\alpha\to\infty$, the sampling design $P_3(s)$ tends to be a simple sampling design.

In the case of our population and $\alpha=0.1$, we have:

$$P(1,2,3)=P(1,2,5)=P(1,3,4)=P(1,4,5)=\frac{1}{16},$$

$$P(1,2,4)=P(2,3,4)=P(2,3,5)=P(3,4,5)=P(1,3,5)=P(2,4\ 5)=\frac{1}{8},$$

$$\pi_1=\frac{13}{35}, \qquad \pi_2=\pi_3=\pi_4=\pi_5=\frac{23}{35}.$$

$$\pi_{12}=\pi_{13}=\pi_{14}=\pi_{15}=\frac{13}{70}, \quad \pi_{23}=\pi_{25}=\pi_{34}=\pi_{45}=\frac{23}{70}, \qquad \pi_{24}=\pi_{35}=\frac{33}{70}.$$

If $\alpha=0.5$ then:

$$P(1,2,3)=P(1,2,5)=P(1,3,4)=P(1,4,5)=\frac{1}{22},$$

$$P(1,2,4)=P(2,3,4)=P(2,3,5)=P(3,4,5)=P(1,3,5)=P(2,4\ 5)=\frac{3}{22}.$$

$$\pi_1=\frac{1}{2}, \qquad \pi_2=\pi_3=\pi_4=\pi_5=\frac{5}{8},$$

$$\pi_{12}=\pi_{13}=\pi_{14}=\pi_{15}=\frac{5}{22}, \qquad \pi_{23}=\pi_{25}=\pi_{34}=\pi_{45}=\frac{7}{22},$$

$$\pi_{24}=\pi_{35}=\frac{9}{22}.$$

In the case of sampling with replacement, probabilities of selecting the k-th population element can be defined as follows:

$$p_{13}(k) = \frac{N + 1 - \sum_{i=1}^{N} a_{ki}}{N(N+1) - \sum_{k=1}^{N}\sum_{i=1}^{N} a_{ki}} \tag{2.75}$$

Hence, the probability $p_{13}(k)$ increases if the sum of the elements which are adjacent to the k-th element decreases.

In the case of the population under our analysis we have:

$$p_2(1)=\frac{1}{9}, \qquad p_2(2)= p_2(3)= p_2(4)= p_2(5)=\frac{2}{9}.$$

Table 2.1

The variances

Strategy	Variance
$\bar{y}_S, P_3$	1.6400
t_{HTS}, P_{10}	0,7706
t_{HTS}, P_{12} ; α=0.1	3.3691
t_{HTS}, P_{12} ; α=0.5	40.4913

Let us define the variable y: y(1)=8, y(2)=1, y(3)=3, y(4)=1, y(5)=3. The population mean value is $\bar{y} = 3.2$ and the population variance is $v_* = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2 = 8.2$.

The population mean $\bar{y}$ is estimated on the basis of the sample mean $\bar{y}_S$ and of the Horvitz-Thompson estimator t_{HTS}, given by the expressions (2.1). Let us remember that P_3 is the sampling design of the simple sample without replacement, defined by the expression (1.28). The variance of the strategy is determined by the form (2.2). The table 2.1 refers the calculated variances of the strategies.

The variance of the strategy (t_{HTS}, P_{10}) takes the lowest value. The design of this strategy prefers selecting elements of the population which are neighbors.

2.3.9. Sampling designs dependent on the determinant of sample co-variances matrix

Let $\mathbf{x}=[x_{ij}]$ be the matrix of dimension m×N. It consists of the values of an m-dimensional auxiliary variable. Let $\mathbf{x}_{*j}^{T}=[x_{1j}...x_{mj}]$, j=1,...,N be an observation of the m-dimensional variable attached to a j-th population element. The vector $\mathbf{x}_{i*}=[x_{i1}...x_{iN}]$ (i=1,...,m) consists of observations of the i-th auxiliary variable. Then:

$$\mathbf{x} = [\mathbf{x}_{*1}\ \mathbf{x}_{*2}\ ...\ \mathbf{x}_{*N}] \quad \text{or} \quad \mathbf{x}=\begin{bmatrix}\mathbf{x}_{1*}\\ ...\\ \mathbf{x}_{m*}\end{bmatrix}.$$

Values of auxiliary variables observed in a sample s of size n can be written as the matrix $\mathbf{x}_S=[\mathbf{x}_{*j_1},...,\mathbf{x}_{*j_n}]$.

Let $\mathbf{z}=[x_{ij}-\bar{x}_i]$, where: $\bar{x}_i=\frac{1}{N}\sum_{j=1}^{N}x_{ij}$ (i=1,...,m; j=1,...,N) and let $\mathbf{z}_S=[z_{*j_1}...z_{*j_n}]$, where: $\mathbf{z}_{*j_k}^{T}=[(x_{1j_k}-\bar{x}_1)...(x_{mj_k}-\bar{x}_m)]$, k=1,..,n. Let $\mathbf{u}=[x_{ij}-\bar{x}_i(s)]$, where: $\bar{x}_i(s)=\frac{1}{n}\sum_{j\in s}x_{ij}$. Moreover, let $\mathbf{u}_S=[\mathbf{u}_{*j_1}...\mathbf{u}_{*j_n}]$, where: $u_{*j_k}^{T}=[(x_{1j_k}-\bar{x}_1(s))..(x_{mj_k}-\bar{x}_m(s))]$, k=1,..,n. Then, $\mathbf{z}_S$ and $\mathbf{u}_S$ are sub-matrices of the matrices $\mathbf{z}$ and $\mathbf{u}$, respectively. Hence, the sub-matrices $\mathbf{z}_S$ and $\mathbf{u}_S$ can be obtained through eliminating all the columns of the matrices $\mathbf{z}$ and $\mathbf{u}$ except those which correspond to the population elements drawn to a sample s.

The population generalised variance is defined by the following expression:

$$g=N^{-m}|\mathbf{z}\,\mathbf{z}^{T}| \tag{2.76}$$

The sample generalized variance is defined by the formula:

$$g_s=n^{-m}|\mathbf{u}_S\,\mathbf{u}_S^{T}| \tag{2.77}$$

All the sampling designs presented in this and next paragraphs were proposed by Wywial (1997a, 1999, 1999a).

Let us consider the following sampling design:

$$P_{14}(s)=\frac{|\mathbf{z}_s\mathbf{z}_s^{T}|}{\sum_{s\in\mathbf{S}}|\mathbf{z}_s\mathbf{z}_s^{T}|}$$

It can be transformed into the following form:

$$P_{14}(s)=\frac{|\mathbf{z}_s\mathbf{z}_s^T|}{\binom{N-m}{n-m}N^m g}, \quad s\in\mathbf{S} \tag{2.78}$$

When m=1, the sampling design $P_{14}(s)$ is reduced to the one considered by Singh and Srivastava (1980).

Let $\mathbf{z}(k_1,....k_r)$ be a sub-matrix obtained through eliminating the columns of numbers $k_1,...,k_r$ from the matrix $\mathbf{z}$. Wywial (1997a, 1999) derived the probabilities of drawing elements $k_1,...,k_r$ from a population for a sample in the r fixed selections. The probabilities of inclusion are as follows:

$$\pi_k^{(14)}=1-\frac{N-n}{N-m}\frac{|\mathbf{z}(k)\mathbf{z}^T(k)|}{N^{(m)}g}, \quad k=1,...,N \tag{2.79}$$

$$\pi_{kh}^{(14)}=1-\frac{N-n}{(N-m)N^m g}\left[\left|\mathbf{z}(h)\mathbf{z}^T(h)\right|+\left|\mathbf{z}(k)\mathbf{z}^T(k)\right|-\frac{N-n-1}{N-m-1}\left|\mathbf{z}(k,h)\mathbf{z}^T(k,h)\right|\right] \tag{2.80}$$

Let us define the following sampling design:

$$P_{15}(s)\propto|\mathbf{z}\mathbf{z}^T|-|\mathbf{z}_s\mathbf{z}_s^T|\geq 0\,.$$

The sampling design is as follows:

$$P_{15}(s)=\frac{|\mathbf{z}\mathbf{z}^T|-|\mathbf{z}_s\mathbf{z}_s^T|}{\left[\binom{N}{n}-\binom{N-m}{n-m}\right]|\mathbf{z}\mathbf{z}^T|} \tag{2.81}$$

or $$P_{15}(s)=\frac{1}{\binom{N}{n}-\binom{N-m}{n-m}}-\frac{\binom{N-m}{n-m}}{\binom{N}{n}-\binom{N-m}{n-m}}P_{14}(s)$$

where $P_{14}(s)$ is defined by the expression (2.78). Wywial (1997a) showed how to determine the inclusion probabilities.

Let us note that the sampling design $P_{14}(s)$ prefers (unlike the sampling design $P_{15}(s)$) the selection of such a sample s that the determinant $|\mathbf{z}_s\mathbf{z}_s^T|$ is large.

2.3.10. Sampling designs dependent on generalized sample variance

Wywiał (1997, 1999, 1999a) considered the sampling design proportional to the sample generalized variance g_s of an m-dimensional auxiliary variable which is defined by the expression (2.77). Let us consider the following sampling design:

$$P_{16}(s)=\frac{g_s}{\sum_{s\in\mathbf{S}} g_s},\ s\in\mathbf{S}$$

This design can be transformed into the following form:

$$P_{16}(s)=\frac{n|\mathbf{u}_s\mathbf{u}_s^T|}{\binom{N-m-1}{n-m-1}N^{m+1}g},\ s\in\mathbf{S} \tag{2.82}$$

When m=1, the sampling design $P_{16}(s)$ is reduced to the sampling design proportional to sample variance considered by Singh and Srivastava (1980).

Let $\mathbf{x}(k_1, ..., k_r)$ be a sub-matrix obtained through eliminating the columns of numbers $k_1, ..., k_r$ from the matrix $\mathbf{x}$. Moreover, let:

$$\mathbf{v}(k_1,...,k_w)=\mathbf{x}(k_1,...,k_w)-\bar{\mathbf{x}}(k_1,...,k_w)\mathbf{J}_{N-w}^T .$$

$\mathbf{J}_{N-w}$ is the column vector with all its (N-w) elements equal to one and:

$$\bar{\mathbf{x}}(k_1,...,k_w)=\frac{1}{N-w}\mathbf{x}(k_1,...,k_w)\mathbf{J}_{N-w} .$$

Wywial (1997a, 1999a) derived the probability of drawing without replacement elements $k_1,...,k_r$ to a sample s during the r fixed selections from a population. The inclusion probabilities of order r=1 and r=2 are as follows:

$$\pi_k^{(16)}=1-\frac{\binom{N-m-2}{n-m-1}(N-1)}{\binom{N-m-1}{n-m-1}N^{m+1}g}\left|\mathbf{v}(k)\mathbf{v}^T(k)\right| . \tag{2.83}$$

$$\pi_{kh}^{(16)} = 1 - \frac{1}{\binom{N-m-1}{n-m-1} N^{m+1} g} \left\{ \binom{N-m-2}{n-m-1} (N-1) \left[\left| \mathbf{v}(k)\mathbf{v}^T(k) \right| + \right.\right.$$

$$\left.\left. + \left| \mathbf{v}(h)\mathbf{v}^T(h) \right| \right] - \binom{N-m-3}{n-m-1} (N-2) \left| \mathbf{v}(k,h)\mathbf{v}^T(k,h) \right| \right\}. \qquad (2.84)$$

Let us considered the following sampling design:

$$P_{17}(s) = \frac{N|\mathbf{z}\mathbf{z}^T| - n|\mathbf{u}_s\mathbf{u}_s^T|}{\sum_{s \in \mathbf{S}} \left(N|\mathbf{z}\mathbf{z}^T| - n|\mathbf{u}_s\mathbf{u}_s^T| \right)}, \; s \in \mathbf{S}.$$

It can be transformed into the following form:

$$P_{17}(s) = \frac{1}{\binom{N}{n} - \binom{N-m-1}{n-m-1}} \left(1 - \frac{n|\mathbf{u}_s\mathbf{u}_s^T|}{N^{m+1} g} \right) \qquad (2.85)$$

or $$P_{17}(s) = \frac{1}{\binom{N}{n} - \binom{N-m-1}{n-m-1}} \left(1 - \binom{N-m-1}{n-m-1} P_{16}(s) \right).$$

where: $P_{16}(s)$ is given by the expression (2.82).

The sampling design $P_{17}(s)$ prefers (opposite to the sampling design $P_{16}(s)$, given by the formula (2.82)) selection of such a sample s that the generalized variance $g_S = n^{-m} \left| \mathbf{u}_S \mathbf{u}_S^T \right|$ is not large.

Wywiał (1997a, 1999a) derived the probability of drawing elements from a population into a sample during the r fixed selections.

The next sampling design is as follows:

$$P_{18}(s) = \frac{|\mathbf{z}\mathbf{z}^T| - |\mathbf{u}_s\mathbf{u}_s^T|}{\sum_{s \in \mathbf{S}} \left(|\mathbf{z}\mathbf{z}^T| - |\mathbf{u}_s\mathbf{u}_s^T| \right)}, \qquad s \in \mathbf{S}.$$

We can prove that:

$$P_{18}(s) = \frac{n\left(gN^m - g_s n^m \right)}{\left[\binom{N-1}{n-1} - \binom{N-m-1}{n-m-1} \right] gN^{m+1}} \qquad (2.86)$$

The probability of drawing elements $k_1,...,k_r$ from a population into a sample s during the r fixed selections can be derived (see Wywiał (1997a, 1999a)). The inclusion probabilities are as follows:

$$\pi_k^{(18)} = \frac{1}{N\left[\binom{N-1}{n-1} - \binom{N-m-1}{n-m-1}\right]}$$

$$\left\{1 - \frac{N}{n}\binom{N-m-1}{n-m-1} + \frac{\binom{N-m-2}{n-m-1}(N-1)|\,\mathbf{v}(k)\mathbf{v}^T(k)\,|}{ngN^m}\right\} \tag{2.87}$$

$$\pi_{kh}^{(18)} = \frac{1}{(n-2)N\left[\binom{N-1}{n-1} - \binom{N-m-1}{n-m-1}\right]}\left\{1 - \frac{N}{n}\binom{N-m-1}{n-m-1} + \right.$$

$$+ \frac{N-1}{ngN^m}\binom{N-m-2}{n-m-1}\left(\mathbf{v}(k)\mathbf{v}^T(k) + \mathbf{v}(h)\mathbf{v}^T(h)\right) +$$

$$\left. - \frac{N-2}{ngN^m}\binom{N-m-3}{n-m-1}\left|\mathbf{v}(k,h)\mathbf{v}^T(k,h)\right|\right\}. \tag{2.88}$$

Let us suppose that the average value of a variable under study is highly dependent on a vector of auxiliary variables. We can expect that some of the considered sampling designs can provide a good sampling strategy of estimation of the population average of the variable under study. The Horvitz–Thompson estimator will be the unbiased estimator of the average but it seems that any analysis of its accuracy should be supported
by a computer simulation. The sampling designs proportional to a generalized variance P_{14} or P_{16} will be used to construct the regression strategies at the end of the chapter 6.

III. STRATIFIED SAMPLING

3.1. Basic properties

Let us assume that the population Ω is divided into non-empty strata denoted by Ω_h (h=1,...,H) and $\Omega_h \cap \Omega_l = \varnothing$ for each pair $h \neq l = 1,...,H$ and $\bigcup_{h=1}^{H} \Omega_h = \Omega$. Let N_h be the size of the stratum Ω_h (h=1,...,H) and $w_h = \frac{N_h}{N}$, where: $N = \sum_{h=1}^{H} N_h$. A simple sample of size $0 < n_h \leq N_h$ drawn from the h-th stratum will be denoted by S_h, h=1,...,H. The sampling design of the stratified sample $S=\{S_1,...,S_H\}$ is as follows:

$$P_w(s) = \prod_{h=1}^{H} \binom{N_h}{n_h}^{-1} \tag{3.1}$$

in the case when the samples are drawn without replacement or

$$P_w^{©}(s) = \prod_{h=1}^{H} N_h^{-n_h} \tag{3.2}$$

in the case when the samples are drawn with replacement.

Let y_{ihk} be the k-th observation of an i-th variable in an h-th stratum.

$$\bar{y}_{ih} = \frac{1}{N_h} \sum_{k=1}^{N_k} y_{ihk}, \qquad v_{*h}(y_i) = c_{*h}(y_i, y_i)$$

$$c_{*h}(y_i, y_j) = \frac{1}{N_h - 1} \sum_{k=1}^{N_h} (y_{ihk} - \bar{y}_{ih})(y_{jhk} - \bar{y}_{jh})$$

$$\bar{\mathbf{y}}_h = [\bar{y}_{1h} \cdots \bar{y}_{mh}], \qquad \mathbf{C}_{*h} = [c_{*h}(y_i, y_j)].$$

The statistics $\bar{\mathbf{y}}_{Sh} = [\bar{y}_{1S_h} \cdots \bar{y}_{mS_h}]$ and $\mathbf{C}_{*S_h} = [c_{*S_h}(y_i, y_j)]$, where:

$$c_{*S_h}(y_i, y_j) = \frac{1}{n_h - 1} \sum_{k \in S_h} (y_{ihk} - \bar{y}_{iS_h})(y_{jhk} - \bar{y}_{jS_h}), \ \bar{y}_{iS_h} = \frac{1}{n_h} \sum_{k \in S_h} y_{ihk} \tag{3.3}$$

are unbiased estimators of parameters: $\bar{\mathbf{y}}_h = [\bar{y}_{1h} \cdots \bar{y}_{mh}]$ and $\mathbf{C}_{*h} = [c_{*h}(y_i, y_j)]$, respectively.

The vector $\bar{\mathbf{y}}$ of population means is the following function of the strata means:

$$\bar{\mathbf{y}} = \sum_{h=1}^{H} w_h \bar{\mathbf{y}}_h .$$

An unbiased estimator of the vector $\bar{\mathbf{y}}$ is defined by the expression:

$$\bar{\mathbf{y}}_{wS} = \sum_{h=1}^{H} w_h \bar{\mathbf{y}}_{hS} . \qquad (3.4)$$

Its covariance matrix in the case of sampling without replacement is as follows:

$$\mathbf{V}(\bar{\mathbf{y}}_{wS}, P_w) = \sum_{h=1}^{H} w_h^2 \mathbf{V}(\bar{\mathbf{y}}_{S_h}) = \sum_{h=1}^{H} w_h^2 \frac{N_h - n_h}{N_h n_h} \mathbf{C}_{*h} . \qquad (3.5)$$

The variance of the i-th element of the vector of estimators $\bar{\mathbf{y}}_{wS}$ is as follows:

$$D^2(\bar{y}_{wiS}, P_w) = \sum_{h=1}^{H} w_h^2 \frac{N_h - n_h}{N_h n_h} v_{*h}(y_i) . \qquad (3.6)$$

An unbiased estimator of the matrix $\mathbf{V}(\bar{\mathbf{y}}_{wS}, P_w)$ is as follows:

$$\mathbf{V}_S(\bar{\mathbf{y}}_{wS}, P_w) = \sum_{h=1}^{H} w_h^2 \frac{N_h - n_h}{N_h n_h} \mathbf{C}_{*S_h} \qquad (3.7)$$

where the elements of the matrix $\mathbf{C}_{*S_h} = [c_{*S_h}(y_i, y_j)]$ are determined by the equation (3.3).

In the case of the sampling with replacement, the covariance matrix of the estimator $\bar{\mathbf{y}}_{wS}$ is as follows:

$$\mathbf{V}(\bar{\mathbf{y}}_{wS}, P_w') = \sum_{h=1}^{H} \frac{1}{n_h} w_h^2 \mathbf{C}_h \qquad (3.8)$$

where:
$$\mathbf{C}_h = \frac{N_h - 1}{N} \mathbf{C}_{*h}$$

An unbiased estimator of the matrix $\mathbf{V}(\bar{\mathbf{y}}_{wS}, P_w')$ is as follows:

$$\mathbf{V}_S(\bar{\mathbf{y}}_{wS}, P_w) = \sum_{h=1}^{H} \frac{w_h^2}{n_h} \mathbf{C}_{*Sh} \qquad (3.9)$$

3.2. Optimization of samples sizes

3.2.1. Proportional determination of sample sizes

Let n be the total size of a sample selected from strata i.e. $n = \sum_{h=1}^{H} n_h$. Let $n_{\#h}$ be size of a sample selected from the h-th stratum. When we assume that $n_{\#h}$ is proportional to the fraction of population elements in the stratum:

$$n_{\#h} = nw_h, \quad h=1,...,H \tag{3.10}$$

The sampling designs for samples of such sizes, selected without replacement from strata, can be derived on the basis of the expressions (3.1), (3.2) and (3.10). It will be denoted by $P_p^{'}$ and P_p in the cases of sampling with replacement and without replacement, respectively. The expressions (3.5) and (3.10) lead to the following variance-covariance matrix of the sampling strategy:

$$\mathbf{V}(\bar{\mathbf{y}}_{wS}, P_p) = \frac{N-n}{Nn} \sum_{h=1}^{H} w_h \mathbf{C}_{*h} \tag{3.11}$$

In the case of sampling with replacement, the equations (3.8) and (3.10) lead to the following variance-covariance matrix:

$$\mathbf{V}(\bar{\mathbf{y}}_{wS}, P_p^{'}) = \frac{1}{n} \sum_{h=1}^{H} w_h \mathbf{C}_h \tag{3.12}$$

Theorem 3.1 [Wywiał (1992)]: The sampling strategy $(\bar{\mathbf{y}}_{wS}, P_p^{'})$ is not worse than the strategy $(\bar{\mathbf{y}}_S, P_1)$ in the sense of the definition 1.18. When the sizes of the strata N_h (h=1,...,H) are sufficiently large, the sampling strategy $(\bar{\mathbf{y}}_{wS}, P_p)$ is not worse than the strategy $(\bar{\mathbf{y}}_S, P_3)$, where $(\bar{\mathbf{y}}_S, P_1)$ and $(\bar{\mathbf{y}}_S, P_3)$ are mean vectors from a simple sample drawn with and without replacement, respectively.

Proof: Let $\mathbf{B} = [\bar{y}_{ih} - \bar{y}_i]$ be the matrix of dimension m×H, where $\bar{y}_{ih}$ is the mean of an i-th variable in an h-stratum. The vector of stratum fraction is $\mathbf{w}=[w_1...w_H]$. Its elements are the diagonal elements of the matrix $\mathbf{D}_w=\text{diag}(\mathbf{w})$. After a simple transformation we have:

$$\mathbf{V}(\bar{\mathbf{y}}_S, P_1) = \mathbf{V}(\bar{\mathbf{y}}_{wS}, P_p^{'}) + \frac{1}{n}\mathbf{B}\mathbf{D}_w\mathbf{B}^T .$$

The matrix $\mathbf{BD}_w\mathbf{B}^T$ is non-negative definite. This and the theorem 1.6 let us say that the strategy $(\bar{\mathbf{y}}_{wS}, P_p')$ is not worse than $(\bar{\mathbf{y}}_S, P_1)$ (the means from the simple sample drawn with replacement). In the case of the samples drawn with replacement the proof is similar.

This result and the theorem 1.7 lead to the conclusion that all considered synthetic measures of the strategy $(\bar{\mathbf{y}}_{wS}, P_p')$ precision are not larger than the appropriate measures of the strategy $(\bar{\mathbf{y}}_S, P_1)$ precision. A similar conclusion deals with strategies $(\bar{\mathbf{y}}_{wS}, P_p)$ $(\bar{\mathbf{y}}_{wS}, P_w)$ and $(\bar{\mathbf{y}}_S, P_3)$.

3.2.2. Minimization of risk function under fixed cost of data observation

Neyman (1934) determined sample sizes selected from strata through minimization of the estimator variance under a fixed total of sample sizes[16]. The solution to this problem is connected with only one variable under study. A generalization of this optimization problem is considered.

Costs of data observation are described by the following linear function[17]:

$$k(\mathbf{n}) = k(n_1,...,n_H) = \sum_{h=1}^{H} k_h n_h \tag{3.13}$$

where: $\mathbf{n} = [n_1...n_h]$ and k_h is the unit cost of data observation in an h-th stratum, h=1,...,H. An admissible level of the total costs of data observation is denoted by K.

Let us consider the following risk function:

$$f_1(\mathbf{n}) = \sum_{i=1}^{m} a_i D^2(t_{wiS}, P_w) \tag{3.14}$$

where $a_i>0$ and $D^2(\bar{y}_{wiS}, P_w)$, i=1,...,m, are determined by the expression (3.6). After some simple transformations:

$$f_1(\mathbf{n}) = \sum_{h=1}^{H} \frac{w_h^2 b_h^2}{n_h} - \frac{1}{N}\sum_{h=1}^{H} w_h b_h^2 \tag{3.15}$$

where:

$$b_h^2 = \sum_{i=1}^{m} a_i v_{*h}(y_i) \tag{3.16}$$

[16] Cochran (1963), p. 97, noted that this problem had been stated and solved by Tschuprow (1923), too.

[17] Beardwood, Halton i Hammersley (1959) considered a non-linear cost function.

If $a_i=1$ for i=1,...,m, $f_1(\mathbf{n})$ is equal to the trace of the covariance matrix of the sampling strategy $(\bar{\mathbf{y}}_{wS}, P_w)$.

The problem is: how to determine such a vector of sample sizes $\underline{\mathbf{n}}=[\underline{n}_1...\underline{n}_H]$ that the risk function takes the minimal value under fixed total costs of observation K.

$$\begin{cases} f_1(\mathbf{n}) = \text{minium} \\ k(\mathbf{n}) \le K, \mathbf{o}_h < \mathbf{n} \le N\mathbf{w} \end{cases} \tag{3.17}$$

where: $k(\mathbf{n})$ is given by the expression (3.13), $\mathbf{o}_H$ is a vector of the dimension 1×m whose all elements are equal to zero, $\mathbf{w}$ is a vector of strata fractions and $N\mathbf{w}=[N_1...N_H]$.

A more general problem was solved by Hughes and Rao (1979). In our case, the algorithm of deriving the solution to the problem (3.17) is as follows: firstly, we determine the quantity:

$$A_h = \frac{b_h}{N\sqrt{k_h}}, \quad h = 1,...,H.$$

Let us assume that the sequence (A_h) is not increasing. Moreover, let

$$\begin{cases} G_0 = 0, \quad G_1 = \dfrac{1}{A_1}\sum\limits_{h=1}^{H} w_h b_h \sqrt{k_h} \\ G_{i+1} = \dfrac{1}{A_{i+1}}\left(\sum\limits_{h=i+1}^{H} w_h b_h \sqrt{k_h}\right) + \sum\limits_{h=1}^{i} k_h N_h, \quad i = 1,..., H-1 \\ G_H = \sum\limits_{h=1}^{H} k_h N_h \end{cases} \tag{3.18}$$

Let z $(z = 0,1,..., H-1)$ be such index that $K \in [G_z; G_{z+1})$. Finally, the solution is shown by the expression:

$$\begin{cases} \underline{n}_h = N_h \qquad \text{dla } h = 1,..., z \\ \underline{n}_h = c_o \dfrac{w_h b_h}{\sqrt{k_h}} \text{ dla } h = z+1,..., H \end{cases} \tag{3.19}$$

where:

$$c_o = \frac{K - \sum\limits_{h=1}^{z} k_h N_h}{\sum\limits_{h=z+1}^{H} w_h b_h \sqrt{k_h}}. \tag{3.20}$$

Particularly, let $K \in (0; G_1]$, what means that $K \le G_1 = \frac{1}{A_1}\sum_{h=1}^{H} w_h b_h \sqrt{k_h}$. So, in this case z=0 and $\underline{n}_h \le N_h$ for all h=1,...,H. Moreover, let us note that z=0, if $N_h \to \infty$ for each h = 1,...,H or

$$N_h > \frac{K}{k_0} \tag{3.21}$$

where: $k_o = \min\limits_{h=1,...,H} \text{imum}\{k_h\}$.

If m=1, solution (3.19) is reduced to the well known Neyman's location of samples in strata.

A particular case of the risk function $f_1(\mathbf{n})$ is equal to $q^2(\bar{\mathbf{y}}_{wS}, P_w)$, if a_i=1 for all i=1,...,m. Greń (1964) proposed $a_i = \frac{1}{\bar{y}_i^2}$ for all i=1,...,m. In this case $f_1(\mathbf{n})$ is the sum of the squared variation coefficients of elements of the estimator vector $\bar{\mathbf{y}}_{wS}$.

Wywiał (1990) considered more general risk function. Auxiliary variables are applied to optimization sample sizes by Dayal (1985).

3.2.3. Minimization of total cost of observation under fixed risk function

Let us consider the following optimization problem:

$$\begin{cases} k(\mathbf{n}) = \text{minimum} \\ f_1(\mathbf{n}) \le f_0, \mathbf{o}_h < \mathbf{n} < N\mathbf{w} \end{cases} \tag{3.22}$$

The constrain $f_1(n) \le f_0$ is equivalent to the following one:

$$f_2(\mathbf{n}) = \sum_{h=1}^{H} \frac{w_h^2 b_h^2}{n_h} \le f_0 + \frac{1}{N}\sum_{h=1}^{H} w_h b_h = f_*$$

The solution is a particular case of the solution to a more general problem formulated and solved by Hughes and Rao (1979). The algorithm of getting the solution is as follows:

$$B_h = \frac{N\sqrt{k_h}}{b_h}, h = 1,...H \tag{3.23}$$

where the parameter b_h is defined by the expression (3.16). Let us assume that the sequence (B_h) is not decreasing. Let

$$\begin{cases} f_{*0} = \infty \\ f_{*1} = \dfrac{b_1}{N\sqrt{k_1}} \sum_{t=1}^{H} w_t b_t \sqrt{k_t} \\ f_{*h+1} = \dfrac{1}{N}\left(\sum_{t=}^{h} w_t b_t^2 + \dfrac{b_{h+1}}{\sqrt{k_{h+1}}} \sum_{t=h+1}^{H} w_t b_t \sqrt{k_t} \right), \quad \text{for} \quad h = 1,...,H-1 \end{cases} \tag{3.24}$$

Let q be such index that $f_* \in (f_{*q+1}; f_{*q}]$, q=0,1,...,H-1. The solution to the problem is as follows:

$$\begin{cases} \underline{n}_h = N_h & \text{dla } h = 1,...,q \\ \underline{n}_h = c_1 \dfrac{w_h b_h}{\sqrt{k_h}} & \text{dla } h = q+1,...,H \end{cases} \tag{3.25}$$

where:

$$c_1 = \frac{f_0 + \dfrac{1}{N} \sum_{t=q+1}^{H} w_t b_t^2}{\sum_{t=q+1}^{H} w_t b_t \sqrt{k_t}}.$$

Particularly, if $f_* > f_{*1}$, then q=0.

Let us note that some other optimisation problems were considered e.g. by Melaku (1987) or Mukerjee and Rao (1985).

3.2.4. Minimization of total risk

The function of the total risk is the sum of the cost function given by the expression (3.13) and the risk function determined by the equation (3.15) and it is shown by the expression[18]:

$$f_3(\mathbf{n}) = \sum_{h=1}^{H} k_h n_h + \sum_{h=1}^{H} \frac{N_h - n_h}{N_h n_h} w_h^2 b_h^2 \tag{3.26}$$

where the parameter b_h^2 is defined by the expression (3.16). The problem has the following solution for $\mathbf{o}_h < \mathbf{n} < N\mathbf{w}$[19]:

[18] Dalenius (1957). noted that this problem was formed by Blythe (1945).

[19] See Yates (1960) and Wywiał (1992).

$$\underline{n}_h = \text{minimum}\left\{\frac{w_h b_h}{\sqrt{k_h}}, N_h\right\}. \tag{3.27}$$

The optimal sample size selected from an h-th stratum can be equal to a stratum size or it is proportional to the product of the fraction of an h-th stratum and to the b_h coefficient.

3.2.5. Optimization of sizes of samples selected from strata under the fixed standard errors of estimators

Dalenius (1957) formed the problem how to determine the sizes of simple samples selected without replacement from strata in such a way that the cost function takes the minimal value under the fixed levels of standard errors of estimators of means. He solved this problem in the case of two strata and two estimated population means. In general, for the samples drawn with replacement, the problem was solved by e.g. Greń (1963, 1966), Hartley (1965), Huddleston, Claypool and Hocking (1970), Jaganathan (1965, 1965a), Kokan (1963) and Yates (1960). Finally, in the general case of sampling without replacement, the problem was solved by Kokan and Khan (1967). The problem described here can be defined as follows:

$$\begin{cases} k(\mathbf{n}) = \text{minimum} \\ D^2(y_{wiS}) \le e_i, \quad i = 1,...,m \\ 1 \le n_h \le N_h, \quad h,...,H \end{cases} \tag{3.28}$$

where: the cost function is defined by the (3.13) and the variance $D^2(.)$ by (3.6). In order to solve this problem the transformation $x_h = \frac{1}{n_h}$, h=1,...,H, leads to the following form of the problem:

$$\begin{cases} z(\mathbf{x}) = \text{minimum} \\ \mathbf{u}_i \mathbf{x}^T \le e_{*i}, i = 1,...,m \\ \frac{1}{N_h} \le x_h \le 1, h = 1,...,H \end{cases} \tag{3.29}$$

where:

$$\mathbf{x} = [x_1 ... x_H], \qquad z(\mathbf{x}) = \sum_{h=1}^{H} \frac{k_h}{x_h}, \qquad \mathbf{u}_i = \left[w_1^2 v_{*1}(y_i) .. w_H^2 v_{*H}(y_i)\right].$$

$$e_{*i} = e_i + \frac{1}{N}\sum_{h=1}^{H} w_h^2 v_{*h}(y_i)$$

Kokan and Khan (1967) proposed an appropriate algorithm leading to a solution to their optimization problem. Bethel (1989) proposed a simplified method of solving the problem (3.28). Chatterjee (1968) considered the dual problem to given by the expression (3.28). In some sense a particular case of the problem (3.28) was formed by Ghosha (1963).

Skibicki and Wywiał (2002) considerd the problem (3.28) in the case when the cost of observation of all population elements is the same. In this case the unit cost function is simplified to the form: $t(n_1,...,n_H)=\sum_{h=1}^{H} n_h$. They stated the following problem of stratifiacation of a population. The population is partitioned into strata and at the same time the sizes of samples drawn from these strata are determined in such a way that the sum $t(n_1,...,n_H)$ takes a minimal value under the fixed variances of estimators of mean values. The variances are the functions of observations of variables in strata and they depend on partition of a population into strata. Hence, the problem (3.28) is generalizing in such a way that additionally the optimal partition of a population into strata is evaluated. This problem can be considered in the situation when we have census data from a population and we look for optimal sampling design for future survey of the population.

3.2.6. Optimization of sample size on the basis of the spectral radius of variance-covariance matrix of estimators

In the paragraph 1.5.1 the spectral radius of the variance-covariance matrix $\mathbf{V}(\overline{\mathbf{y}}_{wS}, P_w)$ was defined as a maximal eigenvalue of this matrix. Let us treat this spectral radius as the function of sample sizes $\mathbf{n}=[n_1...n_H]$ and denote by f(**n**). Let us consider the following optimization problem:

$$\begin{cases} f(\mathbf{n}) = \text{minimum} \\ k(\mathbf{n}) \le K, \quad \mathbf{o}_H < \mathbf{n} \le N\mathbf{w} \end{cases} \tag{3.30}$$

where: $\mathbf{w} = [w_1...w_H]$.

The purpose function f(**n**) can be obtained as maximal value of characteristic polynomial of the matrix $\mathbf{V}(\overline{\mathbf{y}}_{wS}, P_w)$. This polynomial can be written in the following way:

$$F(\mathbf{n}, f) = \sum_{i=0}^{m} g_i f^i \tag{3.31}$$

where:

$$g_i = \sum_{j=1}^{\binom{m}{i}} g_{ij}, \qquad g_m = 1 \tag{3.32}$$

A j-th principal minor of the (m-i) degree of the matrix $\mathbf{V}(\overline{\mathbf{y}}_{wS}, P_w)$ is denoted by g_{ij}.

Hence, the problem (3.30) is equivalent to looking for a conditional minimum of the implicit function F(**n**,f)=0. It is a rather complicated numerical problem.

Wywiał (1988a) proposed the following simplification of the problem (3.30). The well known basic properties of matrix norm lead to the conclusion that the spectral radius f(**n**) is not greater than the function r(**n**) for each **n**>**o**, where:

$$r(\mathbf{n}) = \sum_{h=1}^{H} w_h^2 \frac{N_h - n_h}{N_h n_h} \rho(\mathbf{C}_{*h}) \tag{3.33}$$

$\rho(\mathbf{C}_{*h})$ is the spectral radius of the matrix $\mathbf{C}_{*\mathbf{h}}$ which is the variance-covariance matrix of variables under study in the h-th stratum, h=1,...,H. Our simplified problem is as follows:

$$\begin{cases} r(\mathbf{n}) = \text{minimum} \\ k(\mathbf{n}) \le K, \mathbf{o}_h < \mathbf{n} \le N\mathbf{w} \end{cases} \tag{3.34}$$

The solution to this problem is determined by the expression (3.19), if we state that $b_h = \sqrt{\rho(\mathbf{C}_{*h})}$, for k=1,...,H.

The problem of determining the optimal sample sizes drawn form the strata is formed in such a way that the cost function k(**n**) takes the minimum value under a fixed level of the function r(n) can be solved similarly to the problem defined by the expression (3.22).

Let us consider the following squared form of the matrix $\mathbf{V}(\bar{\mathbf{y}}_{wS}, P_w)$:

$$q(\boldsymbol{\alpha}, \mathbf{n}) = \boldsymbol{\alpha}\mathbf{V}(\bar{\mathbf{y}}_{wS}, P_w)\boldsymbol{\alpha}^T$$

where: $\boldsymbol{\alpha}=[\alpha_1...\alpha_m] \in R^m-\{\mathbf{o}\}$. Hence:

$$f(\boldsymbol{\alpha}, \mathbf{n}) = \max_{\boldsymbol{\alpha}\boldsymbol{\alpha}^T=1} \{q(\boldsymbol{\alpha}, \mathbf{n})\}$$

Let us simplify the set of admissible solutions to the problem (3.30) for the set $\mathbf{B} = \left\{ n_h, h = 1,..., H : \sum_{h=1}^{H} n_h = n_0 > 0 \right\}$, where n_0 is the sum of sizes of the simple samples drawn from the strata. The problem (3.30) can be rewritten in the following way:

$$\begin{cases} \min\limits_{\mathbf{n}\in\mathbf{B}} \max\limits_{\alpha\in\mathbf{A}} q\{(\alpha, \mathbf{n})\} \\ \mathbf{B} = \{\mathbf{n} : \mathbf{J}\mathbf{n}^T = \mathbf{n}_o > \mathbf{0}\} \\ \mathbf{A} = \{\alpha : \alpha\alpha^T = 1\} \end{cases} \tag{3.35}$$

An iteration method of solving to the problem is constructed in the following way[20]:

$$q(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{n}}) = \max_{\boldsymbol{\alpha}\in \mathbf{A}} \min_{\mathbf{n}\in \mathbf{B}} \{q(\boldsymbol{\alpha}, \mathbf{n})\} \tag{3.36}$$

Under the fixed $\boldsymbol{\alpha}$, the function $q(\boldsymbol{\alpha},\mathbf{n})$ takes the minimum value in the point $\mathbf{n}(\boldsymbol{\alpha}) = [n_1(\boldsymbol{\alpha})...n_H(\boldsymbol{\alpha})]$, where:

$$n_h(\boldsymbol{\alpha}) = \frac{n_0}{\overline{q}(\boldsymbol{\alpha})} w_h \sqrt{q_h(\boldsymbol{\alpha})}\,, \qquad h=1,..,H \tag{3.37}$$

where:

$$q_h(\boldsymbol{\alpha}) = \boldsymbol{\alpha}\mathbf{C}_h\boldsymbol{\alpha}^T, \qquad \overline{q}(\boldsymbol{\alpha}) = \sum_{h=1}^{H} w_h \sqrt{q_h(\boldsymbol{\alpha})}\,.$$

The Lagrange function is as follows:

$$F(\boldsymbol{\alpha}, \tilde{\mathbf{n}}) = q(\boldsymbol{\alpha}, \tilde{\mathbf{n}}) - \lambda(\boldsymbol{\alpha}\boldsymbol{\alpha}^T - 1)$$

The necessary condition for the existence the extreme of the function is as follows:

$$\frac{\partial F}{\partial \boldsymbol{\alpha}} = \frac{2}{n_0} q(\boldsymbol{\alpha}) \frac{\partial q}{\partial \boldsymbol{\alpha}} - \lambda\boldsymbol{\alpha} = 0, \tag{3.39}$$

$$\frac{\partial q}{\partial \boldsymbol{\alpha}} = \frac{\boldsymbol{\alpha}}{2}\mathbf{G}(\boldsymbol{\alpha})$$

where:

$$\mathbf{G}(\boldsymbol{\alpha}) = \sum_{h=1}^{H} \frac{w_h}{\sqrt{q_h(\boldsymbol{\alpha})}} \mathbf{C}_h\,.$$

After multiplying the equation (3.39) from the right side by $\boldsymbol{\alpha}^T$, we have:

$$\lambda = \frac{1}{n_0} q(\boldsymbol{\alpha})\boldsymbol{\alpha}\mathbf{G}(\mathbf{a})\boldsymbol{\alpha}^T$$

This and the expression (3.39) lead to the following:

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}\mathbf{P}(\boldsymbol{\alpha}) \tag{3.40}$$

[20] See Wywiał and Kończak (1994) or Wywiał (1995).

where:

$$\mathbf{P}(\boldsymbol{\alpha}) = \frac{\mathbf{G}(\boldsymbol{\alpha})}{\boldsymbol{\alpha}\mathbf{G}(\boldsymbol{\alpha})\boldsymbol{\alpha}^{T}}$$

This leads to the construction of the following iteration expression:

$$\boldsymbol{\alpha}_{t+1}=\boldsymbol{\alpha}_t\mathbf{P}(\boldsymbol{\alpha}_t), \qquad t=1,2,... \tag{3.41}$$

Next on the basis of the expressions (3.37) and (3.38), the elements of the vector $\mathbf{n}_t = [n_1(\boldsymbol{\alpha}_t)...n_H(\boldsymbol{\alpha}_t)]$ are determined. If a norm $\|\mathbf{n}_t-\mathbf{n}_{t-1}\|$ will be lower than an admissible level, the elements of the vector $\mathbf{n}_t$ are treated as a sufficiently good approximation of optimal sample sizes which should be selected from strata.

Skibicki (2003) proved that f(**n**) is a convex function of **n** and proposed the solution to the problem (3.30) based on a gradient method.

3.2.7. Optimal allocation of strata in the case of testing a hypothesis about vector of population means

The hypothesis concerning the vector of population means is considered. It is tested on the basis of data observed in a stratified sample in the case of large sizes of samples as well as large sizes of strata.

Let us consider the sequence of the population $\{U^{(\upsilon)}\}$, $\upsilon = 1,2....$. The size $N^{(\upsilon)}$ increases when $\upsilon \to \infty$. Let us assume that $U^{(\upsilon)}$ is divided into non-empty and disjoint strata $U_1^{(\upsilon)},...,U_H^{(\upsilon)}$ of sizes $N_1^{(\upsilon)},...,N_H^{(\upsilon)}$, respectively, and let $\sum_{h=1}^{H} N_h^{(\upsilon)} = N^{(\upsilon)}$. In each stratum $U_h^{(\upsilon)}$, the array $\mathbf{y}_h^{(\upsilon)} = \left[y_{h,i,j}^{(\upsilon)}\right]$ is observed, where i=1,..., $N_h^{(\upsilon)}$, j =1,...,k, h=1,...,H. Each row of the matrix $\mathbf{y}_h^{(\upsilon)}$ is the observation of a k-dimensional variable. Let

$$\overline{\mathbf{y}}_h = \frac{1}{N_h^{(\upsilon)}}\mathbf{J}^{T}_{N_h^{(\upsilon)}}\mathbf{y}_h^{(\upsilon)}, \quad h = 1,..., H$$

where $\mathbf{J}_a$ is the column vector consisting of a-elements all equal to one. So, $\overline{\mathbf{y}}_h$ is a row-vector consisting of the means in an h-stratum. A row-vector of population means is $\overline{\mathbf{y}} = \sum_{h=1}^{H} w_h \overline{\mathbf{y}}_h$ where: $w_h = \frac{N_h^{(\upsilon)}}{N^{(\upsilon)}}$, $h = 1,..., H$

The variance-covariance matrix in the h-th stratum is:

$$\mathbf{C}_h = \frac{1}{N_h^{(\upsilon)} - 1}\sum_{i\in U_h^{(\upsilon)}} \left(\mathbf{y}_{h,i*}^{(\upsilon)} - \overline{\mathbf{y}}_h\right)^{T}\left(\mathbf{y}_{h,i*}^{(\upsilon)} - \overline{\mathbf{y}}_h\right)$$

where: $\mathbf{y}_{h,i,*}^{(\upsilon)}$ is an i-th row of the matrix $\mathbf{y}_h^{(\upsilon)}$.

The simple sample $s_h^{(\upsilon)}$ of the size $n_h^{(\upsilon)}$ is drawn without replacement from the h-stratum, h = 1, ..., H. Let $S^{(\upsilon)} = \bigcup_{h=1}^{H} s_h^{(\upsilon)}$. The vector of the stratified sample means is as follows:

$$\overline{\mathbf{y}}_{ws^{(\upsilon)}} = \sum_{h=1}^{H} w_h \overline{\mathbf{y}}_{s_h^{(\upsilon)}}$$

where:

$$\overline{\mathbf{y}}_{s_h^{(\upsilon)}} = \frac{1}{n_h^{(\upsilon)}} \sum_{i \in S_h^{(\upsilon)}} \mathbf{y}_{h,i*}^{(\upsilon)} .$$

The variance-covariance matrix of the vector $\overline{\mathbf{y}}_{S^{(\upsilon)}}$ is as follows:

$$\mathbf{V}\left(\overline{\mathbf{y}}_{w,S^{(\upsilon)}}\right) = \sum_{h=1}^{H} \frac{N_h^{(\upsilon)} - n_h^{(\upsilon)}}{n_h^{(\upsilon)} N_h^{(\upsilon)}} w_h^2 \mathbf{C}_h .$$

The sample variance-covariance matrix in an h-stratum is as follows:

$$\mathbf{V}_{S_h^{(\upsilon)}} = \frac{1}{n_h^{(\upsilon)} - 1} \sum_{i \in S_h^{(\upsilon)}} \left(\mathbf{y}_{h,i*}^{(\upsilon)} - \overline{\mathbf{y}}_{S_h^{(\upsilon)}}\right)^T \left(\mathbf{y}_{h,i*}^{(\upsilon)} - \overline{\mathbf{y}}_{S_h^{(\upsilon)}}\right).$$

The unbiased estimator of the matrix $\mathbf{V}\left(\overline{\mathbf{y}}_{wS^{(\upsilon)}}\right)$:

$$\mathbf{V}_{S^{(\upsilon)}}\left(\overline{\mathbf{y}}_{w,S^{(\upsilon)}}\right) = \sum_{h=1}^{H} \frac{N_h^{(\upsilon)} - n_h^{(\upsilon)}}{n_h^{(\upsilon)} N_h^{(\upsilon)}} w_h^2 \mathbf{V}_{S_h^{(\upsilon)}} .$$

Let us assume that $N_h^{(\upsilon)} \to \infty$, $n_h^{(\upsilon)} \to \infty$ and $N_\upsilon - n_\upsilon \to \infty$ as $\upsilon \to \infty$. Moreover, let the arrays $\mathbf{y}_h^{(\upsilon)}$, $h = 1,...,H$ be modified in such a way that the parameters $\overline{\mathbf{y}}_h$, $\mathbf{V}_h$ and w_h are constant for each υ = 1, 2, ... Under the introduced notation, the theorem of Thompson (1997, p. 60), is as follows:

Theorem 3.2. Under the conditions: $\mathbf{V}_{S_h^{(\upsilon)}} \to \mathbf{V}_h$ in probability as $\upsilon \to \infty$ and $\delta_\upsilon^{(h)}(\varepsilon) \to 0$ for any $\varepsilon > 0$, where:

$$\delta_\upsilon^{(h)} = \frac{1}{N_h^{(\upsilon)}} \sum_{\left\{i: i \in U_h^{(\upsilon)}, \left\|\mathbf{y}_{h,i*} - \overline{\mathbf{y}}_h\right\| > \varepsilon n_h^{(\upsilon)} \left(1 - \frac{n_h^{(\upsilon)}}{N_h^{(\upsilon)}}\right)\right\}} \left\|\mathbf{y}_{h,i*} - \overline{\mathbf{y}}_h\right\|$$

and $\|.\|$ is the Euclidian norm, the distribution of the vector statistic:

$$\mathbf{z}_{S_h^{(\upsilon)}} = \left(\bar{\mathbf{y}}_{S_h^{(\upsilon)}} - \bar{\mathbf{y}}_h\right)\sqrt{\frac{N_h^{(\upsilon)} - n_h^{(\upsilon)}}{n_h^{(\upsilon)} N_h^{(\upsilon)}}}$$

approaches the multidimensional normal distribution $N(\mathbf{O}, \mathbf{V}_h)$.

This theorem leads to the following.

Lemma 3.1. When for any h = 1, .., H the assumptions of the theorem 3.2 are fulfilled, the distribution of the vector statistic $\bar{\mathbf{y}}_{wS^{(\upsilon)}}$ approaches the multidimensional normal distribution $N\left(\bar{\mathbf{y}}, \mathbf{V}\left(\bar{\mathbf{y}}_{wS^{(\upsilon)}}\right)\right)$.

Lemma 3.2. The matrix $\mathbf{V}_{S^{(\upsilon)}}^{-1}\left(\bar{\mathbf{y}}_{wS^{(\upsilon)}}\right)$ approaches the matrix $\mathbf{V}^{-1}\left(\bar{\mathbf{y}}_{wS^{(\upsilon)}}\right)$ when $\upsilon \to \infty$.

Proof: $\mathbf{V}_{S^{(\upsilon)}}^{-1}\left(\bar{\mathbf{y}}_{wS^{(\upsilon)}}\right)$ is a continuous function of $\mathbf{V}_{S_h^{(\upsilon)}}$, h=1,...,H, which approaches $\mathbf{V}_h$, respectively, as $\upsilon \to \infty$. This is derived in the following way. Let $c_{*11S}^{(\upsilon)}(i, j)$ be the sample covariance of the i-th and j-th variables. The sample variance-covariance matrix $\mathbf{V}_{S_h^{(\upsilon)}} = \left[c_{*11S}^{(\upsilon)}(i, j)\right]$ approaches $\mathbf{V}_h = \left[c_{*11}(i, j)\right]$ as $\upsilon \to \infty$ when for any $\varepsilon > 0$:

$$\lim_{\upsilon\to\infty} P\left\{\bigcap_{i,j=1}^{k}\left|c_{*11S}^{(\upsilon)}(i, j) - c_{*11}(i, j)\right| \le \varepsilon\right\} = 1 \tag{3.42}$$

On the basis of the well known Bonferroni inequality (see e.g. Miller (1981)) we have:

$$\lim_{\upsilon\to\infty} P\left\{\bigcap_{i,j=1}^{k}\left|c_{*11S}^{(\upsilon)}(i, j) - c_{*11}(i, j)\right| \le \varepsilon\right\} \ge 1 - \sum_{i,j=1}^{k}\lim_{\upsilon\to\infty} P\left\{\left|c_{*11S}^{(\upsilon)}(i, j) - c_{*11}(i, j)\right| > \varepsilon\right\}$$

This and the Chebyshev inequality:

$$P\left\{\left|c_{*11S}^{(\upsilon)}(i, j) - c_{*11}(i, j)\right| > \varepsilon\right\} < \frac{D^2\left(c_{*11S}^{(\upsilon)}(i, j)\right)}{\varepsilon^2}$$

lead to the following:

$$1 \ge \lim_{\upsilon\to\infty} P\left\{\bigcap_{i,j=1}^{k}\left|c_{*11S}^{(\upsilon)}(i, j) - c_{*11}(i, j)\right| \le \varepsilon\right\} \ge 1 - \sum_{i,j=1}^{k}\lim_{\upsilon\to\infty}\frac{D^2\left(c_{*11S}^{(\upsilon)}(i, j)\right)}{\varepsilon^2}.$$

The equation (3.42) is true if for any i,j=1,...,k

$$\lim_{\upsilon\to\infty}\frac{D^2\left(c_{*11S}^{(\upsilon)}(i, j)\right)}{\varepsilon^2} = 0. \tag{3.43}$$

Hence, we have to prove the above expression. Let us simplify the notation of the population and sample covariances in the following way:

$$c_{*11} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}),\ \ \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i,\ \ \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$c_{*11S} = \frac{1}{n-1}\sum_{i=1}^{N}(x_i - \bar{x}_S)(y_i - \bar{y}_S)a_i,\ \bar{x}_S = \frac{1}{n}\sum_{i=1}^{N} x_i a_i,\ \bar{y}_S = \frac{1}{n}\sum_{i=1}^{N} y_i a_i$$

where: $a_i = 1$ if $i \in s$ and $a_i = 0$ if $i \notin s$, $E(a_i) = \frac{n}{N}$,

$$E(a_i a_j) = \frac{n(n-1)}{N(N-1)},\ E(a_i a_j a_h) = \frac{n(n-1)(n-2)}{N(N-1)(N-2)},$$

$$E(a_i a_j a_h a_t) = \frac{n(n-1)(n-2)(n-3)}{N(N-1)(N-2)(N-3)} \quad \text{for } t > h > j > i = 1,\ldots, N.$$

The generality of the following derivation does not depend on the assumption that $\bar{x} = \bar{y} = 0$.

$$E(c_{*11S}) = \frac{n}{n-1}E\left(\frac{1}{n}\sum_{i=1}^{N} x_i y_i a_i - \bar{x}_s \bar{y}_s\right) =$$

$$= \frac{n}{n-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i y_i - \frac{1}{n^2}E\left(\sum_{i=1}^{N} x_i y_i a_i + \sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N} x_i y_j a_i a_j\right)\right) =$$

$$= \frac{n}{n-1}\left(\frac{N-1}{N}c_{*11} - \frac{1}{n}\left(\frac{1}{N}\sum_{i=1}^{n} x_i y_i + \frac{n-1}{N(N-1)}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N} x_i y_j\right)\right) =$$

$$= \frac{n}{n-1}\left(\frac{N-1}{N}c_{*11} - \frac{N-1}{nN}c_{*11} - \frac{n-1}{nN(N-1)}\left(\sum_{i=1}^{N} x_i \sum_{j=1}^{N} y_j - \sum_{i=1}^{N} x_i y_i\right)\right) =$$

$$= \frac{n}{n-1}\left(\frac{N-1}{N}\left(1 - \frac{1}{n}\right)c_{*11} + \frac{n-1}{nN}c_{*11}\right) =$$

$$= \frac{n}{n-1}\left(\frac{(N-1)(n-1)}{nN} + \frac{n-1}{Nn}\right)c_{*11} = \frac{n(n-1)}{(n-1)Nn}(N-1+1)c_{*11} = c_{*11}.$$

Hence, the statistic c_{*11S} is the unbiased estimator of the covariance c_{*11}. Now, we are going to determine the variance of the statistic c_{*11S}. Firstly, we are going to derive the second moment of the estimator c_{*11S}:

$$E\left(c_{*11S}^2\right)=\frac{1}{(n-1)^2}E\left(\sum_{i=1}^{N}x_i y_i a_i\right)^2-\frac{2n}{(n-1)^2}E\left(\bar{x}_s\bar{y}_s\sum_{i=1}^{N}x_i y_i a_i\right)+$$
$$+\left(\frac{n}{n-1}\right)^2 E\left(\bar{x}_s^2\bar{y}_s^2\right). \tag{3.44}$$

$$\frac{1}{(n-1)^2}E\left(\sum_{i=1}^{N}x_i y_i a_i\right)^2=\frac{1}{(n-1)^2}E\left(\sum_{i=1}^{N}x_i^2 y_i^2 a_i+\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}x_i y_i x_j y_j a_i a_j\right)=$$

$$=\frac{n}{(n-1)^2 N}\left(\sum_{i=1}^{N}x_i^2 y_i^2+\frac{n-1}{N-1}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}x_i y_i x_j y_j\right)=$$

$$=\frac{n}{(n-1)^2 N}\left(1-\frac{n-1}{N-1}\right)\sum_{i=1}^{N}x_i^2 y_i^2+\frac{n}{N(N-1)(n-1)}\left(\sum_{i=1}^{N}x_i y_i\right)^2=$$

$$=\frac{n(N-n)}{(n-1)^2 N}c_{*22}+\frac{n(N-1)}{(n-1)N}c_{*11}^2=\frac{1}{n}c_{*22}+c_{*11}^2+\frac{1}{n}c_{*11}^2+0\left(N^{-1}\right)+$$
$$+0\left(n^{-1}N^{-1}\right)+0\left(n^{-2}\right), \tag{3.45}$$

$$\frac{2n}{(n-1)^2}E\left(\bar{x}_s\bar{y}_s\sum_{i=1}^{N}x_i y_i a_i\right)=\frac{2}{n(n-1)^2}\left(\sum_{i=1}^{N}x_i^2 y_i^2 a_i+\right.$$

$$+\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}x_i^2 y_i y_j a_i a_j+\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}x_i y_i^2 x_j a_i a_j+$$

$$\left.+\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}x_i y_i x_j y_j a_i a_j+\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{\substack{t=1\\ t\neq i\\ t\neq j\\ i\neq j}}^{N}x_i y_i x_j y_t a_i a_j a_t\right)=$$

$$= \frac{2}{(n-1)^2}\left(\frac{N-3n+2}{N}c_{*22} + \frac{(n-1)(N-1)}{N}c_{*11}^2 + \frac{(n-1)(n-2)}{N(N-1)(N-2)}\left(2\sum_{i=1}^{N}x_i^2y_i^2 + \right.\right.$$

$$\left.\left. - \sum_{i=1}^{N}x_iy_i^2\sum_{i=1}^{N}x_i - \left(\sum_{i=1}^{N}x_iy_i\right)^2\right)\right) = \frac{2\left(N^2-3Nn+2N+2n^2\right)}{(n-1)^2N(N-2)}c_{*22} +$$

$$+ \frac{2(N-1)(N-n)}{(n-1)N(N-2)}c_{*11}^2 = \frac{2}{n}c_{*11}^2 + 0\left(N^{-1}\right) + 0\left(n^{-1}N^{-1}\right) + 0\left(n^{-2}\right), \qquad (3.46)$$

$$\frac{n^2}{(n-1)^2}E\left(\bar{x}_s^2\bar{y}_s^2\right) =$$

$$= \frac{1}{n(n-1)^2}\left(\frac{1}{N}\sum_{i=1}^{N}x_i^2y_i^2 + \frac{n-1}{N(N-1)}\sum_{i=1}^{N}x_i^2y_i^2 + \frac{2(n-1)}{N(N-1)}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}x_iy_i^2x_j + \right.$$

$$+ \frac{(n-1)(n-2)}{N(N-1)(N-2)}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\sum_{\substack{t=1\\t\neq i\\t\neq j\\j\neq i}}^{N}x_iy_j^2x_t + \frac{2(n-1)}{N(N-1)}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}x_iy_ix_jy_j +$$

$$+ \frac{2(n-1)}{N(N-1)}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}x_i^2y_iy_j + \frac{4(n-1)(n-2)}{N(N-1)(N-2)}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\sum_{\substack{t=1\\t\neq i\\t\neq j\\j\neq i}}^{N}x_iy_iy_jx_t +$$

$$+ \frac{(n-1)(n-2)}{N(N-1)(N-2)}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\sum_{\substack{t=1\\t\neq i\\t\neq j\\j\neq i}}^{N}x_i^2y_jy_t +$$

$$\left. + \frac{(n-1)(n-2)(n-3)}{N(N-1)(N-2)(N-3)}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\sum_{\substack{t=1\\t\neq i\\t\neq j\\j\neq i}}^{N}\sum_{\substack{h=1,h\neq t\\h\neq j,h\neq i\\t\neq i,t\neq j\\j\neq i}}^{N}x_ix_jy_ty_h\right) =$$

$$= \frac{1}{n(n-1)^2N}\left(\left(N-7n+6+\frac{8(n-1)(n-2)}{N-2} - \frac{4(n-1)(n-2)(n-3)}{(N-2)(N-3)}\right)c_{*22} + \right.$$

$$+\frac{(n-1)(N-1)}{N-2}\left(N-2n+2\frac{(n-2)(n-3)}{N-3}\right)c_{*20}c_{*02}+$$

$$+\frac{2(n-1)(N-1)}{N-2}\left(N-2n+2+\frac{(n-2)(n-3)}{N-3}\right)c_{*11}^2\Bigg)=0\left(n^{-2}\right).$$

This and the expressions (3.43)-(3.46) lead to the following:

$$E\left(c_{*11S}^2\right)=\frac{1}{(n-1)^2N}\Bigg(\Bigg(n(N-n)-\frac{2\left(N^2-3Nn+2N+2n^2\right)}{N-2}+\frac{N-6n+5}{n}+$$

$$+\frac{8(n-1)(n-2)}{n(N-2)}-\frac{4(n-1)(n-2)(n-3)}{n(N-2)(N-3)}\Bigg)c_{*22}+$$

$$\frac{(n-1)(N-1)}{n(N-2)}\left(N-2n+2+\frac{(n-2)(n-3)}{N-3}\right)c_{*20}c_{*02}\Bigg)+c_{*11}^2+$$

$$+\frac{N-n}{N(n-1)}\left(-\frac{N}{N-2}+\frac{2(N-1)(N-n-1)}{n(n-1)(N-2)}\right)c_{*11}^2=$$

$$=\frac{1}{n}\left(c_{*22}-c_{*11}^2\right)+c_{*11}^2+0\left(N^{-1}\right)+0\left(n^{-1}N^{-1}\right)+0\left(n^{-2}\right).$$

Hence:

$$D^2\left(c_{*11S}\right)=E\left(c_{*11S}^2\right)-c_{*11}^2=\frac{1}{n}\left(c_{*22}-c_{*11}^2\right)+$$

$$+0\left(N^{-1}\right)+0\left(n^{-1}N^{-1}\right)+0\left(n^{-2}\right).$$

This result leads to the expression (3.43).

Theorem 3.3. Under the assumption of the theorem 3.2, the statistic

$$\hat{U}_{S^{(\upsilon)}}=\left(\bar{\mathbf{y}}_{wS^{(\upsilon)}}-\bar{\mathbf{y}}+\boldsymbol{\Delta}\right)\mathbf{V}_{S^{(\upsilon)}}^{-1}\left(\bar{\mathbf{y}}_{wS^{(\upsilon)}}\right)\left(\bar{\mathbf{y}}_{wS^{(\upsilon)}}-\bar{\mathbf{y}}+\boldsymbol{\Delta}\right)^T, \tag{3.47}$$

where: $\boldsymbol{\Delta}$ is a constant row vector, approaches the non-central chi-square distribution $\chi_k^2(\kappa)$ with k-degree of freedom and the non-centrality parameter:

$$\kappa=\boldsymbol{\Delta}\mathbf{V}^{-1}\left(\bar{\mathbf{y}}_{wS^{(\upsilon)}}\right)\boldsymbol{\Delta}^T \tag{3.48}$$

Proof. The lemma 3.2 leads to the conclusion that the statistic $U_{S^{(\upsilon)}}$ approaches the following Wald's statistic (see, e.g. Everitt (1998)) when $\upsilon \to \infty$

$$U_{S^{(\upsilon)}} = \left(\overline{\mathbf{y}}_{S^{(\upsilon)}} - \overline{\mathbf{y}} + \boldsymbol{\Delta}\right)\mathbf{V}^{-1}\left(\overline{\mathbf{y}}_{S^{(\upsilon)}}\right)\left(\overline{\mathbf{y}}_{S^{(\upsilon)}} - \overline{\mathbf{y}} + \boldsymbol{\Delta}\right)^{T}.$$

It is the square form of the vector $\left(\overline{\mathbf{y}}_{S^{(\upsilon)}} - \overline{\mathbf{y}} + \boldsymbol{\Delta}\right)$ which, on the basis of the lemma 3.1, approaches the normal distribution N($\boldsymbol{\Delta}$, **V**), as $\upsilon \to \infty$. Hence, the well known limit theorems, see e.g. Rao (1982), lead to the conclusion that the Wald's statistic $U_{S^{(\upsilon)}}$ approaches the $\chi^2_k(\kappa)$ distribution, as $\upsilon \to \infty$.

We assume that υ takes such a sufficiently large level that the statistic U_S has approximately $\chi^2_k(\kappa)$ distribution. For the sake of simplification, let us introduce the following notation:

$$n_h^{(\upsilon)} = n_h,\ N_h^{(\upsilon)} = N_h,\ \text{for } h = 1,...,H \text{ and } n^{(\upsilon)} = n,\ N^{(\upsilon)} = N,\ S^{(\upsilon)} = S,$$

$$\overline{\mathbf{y}}_{S^{(\upsilon)}} = \overline{\mathbf{y}}_S,\ \hat{U}_{S^{(\upsilon)}} = \hat{U}_S.$$

Let us consider the following hypothesis:

$$H_0 : \boldsymbol{\Delta} = \mathbf{O}, \qquad H_1\ \boldsymbol{\Delta} \neq \mathbf{O}$$

If the hypothesis H_0 is true, the statistic U_S has approximately the central χ^2_k distribution with the k-degree of freedom. In the case when H_1 is true, the statistic U_S has the non-central $\chi^2_k(\kappa)$ distribution with $\kappa > 0$. The non-centrality parameter shown by the expression (3.48) can be rewritten in the following way:

$$\kappa = \boldsymbol{\Delta}\boldsymbol{\Delta}^{T}\boldsymbol{\gamma}\mathbf{V}^{-1}\left(\overline{\mathbf{y}}_S\right)\boldsymbol{\gamma}^{T} \tag{3.49}$$

where:

$$\boldsymbol{\gamma} = \frac{\boldsymbol{\Delta}}{\sqrt{\boldsymbol{\Delta}\boldsymbol{\Delta}^{T}}},\ \boldsymbol{\gamma}\boldsymbol{\gamma}^{T} = 1.$$

If λ is the maximal eigenvalue of the variance-covariance matrix $\mathbf{V}(\overline{\mathbf{y}}_S)$, $\frac{1}{\lambda}$ is the minimal eigenvalue of the matrix $\mathbf{V}^{-1}\left(\overline{\mathbf{y}}_{wS^{(\upsilon)}}\right)$. This and the expression (3.49) lead to the following:

$$\frac{1}{\lambda} = \underset{\boldsymbol{\gamma}\boldsymbol{\gamma}^{T}=1}{\text{minimum}}\left\{\boldsymbol{\gamma}\mathbf{V}^{-1}\left(\overline{\mathbf{y}}_{wS^{(\upsilon)}}\right)\boldsymbol{\gamma}^{T}\right\}, \tag{3.50}$$

$$\kappa \geq \kappa_0 = \frac{\mathbf{\Delta}\mathbf{\Delta}^T}{\lambda} .$$

The power of the test increases when the parameter κ is larger and larger. So, as the maximal eigenvalue λ becomes shorter and shorter, the power of the test becomes bigger and bigger. The maximal eigenvalue λ of the matrix $\mathbf{V}\left(\bar{\mathbf{y}}_{wS^{(v)}}\right)$ is the function of the sample sizes n_1, .., n_H. Hence, we have the optimization problem of determining such sample sizes $\{n_{\#h}\}$ that the function $\lambda(n_{\#1}, ..., n_{\#H}\}$ takes the minimal value under the limited sum of these sizes or the limited level of the following cost function:

$$c\left(n_1,..,n_H\right) = \sum_{h=1}^{H} c_h n_h . \tag{3.51}$$

The unit cost of observation of variable values in an h-stratum is denoted by c_h. More precisely, we have the following optimizing problem:

$$\begin{cases} \lambda\left(n_1,...,n_H\right) = \text{minimum} \\ c\left(n_1,...,n_H\right) \leq c_0 \\ 1 \leq n_h \leq N_h \quad \text{for} \quad h = 1,...,H \end{cases} \tag{3.52}$$

In some situations we expect that the power of the test, under both fixed hypotheses, should be not less than β. This means that $\kappa \geq \kappa_\beta$ and $\kappa_0 = \frac{\mathbf{\Delta}\mathbf{\Delta}^T}{\lambda} \geq \kappa_\beta$. So: $\lambda \leq \frac{\mathbf{\Delta}\mathbf{\Delta}^T}{\kappa_\beta} = \lambda_\beta$. Then, we have the following optimization problem:

$$\begin{cases} c\left(n_1,...,n_H\right) = \text{minimum} \\ \lambda\left(n_1,...,n_H\right) \leq \lambda_\beta \\ 1 \leq n_h \leq N_h \quad \text{for} \quad h = 1,...,H \end{cases} \tag{3.53}$$

Now we determine such sample sizes that the cost function should take a minimal value under the assumed level β of the power of the test of the hypotheses H_0 and H_1.

In order to handle both optimization problems the appropriate numerical methods should be applied, see the unit 3.2.6.

3.2.8. Optimization of sample sizes on the basis of Neyman's location

Neyman (1934) determined the optimal sizes of samples selected from strata in the case when only one variance is minimized. These optimal samples are the solution to the problem (3.17) in the case when only one variance of the i-th variable is minimized, that is when the purpose function has become variance of an i-th variable. If the standard deviation $\sqrt{v_{*h}(y_i)}$ substituted for b_h, h=1,...,H, the optimal sample sizes are shown by the expressions (3.18)-(3.20). Let us denote by $\underline{\mathbf{n}}^{(i)}$ (i=1,...,m) the vectors of the optimal sample sizes determined in such a way. Substituting $\sqrt{v_{*h}(y_i)}$, $\underline{\mathbf{n}}^{(i)}$ for b_h, $\mathbf{n}$, resperctively, in the expression (3.15) we obtain the minimal value of the variance of the i-th estimator which is denoted by f_{1i} ($\underline{\mathbf{n}}^{(i)}$) under the vector $\underline{\mathbf{n}}^{(i)}$.

On the basis of these Neyman's optimal sample solutions $\underline{\mathbf{n}}^{(i)}$ (i=1,...,m), compromise sample sizes can be found. This problem was considered e.g. by: Dalenius (1953, 1957), Geary (1949), Greń (1963, 1964), Kish (1961), Neyman (1934), Mahalanobis (1944), Srikantan (1963). Generally, their results can be treated as solutions obtained on the basis of goal optimization. Almost all the authors considered the following problem:

$$\begin{cases} \sum_{i=1}^{m} \dfrac{1}{e_i} = \text{minimum} \\ k(\mathbf{n}) \le K, \quad \mathbf{J}_H^T < \mathbf{n} \le N\mathbf{w} \end{cases} \tag{3.54}$$

where e_i is the relative efficiency coefficient defined by the expression

$$e_i = \frac{f_1(\underline{\mathbf{n}}^{(i)})}{D^2(\bar{y}_{wiS})}$$

The variance $D^2(\bar{y}_{wiS}, P_w)$ is determined by the equation (3.6). The optimal solution to the problem (3.54) is determined by the equations (3.18)-(3.20) when b_h is defined by

$$b_h = f_1^{-1}(\underline{\mathbf{n}}^{(i)}) \sum_{i=1}^{m} v_{*h}(y_i)$$

Let us show the following problem considered by Greń (1964):

$$\begin{cases} \sum_{h=1}^{H} \sum_{i=1}^{m} (n_h - \underline{n}_h^{(i)})^2 k_h = \text{minimum} \\ k(\mathbf{n}) \le K, \quad \mathbf{o}_h < \mathbf{n} \le N\mathbf{w} \end{cases} \tag{3.55}$$

When $N_h \to \infty$ for each h=1,...,H the compromise solution is as follows:

$$\underline{n}_h = \underline{\overline{n}}_h + \frac{K - \sum_{t=1}^{H} k_t \underline{\overline{n}}_t}{H\bar{k}} \tag{3.56}$$

where:

$$\underline{\overline{n}}_h = \frac{1}{m}\sum_{i=1}^{m} n_h^{(i)}, \qquad \bar{k} = \frac{1}{H}\sum_{h=1}^{H} k_h .$$

Let us note that other problems involving Neyman's optimal sample sizes can be formed and solved on the basis of multipurpose optimization methods.

3.2.9. Optimization of sample sizes on the basis of generalized variance

The generalized variance is a measure of precision of a vector estimation based on the ellipsoid confidence set as it was noted in paragraph 1.5.1. The generalized variance is an increasing function of the volume of such a confidence set. Hence, the sizes of a sample (selected from strata) should be determined in the such a way that the generalized variance of the vector $\bar{\mathbf{y}}_{wS}$ takes the minimal value. This problem was formulated by Dalenius (1953) and can be rewritten as follows:

$$\begin{cases} f(\mathbf{n}) = \text{minimum} \\ k(\mathbf{n}) \le K, \quad \mathbf{o}_h < \mathbf{n} \le N\mathbf{w} \end{cases} \tag{3.57}$$

where: the cost function is defined by the expression (3.13) and

$$f(\mathbf{n}) = \det \mathbf{V}(\bar{\mathbf{y}}_{wS}, P_w).$$

The iteration method was used to determine the solution to this problem by Ghosh (1958). Greń and Koźniewska (1964) proved that Ghosh's solution is convergent on the true solution but in the case when H=m=2 and $\mathrm{Cov}(\bar{\mathbf{y}}_{w1S}, \bar{\mathbf{y}}_{w2S}, P_w) = 0$. Arwanitis and Afonia (1971) adapted the gradient method to solve the problem (3.57) for m<4 and H>2.

Lemat 3.3 [Wywiał (1989, 1992)]: If at least one intra-stratum variance-covariance matrix $\mathbf{C}_{*h}$ (h=1,..,H) is positive definite, the generalized variance f(**n**) is a strictly convex function in the set D_b for sampling without replacement or in the set D_z for sampling with replacement, where:

$$D_b = \{\mathbf{n}: \mathbf{o}_H < 2\mathbf{n} < N\mathbf{w}\}$$

$$D_z = \{\mathbf{n}: \mathbf{n} > \mathbf{o}_H\}$$

Wywiał (1992) showed that the vector of optimal sample sizes is a solution to the following set of equations:

$$\begin{cases} \dfrac{1}{k_h}\dfrac{\partial f(\mathbf{n})}{\partial n_h} = \dfrac{1}{k_h}\dfrac{\partial f(\mathbf{n})}{\partial n_H}, h = 1,...,H-1 \\ k(\mathbf{n}) = K \end{cases} \tag{3.58}$$

where[21] :

$$\frac{\partial f(\mathbf{n})}{\partial n_h} = -\frac{w_h^2}{n_h^2}\sum_{i=1}^{m}\det\begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,m} \\ \cdot & \cdot & \cdots & \cdot \\ v_{h,i-1,1} & v_{h,i-1,2} & \cdots & v_{h,i-1,m} \\ c_{h,i,1} & c_{h,i,2} & \cdots & c_{h,i,m} \\ v_{h,i+1,1} & v_{h,i+1,2} & \cdots & v_{h,i+1,m} \\ \cdot & \cdot & \cdots & \cdot \\ v_{m,1} & v_{m,2} & \cdots & v_{m,n} \end{bmatrix}$$

where the elements of the matrices $\mathbf{V}(\bar{\mathbf{y}}_{wS}, P'_w)$ and $\mathbf{C}_h$ are denoted by v_{ij} and $c_{hij} = c_h(y_i, y_j)$, respectively.

Let us suppose that the admissible level f_o of a generalized variance $f(\mathbf{n})$ is determined. Then, we can find such sizes of samples selected from strata that the cost function $k(\mathbf{n})$ takes a minimal value. This problem can be written as follows:

$$\begin{cases} k(\mathbf{n}) = \text{minimum} \\ f(\mathbf{n}) \le f_o, \quad \mathbf{o}_h < \mathbf{n} \end{cases} \tag{3.59}$$

Wywiał (1992) showed that the solution to the following system of equation

$$\frac{1}{mf_o}\frac{\partial f(\underline{\mathbf{n}})}{\partial n_h} = \frac{k_h}{k(\underline{\mathbf{n}})}, \qquad h = 1,...,H$$

is the vector of optimal sample sizes.

[21] See e.g. Kubik and Krupowicz (1982), p. 445-446.

3.3. Stratification of population on the basis of auxiliary variables

The vector of stratified sample means is usually a more accurate estimator of population averages than the vector of simple sample means. Precision of the vector of the stratified samples is good when the intra-stratum spread of variables is short. Hence, a population should be divided into such strata that intra-stratum spread of variables under study is short as possible. In a one-dimensional case, the values of a variable under study are divided into such subsets (strata) that the intra-stratum variance takes the value as low as possible. This approach was considered and developed e.g. by: Cochran (1961, 1963), Dalenius (1957), Dalenius and Gurney (1951), Hess, Sethi i Balakrishnan (1966), Jonin, Jonina and Zhuravlev (1978), Serfling (1968). Let us add that Kish (1965) presented practical rules which should lead to good stratification of a population. When dividing a population into strata is not possible or too expensive, so called stratification after selecting a sample is considered.

3.3.1. Non-dominated partitions of population into strata

Let us consider the proportional location of the sample in strata. So, $n_h=nw_h$ and $N_h=Nw_h$ for all h=1,...,H where n_h is the size of a simple sample s_h drawn without replacement from h stratum. Under this assumption the vector of variances of the stratified sample means is as follows:

$$\mathbf{d}(\mathbf{a}) = \begin{bmatrix} d_1(\mathbf{a}) \\ ... \\ d_m(\mathbf{a}) \end{bmatrix} \tag{3.60}$$

where:

$$d_i(\mathbf{a}) = D^2\left(\bar{y}_{wSi} \mid \mathbf{a}\right) = \frac{N-n}{Nn}\sum_{h=1}^{H} w_h(\mathbf{a}) v_{*h}(y_i \mid \mathbf{a}) \text{ , i=1,...,m.} \tag{3.61}$$

$\mathbf{a} = \{U_1,..., U_H\}$ is a partition of a population U into a set of strata. Let us assume that $nw_h>1$ for each h=1,...,H. Let A be a set of admissible partitions of a population. Our purpose is determining such a partition $\underline{\mathbf{a}} \in A$ that $\mathbf{d}(\underline{\mathbf{a}})$=minimum. The set of non-dominated partitions, denoted by $\underline{A} \subset A$, is a solution to this problem. Partitions $\mathbf{a},\mathbf{b} \in \underline{A}$ if it is not true that $d(\mathbf{a}) > d(\mathbf{b})$ and $d(\mathbf{a}) < d(\mathbf{b})$ and $d(\mathbf{a}) < d(\mathbf{c})$ and $d(\mathbf{b}) < d(\mathbf{c})$ for $\mathbf{c} \in A\text{-}\underline{A}$. Usually, the set $\underline{A}$ includes more than one partition. In order to select the only one reasonable partition $\mathbf{a}_*$ from the set $\underline{A}$ the additional optimization criterion should be constructed. For instance, the sum $d(\mathbf{a}) = \sum_{h=1}^{m} d_i(\mathbf{a})$, $\mathbf{a} \in \underline{A}$ can be minimized.

Let us remember that the variance covariance matrix of the simple sample (drawn without replacement) was denoted by $\mathbf{V}(\bar{\mathbf{y}}_S) = \frac{N-n}{Nn}\mathbf{C}_*$. The variance-covariance matrix of the vector of stratified sample mean is denoted by $\mathbf{V}(\bar{\mathbf{y}}_{wS})$ and expressed by the form (3.11) in the case when the samples are drawn without replacement from strata and their sizes are proportional to the sizes of appropriate strata. We can postulate that such a partition of a population into strata should be evaluated that the stratified sample vector is better than the vector of the simple sample. It means that we postulate that such a set G of partitions of a population into strata is formed that the matrix

$$F(\mathbf{a}) = \mathbf{V}(\bar{\mathbf{y}}_S) - \mathbf{V}(\bar{\mathbf{y}}_{wS} \mid \mathbf{a}), \text{ for } \mathbf{a} \in G \tag{3.62}$$

is positive definite. The variance-covariance matrix of the vector of means drawn from the partition **a** of a population into strata is denoted by $\mathbf{V}(\bar{\mathbf{y}}_{wS} \mid \mathbf{a})$. Now, we can look for the set $\underline{G} \subseteq G$ of non-dominated partitions determined by the maximization of the following criterion function:

$$\boldsymbol{\lambda}(\mathbf{a}) = \begin{bmatrix} \lambda_1(\mathbf{a}) \\ \ldots \\ \lambda_m(\mathbf{a}) \end{bmatrix} \tag{3.63}$$

where $\boldsymbol{\lambda}(\mathbf{a})$ is the vector of all eigenvalues of the matrix $\mathbf{F}(\mathbf{a})$ and $\lambda_1 \geq \lambda_{2\geq\ldots\geq} \lambda_m$. Usually, the set $\underline{G}$ consists of more than one partition. In this case the new super-criterion function should be constructed in order to obtain the unique partition. For instance, such functions can be determined by

$$\lambda_*(\mathbf{a}) = \sum_{h=1}^{m} \lambda_i(\mathbf{a}) = \mathrm{tr}\mathbf{F}(\mathbf{a}), \qquad \lambda_{**}(\mathbf{a}) = \lambda_1(\mathbf{a})$$

Determination of the set of all non-dominated partitions is rather difficult even in the case of not too large population. That is why some clustering methods can be suggested in order to stratify a population. Some modification of the well known k-means method presented in the paragraph 3.7 can be adapted for this purpose.

3.3.2. Stratification of population through minimization of the spectral radius of the variance-covariance matrix of auxiliary variables

Let us consider the regression superpopulation model defined in the unit 1.2 by the probability distribution of the random vector $\mathbf{Y}=[Y_1 \ldots Y_N]$, where Y_k is attached to the k-th element of a population $\Omega=\{1,\ldots,N\}$. It has the following properties:

$$Y_k=\beta_0+\boldsymbol{\beta}\mathbf{x}_k^T+Z_k, \qquad k=1,...,N \tag{3.64}$$

where: $\mathbf{x}_k=[x_{k1}...x_{km}]$ is the vector of the values of the fixed auxiliary variables observed on the k-th population element. The vector of regression parameters is denoted by $\boldsymbol{\beta}=[\beta_1... \beta_m]$. The operators of the expected value, variance and covariance calculated on the basis of the distribution function which defines a superpopulation model are denoted by the symbols $E(.)$, $D^2(.)$, $Cov(.)$. Let us introduce the following assumptions:

$$\begin{cases} E(Y_k)=\mu_k=\beta_0+\boldsymbol{\beta}\mathbf{x}_k^T \\ E(Z_k),\ D^2(Y_k)=D^2(Z_k)=\sigma^2 \\ Cov(Y_k,Y_t)=Cov(Z_k,Z_t)=0 \end{cases} \tag{3.65}$$

where: $k\neq l=1,...,N$.

Let us assume that the population Ω is partitioned into non-empty and disjoint strata Ω_h, h=1,...H.

The predictor $\hat{Y}_{wS}$ of a value of the population mean $\overline{Y}=\frac{1}{N}\sum_{k=1}^{N}Y_k$ is as follows:

$$\hat{Y}_{wS}=\sum_{h=1}^{H}w_h\overline{Y}_{S_h} \tag{3.66}$$

where:

$$\overline{Y}_{S_h}=\frac{1}{n_h}\sum_{k\in S_h}Y_k\ .$$

The strategy $(\hat{Y}_{wS},P_p)$ is p-ξ unbiased predictor of the mean $\overline{Y}$. Anderson, Kish and Cornell (1980) derived:

$$E(V_{*h})=\beta\mathbf{C}_{*h}\beta^T+\sigma^2. \tag{3.67}$$

where:

$$V_{*h}=\frac{1}{N_h-1}\sum_{k\in\Omega_h}(Y_k-\overline{Y}_h),\ \overline{Y}_h=\frac{1}{N_h}\sum_{k\in\Omega_h}Y_k$$

This and the expression (3.11) lead to the following one:

$$ED^2(\hat{Y}_{wS},P_p)=\frac{N-n}{Nn}[\boldsymbol{\beta}\mathbf{C}_w\boldsymbol{\beta}^T+\sigma^2] \tag{3.68}$$

where: $\mathbf{C}_w=\sum_{h=1}^{H}w_h\mathbf{C}_{*h}$ is an intra-stratum variance-covariance matrix of the auxiliary variables. Let ρ_w be the maximal eigenvalue (the spectral radius) of the matrix $\mathbf{C}_w$. Then:

$$\rho_w = \underset{\alpha\alpha^T=1}{\text{maximum}}\left\{\boldsymbol{\alpha}\,\mathbf{C}_w\boldsymbol{\alpha}^T\right\}, \quad \text{where:} \quad \boldsymbol{\alpha} = \frac{\boldsymbol{\beta}}{\sqrt{\boldsymbol{\beta}\boldsymbol{\beta}^T}}$$

$$ED^2\left(\hat{Y}_{wS}\right) \le \frac{N-n}{Nn}[\boldsymbol{\beta}\boldsymbol{\beta}^T\rho_w + \sigma^2] = M(\rho_w) \tag{3.69}$$

Hence, the parameter ρ_w can be the criterion function of the cluster method used to stratify the population Ω into strata of the same size. The strata were determined in such a way that the maximal eigenvalue of the intra-stratum variance-covariance matrix of auxiliary variables takes a minimal value. The clustering algorithm can be an adaptation of the well-known Ward method. That is why this clustering algorithm will be called the modified Ward method. Let us note that the order method of Ward leads to such strata that the trace of the intra-stratum variance-covariance matrix of auxiliary variables takes a minimal value.

Anderson, Kish and Cornell (1980) compared the $ED^2(\hat{Y}_{wS}, P_p)$ with the mean square prediction errors of the mean $\overline{Y}$ from a simple sample of size n. They stratified the population through partition values of each auxiliary variable into the class boundaries by the well-known rule of Dalenius and Hodges (1959) and then they explored how the efficiency of stratification depended on the correlation coefficients between the auxiliary variables and the main variable.

Let the matrix $\mathbf{C}_w=\mathbf{C}_w(\mathbf{X})$ be the intra-stratum variance-covariance matrix for the partition of a population into the strata Ω_1, ..., Ω_H represented by the block matrix $\mathbf{X} = [X_1, ..., X_H]$ of the auxiliary variables observations. The sub-matrix X_h of dimension $m\times N_h$ consists of the N_h observations of the m-dimensional auxiliary variable. Let $\rho_w=\rho_w(\mathbf{X})$ be the spectral radius of the matrix $\mathbf{C}_w(\mathbf{X})$ for the partition $\mathbf{X}$.

On the basis of the expression (3.69) we infer that $M(\rho_w(\mathbf{X}))$ is minimal if the function $\rho_w(\mathbf{X})$ is minimal. Then it leads to the following optimization problem:

$$\rho_w\left(\underline{\mathbf{X}}\right) = \min_{\mathbf{X}\in\aleph}\operatorname{imum}\{\rho_w(\mathbf{X})\} \tag{3.70}$$

where: $\aleph$ is the set of block matrices of type $\mathbf{X}$ and each of them represents the admissible partition of a population into strata.

Hence, the optimal stratification is equivalent to the determination of such a partition of the population represented by the block matrix $\underline{\mathbf{X}}$, which minimizes the spectral radius $\rho_w(\mathbf{X})$ of the intra-stratum matrix of variance--covariance $\mathbf{C}_w(\mathbf{X})$ in the set $\aleph$.

Let us note that the partition $\underline{\mathbf{X}}$ is optimal in the case when a vector of means of variables under study is estimated and variables are described by the regression models dependent on the same set of auxiliary variables.

It is difficult to find the solution of the optimization problem directly through minimization of $\rho_w(\mathbf{X})$ because the size of the admissible solution set $\aleph$ is very large. That is why the well-known agglomeration clustering method of Ward was adapted in order to form the partitions of a population into strata. When the number of the algorithm stage increases, the quantity of groups (strata) decreases. At each stage of the algorithm, groups are joined in such a way that the spectral radius of the intra-stratum matrix of variance-covariance achieves a minimum value. Before proceeding with the clustering algorithm a population is treated as a collection of one-element groups. At each stage of the algorithm there are three ways of forming a new cluster. Two elements of a population can be clustered into one group. A population element can be joined to a multi-element group. It is possible to join two multi-element groups.

Let us suppose that the following block matrix represents a collection of groups resulting from the t-th stage of the algorithm:

$$\underline{\mathbf{X}}(t) = \{\underline{\mathbf{X}}_h(t) : h \in \boldsymbol{\eta}(t)\}, \qquad t = 0, 1, ..., N-1 \tag{3.71}$$

where: $\eta(t) = \{h: 1 \le h \le N\}$ is the set of $(N - t)$ indices of the clusters of size $N_h(t) \ge 1$ represented by the submatrices $X_h(t)$ of dimension $m \times N_h(t)$ of auxiliary variables data. The block matrix $\underline{\mathbf{X}}(t)$ consists of $(N - t)$ submatrices.

Let $\mathbf{X}_k(t+1)$ be a submatrix representing a new group formed at the $(t+1)$-th stage of the algorithm, then $\mathbf{X}_k(t+1) = [\underline{\mathbf{X}}_i(t) \cup \underline{\mathbf{X}}_j(t)]$ for $k = \min\{i, j\}$. Then $\eta(t+1) = \eta(t) - \max\{i,j\}$. Denoting by $\mathbf{X}_h(t+1) = \underline{\mathbf{X}}_h(t)$ the remaining submatrices for $h \in \eta(t)$-i-j, we obtain the admissible set of groups for the $(t+1)$-th stage represented by the following block matrix:

$$\mathbf{X}(t+1) = \{\mathbf{X}_h(t+1) : h \in \eta(t+1)\}.$$

Let $\aleph(t+1)$ be the set of all admissible matrices of type $\mathbf{X}(t+1)$. For all $\mathbf{X}(t+1) \in \aleph(t+1)$ the increment of the spectral radius will be denoted by:

$$d(\mathbf{X}(t+1) = \rho_w(\mathbf{X}(t+1)) - \rho_w(\underline{\mathbf{X}}(t)). \tag{3.72}$$

Finally, we select an optimal partition of the population represented by such a matrix $\underline{\mathbf{X}}(t+1)$ that:

$$d\left(\underline{\mathbf{X}}(t+1)\right) = \min_{\mathbf{X}(t+1) \in \aleph(t+1)} \operatorname{imum} \{d(\mathbf{X}(t+1))\}. \tag{3.73}$$

The algorithm will be completed at the stage number $t = N - 1$ when due to the agglomeration process only one cluster equal to the population is left.

It can be proved almost immediately that if $m = 1$, then the expression (3.72) can be reduced to the form:

$$d(\mathbf{X}(t+1)) = N_i N_j \left(N_i + N_j\right)^{-1} \left(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\right)^T \left(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\right)$$

where: $\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j$ are the vectors of means of variables of the joint groups represented by the submatrices $\mathbf{X}_i$, $\mathbf{X}_j$, respectively. Hence, d(**X**(t+1)) becomes the well-known clustering criterion proposed by Ward (1963). Therefore, the Ward's choice rule of optimal population partition can also be applied in this case. The partition represented by $\underline{\mathbf{X}}(g)$ and obtained at the g-th stage of the algorithm is optimal if it fulfils the following expression:

$$d\left(\underline{\mathbf{X}}(g+1)\right) = \max_{t=1,2,\ldots} \text{imum}\left\{d(\underline{\mathbf{X}}(t+1))\right\}$$

Then the partition represented by the block matrix $\underline{\mathbf{X}}(g)$ is chosen as an optimal one if the increment of the criterion function reaches the maximal value at the next stage of the algorithm.

Let us note that the problem of the optimal stratification has been formulated by Dalenius (1950). A given population is partitioned into strata on the basis of clustering outcomes of a multidimensional auxiliary variable. The problem has been developed by many statisticians. Among others, Singh (1971), Anderson, Kish, Cornell (1980), Wywial (1991b, 1995a), Skibicki and Wywiał. (2001) have used an auxiliary variable for the stratification. Schneaberger and Pollot (1985) have obtained an optimal division of values of a two-dimensional normal random variable. Bracha (1991) has presented the survey of stratifying methods and their modifications while Thomsen (1976) has compared several methods of stratification.

Wywial (1998) considered an example of stratification of a population by means of the three methods defined above. Data consist of Swedish municipalities. They were published by Sarndal, Swensson and Wretman (1992). We consider two auxiliary variables p75 - the size of the

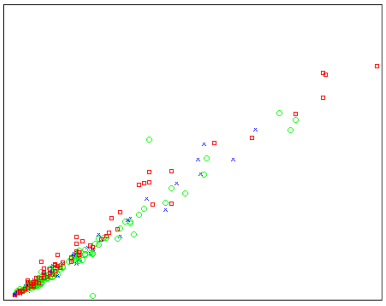


Figure 3.1. The geographic. strat.

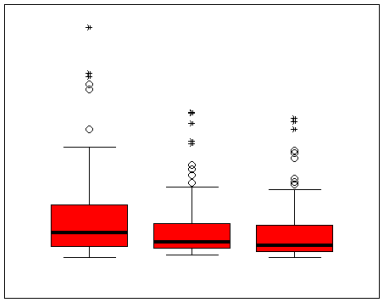


Figure 3.2. The geographic. strat.

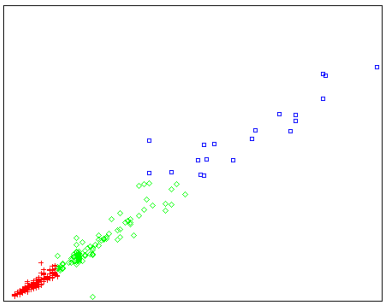


Figure 3.3. The meth. of Ward.

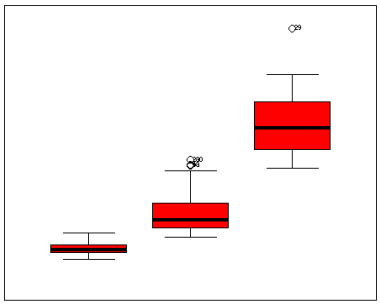


Figure 3.4. The meth. of Ward.

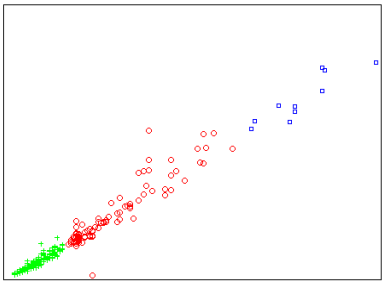


Figure 3.5. The mod. meth. of Ward.

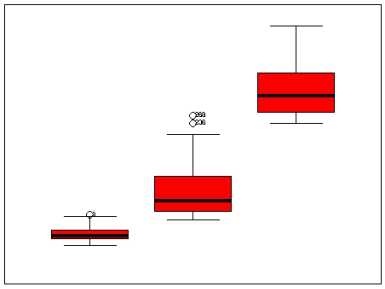


Figure 3.6. The mod. meth. of Ward.

population in 1975 (in thousands), ME84 - number of municipal employees in 1984 and the variable under study: RMT85 - revenues from the 1985 municipal taxation (in millions of kronor). The number of data is 284. After studying the distribution of the data, three non-typical observations of the variables were found. Their values were too large. They create the first stratum and they all have to be sampled. The rest of the data were clustered into three strata by means of the three mentioned. Firstly, the strata were determined according to the geographical regions of Sweden. Next, these strata were obtained by means of the clustering method of Ward on the basis of the auxiliary variables. The third stratification was developed by means of the modified method of Ward. The scatters of the two auxiliary variables in the strata obtained by means of these three methods are represented by figures 3.1, 3.3 and 3.5 respectively. The distributions in the particular strata of the variable (under study) RMT85 are represented by the figures 3.2, 3.4 and 3.6 according to the methods of stratification. The analysis of the figures leads to the conclusion that the sets of strata which have resulted from the two Ward methods are better than geographical stratification because the strata means obtained by the Ward methods are not homogenous and the intra-strata distribution are not spread too much.

Table 3.1

The accuracy comparisons

The method of stratification	The optimal location of the samples			The optimal values of the variances	The ratio of the variances to the variance in the case of the stratification:	
	I	II	III		the geograph.	the modified Ward
1	2	3	4	5	6	7
The geographical	10	12	8	1085	-	-
The order method of Ward	9	15	6	124	0.114	0.544
The modified method of Ward	10	18	2	228	0.210	-

The optimal location of a sample of size 30 was considered on the basis of the standardized auxiliary variables. The size of a sample drawn from a stratum was proportionate to the product of the stratum fraction and the square root of the trace of the stratum variance-covariance matrix of the standardized auxiliary variables. Next, the variance of the stratified sample mean of the RMT85 variable was computed for each location of a sample, separately. The results are shown in the table 3.1. The optimal location of the samples in the strata obtained by means of Ward methods leads to a shorter variance of the estimator than the optimal location in the geographical strata. Moreover, in this sense the order method of Ward leads to a better estimation accuracy of the population mean of the RMT85 variable than the modified method of Ward.

3.3.3. On stratification of population in order to optimize the sample sizes in the case of estimation of mean vector

Let $\Omega=\{1,...,N\}$ be a fixed and identifiable population. It is partitioned into H strata: Ω_h, where: $\bigcup_{h=1}^{H}\Omega_h = \Omega$ and $\Omega_h \cap \Omega_k = \varnothing$ for $k \neq h = 1,...,H$. The size of an h-th stratum is denoted by N_h and $N = \sum_{h=1}^{H} N_h$. Let s_h be a simple sample of size n_h drawn without replacement from an h-th stratum and let $s = \bigcup_{h=1}^{H} s_h$. The considered population parameters are as follows:

$$\bar{y}_i = \frac{1}{N}\sum_{k\in\Omega} y_{ik}\ ,\ \bar{y}_{ih} = \frac{1}{N_h}\sum_{k\in\Omega_h} y_{ik}\ ,\ v_{ih} = \frac{1}{N_h - 1}\sum_{k\in\Omega_h}\left(y_{ik} - \bar{y}_{ih}\right)^2 .$$

As it is well-known, the unbiased estimator of the average $\bar{y}_i$ (i=1,...,m) is the following statistic:

$$\bar{y}_{is} = \sum_{h=1}^{H} w_h \bar{y}_{is_h} \tag{3.74}$$

where:

$$w_h = \frac{N_h}{N},\ \bar{y}_{is_h} = \frac{1}{n_h}\sum_{k\in s_h} y_{ik}$$

Its variance is as follows:

$$D_i^2 = D^2\left(\bar{y}_{is_h}\right) = \sum_{h=1}^{H} w_h^2 \frac{N_h - n_h}{N_h n_h} v_{ih} \tag{3.75}$$

The problem is how to determine the sizes n_h, h = 1, ..., H, in such a way that the sum $n = \sum_{h=1}^{H} n_h$ takes the minimal value if variances of stratified sample means are fixed. This is explained by the following expression:

$$\begin{cases} \sum_{h=1}^{H} n_h = \text{minimum} \\ D_i^2 \le e_i^2,\ i = 1,...,m \\ 2 \le n_h \le N_h,\ h = 1,...,H \end{cases} \tag{3.76}$$

Let us assume that $x_h = \frac{1}{n_h}$, for h = 1, ..., H. This leads to the following equivalent problem (see e.g. Kokan and Khan (1967)):

$$\begin{cases} f\left(x_1,...,x_H\right) = \sum_{h=1}^{H} \frac{1}{x_h} = \text{minimum} \\ \sum_{h=1}^{H} a_{ih} x_h \le b_i,\ \ i = 1,...,m \\ \frac{1}{N_h} \le x_h \le 0.5,\ \ h = 1,...,H \end{cases} \tag{3.77}$$

where:

$$a_{ih} = w_h^2 v_{ih}, \quad b_i = e_i^2 + \sum_{h=1}^{H} \frac{w_h^2 v_{ih}}{N_h} \tag{3.78}$$

Wywiał (2000a) noted that the conditions, defined by the inequalities in the expression (3.77), determine the simplex R consisting of the admissible solutions to the problem. The purpose function $f(x_1, ..., x_H)$ is strictly convex and decreasing in the simplex R. Hence, the solution to the problem is determined by the co-ordinates of the point lying on one of the walls of the simplex.

The shape of the simplex R depends on the fractions $\{w_h\}$ and variances $\{v_{ih}\}$. So, it depends on a partition of a population into strata. Let R_t, R_{t+1} be simplexes and let g_t, g_{t+1} be their volumes. The simplexes are determined by two different partitions of the population into the strata. These partitions are indexed by t and t+1. Moreover, let f_t and f_{t+1} be optimal values of the purpose function of the problem (3.77). If $R_t \subset R_{t+1}$, both simplexes have the same point A which is the origin point of the edges of the simplexes. These edges are parallel to the axis of the co-ordinate system. Hence, if $R_t \subset R_{t+1}$ then $g_t \le g_{t+1}$ and $f_t \ge f_{t+1}$. In this situation Wywiał (2000a) concluded that the strata should be determined in such a way that the volume g_t of the simplex R_t should be as large as possible.

Let us note that it is possible, $g_t > g_{t+1}$, even if $R_t \not\subset R_{t+1}$. In this case the inequality $f_t \ge f_{t+1}$ can be true but not necessarily.

Let $\mathbf{u}$ be the matrix of dimension H×k and $\mathbf{A}$ be the column vector of dimension H×1. Let R_t be the simplex spanned on the points whose co-ordinates are elements of the vector $\mathbf{A}$ and the columns of the matrix $\mathbf{u}$. The number of all these vertexes is denoted by k. If m≥H, the simplex R_t can be decomposed (triangularised) into such disjoint simplexes that their sum is equal to the simplex R_t. Let $R_t\left(\mathbf{A}, \mathbf{u}_{j_1,...,j_H}\right)$ be the simplex spanned on the points whose co-ordinates are elements of the vector $\mathbf{A}$ and the columns of the matrix $\mathbf{u}_{j_1...j_H}$. This matrix consists of the columns of the matrix $\mathbf{u}$ identified by the indexes $j_1,...,j_H$. Let us assume that $\{j_1,...,j_H\}$ is such a set of all H-element combinations of the column indexes of the matrix $\mathbf{u}$ that:

$$\begin{cases} R_t\left(\mathbf{A}, \mathbf{u}_{j_1...j_H}\right) \cap R_t\left(\mathbf{A}, \mathbf{u}_{e_1...e_H}\right) = \varnothing \ \text{ for } \left(j_1,..., j_H\right) \neq \left(e_1,..., e_H\right) \\ \bigcup_{\{j_1,...,j_H\}} R_t\left(\mathbf{A}, \mathbf{u}_{j_1...j_H}\right) = R_t(\mathbf{A}, \mathbf{u}) = R_t \end{cases} \tag{3.79}$$

where: m≥H and $k = \mathrm{Card}\left(\{j_1,.., j_H\}\right) \le \binom{m}{H}$. The volume of the simplex $R_t\left(\mathbf{A}, \mathbf{u}_{j_1...j_H}\right)$ is determined by the expression (see e.g. Borsuk (1969)):

$$g_t\left(\mathbf{A},\mathbf{u}_{j_1\ldots j_H}\right)=\left|\det\begin{bmatrix}1 & \mathbf{J}\\ \mathbf{A} & \mathbf{u}_{j_1\ldots j_H}\end{bmatrix}\right|,$$

where each element of the row vector **J** is equal to one. Hence:

$$g_t=\sum_{\{j_1,\ldots,j_H\}} g_t\left(\mathbf{A},\ \mathbf{u}_{j_1\ldots j_H}\right). \tag{3.80}$$

In the case when H=2 and m≥2, the columns of the matrix **u** can be rewritten as follows: $\mathbf{u}_j=\begin{bmatrix}u_{1j}\\ u_{2j}\end{bmatrix}$, $j=1,\ldots,k\le\binom{m}{2}$. Let the columns of the matrix **u** be ordered in such a way that $\mathbf{u}_j \succ \mathbf{u}_{j+1}$ if and only if $u_{1j}\ge u_{1,j+1}$ and $u_{2j}\le u_{2,j+1}$ and at least one of these two inequalities is sharp. Let $\Delta_{i,j}=R\left(\mathbf{A},\ \mathbf{u}_i,\mathbf{u}_j\right)$ be a triangle spanned on three points $\left\{\mathbf{A},\ \mathbf{u}_i,\mathbf{u}_j\right\}$. The triangularisation of the simplex R(**A**,**u**) is as follows:

$$\begin{cases}\Delta_{j-1,j}\cap\Delta_{j,j+1}=\varnothing \text{ for } j=2,\ldots,k-1\\ \bigcup_{j=1}^{k-1}\Delta_{j,j+1}=R(\mathbf{A},\mathbf{u})\end{cases}$$

In order to maximize the determinant g_t, the well-known clustering method of Hartigan (1975), called the k-means clustering method, can be adopted. The starting point of the clustering algorithm is an arbitrary partition of the population into strata. At the t-th stage of the clustering iteration, each element of the population is moved from one stratum to another and the values of the determinant g_t are evaluated. The new partition, obtained in this way is optimal if the determinant takes a maximal value and if $g_t>g_{t-1}$. Let us notice that during the (t+1)-th iteration such a possible partition is admissible if for which the simplex includes the previous optimal simplex (determined in the t-th iteration). The iteration process is continued until the T-th iteration, when there is no new partition that leads to a greater value of the simplex volume than g_T. The next stop rule is assigning the admissible number of the iterations.

The problem considered above can be considered in practice when the census data are available. In this case, it is possible to stratify a population through the maximization of the simplex volume. Next, the optimal sample sizes can be evaluated. The obtained partition of the population into strata as well as the optimal sample sizes can be useful to make up projects of sample surveys of the population before the next census.

Let us note that Skibicki (2002) considered the problem (3.76) in more general case when costs of observation of data are not the same.

3.4. Two-phase sampling for stratification

Wywiał (1996c) considered the following problem. A two-phase sample is drawn from a superpopulation. In the first phase the simple sample is selected without replacement. Next, it is stratified by means of an appropriate cluster method. The stratification is based on auxiliary variables. In the second phase the simple samples are selected without replacement from the just created strata.

In the first phase a simple sample s' of the size n' is drawn without replacement from a fixed population Ω. Values of auxiliary variables are observed in the sample s'. The sample s' is clustered into strata denoted by s'_h , h=1,...,H, of the sizes n_h', respectively. The sample can be stratified by means of a well-known cluster method like Ward's or k-means on the basis of auxiliary variable observations.

Let $S=\{S_1,...S_H\}$ be the sequence of samples, where S_h is a simple sample of the size n_h drawn form the strata s'_h. The value of the population mean $\bar{y}$ is estimated on the basis of the following statistic:

$$\hat{y}_{wS} = \sum_{h=1}^{H} w'_h \, \bar{y}_{S_h} \tag{3.81}$$

where:

$$\bar{y}_{S_h} = \frac{1}{n_h} \sum_{k \in S_h} y_k \,, \qquad w'_h = \frac{n'_h}{n'}, \qquad n' = \sum_{h=1}^{H} n'_h \,.$$

Sarndall, Swenson and Wretman (1992), p. 353, proved that the statistic $\bar{y}_{wS}$ is an unbiased estimator of the mean and they derived the following formula for its variance:

$$D^2(\hat{y}_{wS}) = \frac{N-n'}{Nn'} v_*(y) + E_{S'}\left(\sum_{h=1}^{H} (w'_h)^2 \frac{n'_h - n_h}{n'_h n_h} v_{*S'_h}(y \mid S') \right) \tag{3.82}$$

where:

$$v_*(y) = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})^2 \,, \qquad v_{*S_h}(y|s') = \frac{1}{n'_h - 1} \sum_{k \in S_h} \left(y_k - \bar{y}_{S_h} \right)^2 \,.$$

where $E_{S'}(.)$ is the operator of the expected value calculated on the basis of the probability distribution of the simple random sample S'.

Let us assume that $n'_h = \frac{n'}{H}$ and that $n_h = \frac{n}{H}$ for h=1,...,H. for h=1,...,H. Hence, the variance given by the expression (3.82) takes the following form:

$$D_p^2(\hat{y}_{wS}) = \frac{N-n'}{Nn'} v_*(y) + \frac{1}{H}\frac{n'-n}{n'n} E_{S'}\left(\sum_{h=1}^{H} v_{*S_h}(y|s')\right). \quad (3.83)$$

Let s be the simple sample of size n drawn without replacement from a finite and fixed population. The sample mean is denoted by $\overline{Y}_S = \frac{1}{n}\sum_{k\in S} Y_k$. Its variance is as follows:

$$D^2(\overline{Y}_S) = \frac{N-n}{Nn} v_*(y).$$

The sample variance

$$v_{*S'}(y) = \frac{1}{n'-1}\sum_{k\in S'}\left(y_k - \overline{y}_{S'}\right)^2, \qquad \overline{y}_{S'} = \frac{1}{n}\sum_{k\in S'} y_k$$

can be decomposed in the following way:

$$v_{*S'}(y) = \frac{1}{n'-1}\left[\left(\frac{n'}{H}-1\right)\sum_{h=1}^{H} v_{*S'_h}(y \mid S'=s') + \frac{n'}{H}\sum_{h=1}^{H}\left(\overline{y}_{S'_h} - \overline{y}_{S'}\right)^2\right]. \quad (3.84)$$

Then:

$$\frac{1}{H}\sum_{h=1}^{H} v_{*S_h}(y|S'=s') = \frac{n'-1}{n'-H} v_{*S'}(y) - \frac{n'}{H(n'-H)}\sum_{h=1}^{H}\left(\overline{y}_{S_h} - \overline{y}_{S'}\right)^2.$$

The obtained result and the expression (3.83) lead to the following formula:

$$D_p^2(\hat{y}_{wS}) = \left[\frac{H-1}{n'-H}\frac{n'-n}{n'n} + \frac{N-n}{Nn}\right] v_*(y) - \frac{n'-n}{nH(n'-H)} E_{S'}\left(\sum_{h=1}^{H}(\overline{y}_{S_h} - \overline{y}_{S'})^2\right). \quad (3.85)$$

This result allow us to calculate the following difference:

$$D^2(\overline{y}_S) - D_p^2(\hat{y}_{wS}) = \frac{n'-n}{nH(n'-H)} E_{S'}\left(\sum_{h=1}^{H}(\overline{y}_{S_h} - \overline{y}_{S'})^2\right) - \left[\frac{H-1}{n'-H}\frac{n'-n}{n'n}\right] v_*(y) \quad (3.86)$$

or

$$D^2(\bar{y}_S) - D_p^2(\hat{y}_{wS}) = \frac{1-\frac{n}{n'}}{nH(1-\frac{H}{n'})} E_{S'}\left(\sum_{h=1}^{H}(\bar{y}_{S_h'} - \bar{y}_{S'})^2\right) - \frac{(H-1)\left(1-\frac{n}{n'}\right)}{\left(1-\frac{H}{n'}\right)n'n} v_*(y) \tag{3.87}$$

This result leads us to the following conclusion: If $N\to\infty$ and $n'\to\infty$ and $N-n'>0$, then

$$D^2(\bar{y}_S) - D_p^2(\hat{y}_{wS}) = \frac{1}{nH} E_{S'}\left(\sum_{h=1}^{H}(\bar{y}_{S_h'} - \bar{y}_{S'})^2\right). \tag{3.88}$$

Hence, if the sample size n is fixed and the sample size n' and the population size N are sufficiently large, the estimator $\bar{y}_{wS}$ is not less accurate then the sample mean $\bar{y}_S$.

Let us consider an example of stratification on the basis of the variable under research. A population consists of eight elements (N=8). The following values of a variable *y* are observed: 1,2,3,4,10,11,12,13. The value of the population mean is $\bar{y} = 7.0$ and the value of the variance $v_*(y) = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2$ =24.5714. The sample s' whose size is six elements, is drawn without replacement from the population. Then, the number of different samples is $\binom{8}{6} = 28$. Each sample has been divided into two strata of the same size. The two strata are optimal if the wihin-strata variance takes the minimal value. After some calculations we receive: $\sum_{s'\in S}\sum_{h=1}^{2} v_{*S_h'}(y|s') = 328.8322$ and

$$D_p^2(\hat{y}_{ws}) = \frac{5.8720}{n} + 0.0451$$

where $n=n_1+n_2$ and $n_1=n_2$. In our case $2\le n\le 6$. Moreover:

$D_p^2(\hat{y}_{ws}) = 2.9811$, for n=2, $D_p^2(\hat{y}_{ws}) = 1.5131$, for n=4.

The variance of the simple sample means takes the value:

$$D^2(\bar{y}_S) = \frac{24.5714}{n} - 3.0714.$$

Hence:

$D^2(\bar{y}_S) = 9.2142$, for n=2, $D^2(\bar{y}_S) = 3.0715$, for n=4

Let the ratio $e = \frac{D^2(\hat{y}_{ws})}{D^2(\bar{y}_s)}$ be a relative efficiency coefficient:

e=0.324 for n=2; e=0.493 for n=4.

The variance of the stratified sample mean is less than the variance of the simple sample mean. This result was obtained when the first phase sample was stratified on the basis of the researched variable. In practice, the stratification of the sample should be based on auxiliary variables. Some accuracy analysis of the estimator $\hat{y}_{ws}$ based on simulation method was considered by Wywiał (2002b). From the other point of view the problem is considered in the next paragraph:

3.5. Estimation of population average on the basis of strata formed by means of discrimination functions

Let $\Omega=\{e_1,...,e_N\}$ be a fixed population of size N. The observations of a k-dimensional variable under study are denoted by $\mathbf{y}_i = [y_{i1} \dots y_{ik}]$, i = 1, ..., N. Let $\mathbf{x}_i = [x_{i1} \dots x_{ia}]$, i = 1, ..., N, be an i-th observation of an a-th dimensional auxiliary variable. We assume that the variables under study and auxiliary variables are highly dependent on each other. A simple sample s of size m is drawn without replacement from the population Ω. The sample s is clustered into mutually disjoint and nonempty subsets $s_1, s_2, ..., s_H$. Hence $\bigcup_{h=1}^{H} s_h = s$. The set s_h, h = 1, ..., H, is of the size m_h and $1 \le m_h \le m - H$ and $m = \sum_{h=1}^{H} m_h$. The sets $s_1, s_2, ..., s_H$ are obtained on the basis of the observation of the variables under study by means of an appropriate clustering method. Particularly, we can propose the well known clustering method by Ward or the k-means method. They provide a short intracluster spread of observations of a multidimensional variable.

Let $\mathbf{x}_{s_h} = [\mathbf{x}_{i_h}]$ be the matrix of dimensions m_h x a. The row $\mathbf{x}_{i_h}$ of the matrix $\mathbf{x}_{s_h}$ is the outcome of an a-dimensional auxiliary variable observed in the set s_h, so $i_h \in s_h$. The population Ω will be partitioned into nonempty and mutually disjoint strata (subpopulations) Ω_h, h = 1,...,H, and $\bigcup_{h=1}^{H} \Omega_h = \Omega$. The row $\mathbf{x}_{i_h}$ of the matrix $\mathbf{x}_{\Omega_h}$ is the outcome of an a-dimensional auxiliary variable observed in the stratum Ω_h, so $i_h \in \Omega_h$. The matrices $\mathbf{x}_{s_1}, ..., \mathbf{x}_{s_H}$ can be a base of construction of a discrimination function which divides the population Ω into nonempty and mutually disjoint

strata. Particularly, the well known linear discrimination functions can be obtained when the auxiliary variables have approximately normal multidimensional distributions with the same covariance matrices in the subpopulations, see e.g. Rao (1973). These quite strong assumptions are not necessary in other discrimination methods.

Let d_{ij} be the Euclidean distance between the vectors $\mathbf{x}_i,\mathbf{x}_j \in \mathbf{X}$, where $\mathbf{X}$ is the matrix of auxiliary variable observed in a population Ω. Hence:

$$d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T} \tag{3.89}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are observations of auxiliary variables attached to the population elements i and j respectively, where $i \in \Omega$ and $j \in \Omega$.

Let us assume that $s_h \subseteq \Omega_h$ for all h = 1, ..., H. The distance between a vector $\mathbf{x}_i \in \mathbf{X}$ and a set of vectors $\mathbf{X}_{s_j}$ will be denoted by $\rho(\mathbf{x}_i, \mathbf{X}_{s_j})$. For example:

$$\rho(\mathbf{x}_i, \mathbf{X}_{s_j}) = \frac{1}{m_j} \sum_{p \in s_j} d_{ip} \tag{3.90}$$

$$\rho(\mathbf{x}_i, \mathbf{X}_{s_j}) = \min_{p \in s_j} imum\{d_{ip}\} \tag{3.91}$$

$$\rho(\mathbf{x}_i, \mathbf{X}_{s_j}) = \max_{p \in s_j} imum\{d_{ip}\} \tag{3.92}$$

The criterion function is defined in the following way. We assume that $\rho(\mathbf{x}_i, \mathbf{X}_{s_h}) = 0$, if $i \in s_h$. The i-th element of the population is attached to the t-th stratum if and only if $i \in \Omega_t - s_t$ and

$$\rho(\mathbf{x}_i, \mathbf{X}_{s_t}) = \min_{j=1,...,H} imum\{\rho(\mathbf{x}_i, \mathbf{X}_{s_j})\} \tag{3.93}$$

This discrimination method can be named the conditional k-means method. Its algorithm is as follows. Firstly a population Ω is arbitrarily divided into such subpopulations $\Omega_h^{(o)}$, h = 1, ..., H, that $s_h \subseteq \Omega_h^{(o)}$ for all h = 1, ..., H. Next, the well known algorithm of the k-means method is implemented but under the condition that the subset s_h is the constant object in the subpopulation $\Omega_h^{(t)}$ for each h = 1,...,H and for each iteration index t=0,1,2,....

Let Ω_1, ..., Ω_H be the strata (subpopulations) provided by the conditional k-means method or some other method of discrimination. Hence, $s_h \subseteq \Omega_h$ and the size of the stratum Ω_h is equal to $N_h \geq m_h$ for all h = 1, ..., H and $\sum_{h=1}^{H} N_h = N$. We can expect that the spread of observations of the multidimensional auxiliary variable as well as the spread of observations

of the multidimensional variable under study to be low in the strata Ω_1, ..., Ω_H. More specifically, we can expect that intrastratum variances of the auxiliary variables and the intra-stratum variances of the variables under the study should be small. This suggests drawing the samples from the strata Ω_1, ..., Ω_H.

Let z_h be a simple random sample of the size n_h drawn without replacement from the set Ω_h-s_h, h=1, ..., H. The sum of these samples will be denoted by $z = \bigcup_{h=1}^{H} z_h$. Moreover, let g = (s, z). Hence the sample s can be treated as a pilot sample which lets us divide the population Ω into the strata. Next, the stratified sample z is selected. In another situation, the sample s can be drawn on one occasion and the sample z on another occasion of statistical research.

In order to simplify our study, let us assume that an estimation of the average of only one variable under study is considered. It does not limit the generality of the following results: the population mean is defined by the expression:

$$\bar{y} = \bar{y}_\Omega = \frac{1}{N}\sum_{i\in\Omega} y_i \ .$$

The sample treated as a random set will be denoted by a capital letter, e.g. S. The outcome of the sample S will be denoted by a small letter s. The simple sample mean is as follows:

$$\bar{y}_S = \frac{1}{m}\sum_{i\in S} y_i \ .$$

Let us define the following statistics:

$$\bar{y}_{S_h} = \frac{1}{m_h}\sum_{i\in S_h} y_i \ , \qquad \bar{y}_{\Omega_h - S_h} = \frac{1}{N_h - m_h}\sum_{i\in\Omega_h - S_h} y_i \ ,$$

$$\bar{y}_{\Omega - S} = \frac{1}{N - m}\sum_{i\in\Omega - S} y_i \ , \qquad \bar{y}_{Z_h} = \frac{1}{n_h}\sum_{i\in Z_h} y_i \ .$$

The estimator of population mean is as follows:

$$\hat{y}_G = \alpha\bar{y}_S + (1-\alpha)\sum_{h=1}^{H} w_h \bar{y}_{Z_h} \tag{3.94}$$

where:

$$0<\alpha<1,\ w_h = \frac{N_h - m_h}{N - m} \ .$$

The conditional expected value of the statistic $\hat{y}_G$ is as follows:

$$E_{Z/S}(\hat{y}_G) = \alpha\bar{y}_S + (1-\alpha)\sum_{h=1}^{H} w_h \bar{y}_{\Omega_h - S_h} ,$$

$$E_{Z/S}(\hat{y}_G) = \alpha\bar{y}_S + (1-\alpha)\bar{y}_{\Omega - S} \tag{3.95}$$

or

$$E_{Z/S}(\bar{y}_G) = \frac{N(1-\alpha)}{N-m}\bar{y} + \frac{N\alpha - m}{N-m}\bar{y}_S . \tag{3.96}$$

Particularly

$$E_{Z/S}(\hat{y}_G) = \bar{y}, \text{if}\ \alpha = \frac{m}{N} . \tag{3.97}$$

On the basis of the expressions (3.95) or (3.96) we infer that the unconditional expected value of the estimator is:

$$E(\hat{y}_G) = E_S E_{Z/S}(\hat{y}_G) = \bar{y} . \tag{3.98}$$

The variance of the estimator is derived according to the expression:

$$D^2(\hat{y}_G) = D_S^2(E_{Z/S}(\hat{y}_G)) + E_S(D_{Z/S}^2(\hat{y}_G)) . \tag{3.99}$$

Wywial (1998a) derived the following formula:

$$D^2(\hat{y}_G) = \frac{2(N\alpha - m)^2}{(N-m)Nm} v_* + (1-\alpha)^2 E_S\left(\sum_{h=1}^{H} w_h^2 \frac{N_h - m_h - n_h}{(N_h - m_h)n_h} v_{*\Omega_h - S_h}\right) . \tag{3.100}$$

This expression can be rewritten in the following way:

$$D^2(\hat{y}_G) = \frac{2(N\alpha - m)^2}{(N-m)Nm} v_* + (1-\alpha)E_S\left(\sum_{h=1}^{H} w_h^2 \frac{N_h - m_h - n_h}{(N_h - m_h)n_h}\frac{N_h - 1}{N_h - m_h - 1} v_{*\Omega_h}\right) +$$

$$-(1-\alpha)^2 E_S\left(\sum_{h=1}^{H} w_h^2 \frac{m_h - 1}{N_h - m_h - 1} v_{*S_h}\right) +$$

$$-(1-\alpha)^2 E_S\left(\sum_{h=1}^{H} w_h^2 \frac{m_h N_h}{(N_h - m_h)(N_h - m_h - 1)}\left(\bar{y}_{\Omega_h} - \bar{y}_{S_h}\right)^2\right) \tag{3.101}$$

where:

$$v_{*\Omega_h} = \frac{1}{N_h - 1}\sum_{i\in\Omega_h}\left(y_i - \bar{y}_{\Omega_h}\right)^2, \bar{y}_{\Omega_h} = \frac{1}{N_h}\sum_{i\in\Omega_h} y_i .$$

$$v_{*S_h} = \frac{1}{m_h - 1}\sum_{i\in S_h}\left(y_i - \bar{y}_{S_h}\right)^2$$

Particularly, on the basis of the expression (3.100) we have:

$$D^2\left(\hat{y}_G\right) = \frac{(N-m)^2}{N^2} E_S\left(\sum_{h=1}^{H} w_h^2 \frac{N_h - m_h - n_h}{(N_h - m_h)n_h} v_{*\Omega_h - S_h}\right), \text{ if } \alpha = \frac{m}{N}. \tag{3.102}$$

The unbiased estimator of the variance of the estimator $\hat{y}_G$ is as follows:

$$D_G^2\left(\hat{y}_G\right) = \frac{(N\alpha - m)^2}{(N-m)Nm} v_{*S} + (1-\alpha)^2 \sum_{h=1}^{H} w_h^2 \frac{N_h - m_h - n_h}{(N_h - m_h)n_h} v_{*Z_h} \tag{3.103}$$

where:

$$v_{*S} = \frac{1}{m-1}\sum_{i\in S}\left(y_i - \bar{y}_S\right)^2 \qquad v_{*Z_h} = \frac{1}{n_h - 1}\sum_{i\in Z_h}\left(y_i - \bar{y}_{Z_h}\right)^2 .$$

Let us assume the proportional allocation of the samples in the strata. If $n_h = nw_h$ for all h=1,...,H, then $n = \sum_{h=1}^{H} n_h$ and the expression (3.100) can be simplified to the form:

$$D^2\left(\hat{y}_G^{(P)}\right) = \frac{2(N\alpha - m)^2}{(N-m)Nm} v_* + (1-\alpha)^2 \frac{N-m-n}{(N-m)n} E_S\left(\sum_{h=1}^{H} w_h v_{*\Omega_h - S_h}\right). \tag{3.104}$$

When N is large and m is small (that is when $N \to \infty$ and $\frac{m}{N} \to 0$), then

$$D^2\left(\hat{y}_G^{(P)}\right) = \frac{\alpha^2 v_*}{m} + \frac{(1-\alpha)^2}{n} E_S\left(\sum_{h=1}^{H} w_h v_{*\Omega_h - S_h}\right). \tag{3.105}$$

Moreover, if $N \to \infty$ and $\alpha = \frac{m}{N} \to 0$, then

$$D^2\left(\hat{y}_G^{(P)}\right) = \frac{1}{n} E_S\left(\sum_{h=1}^{H} w_h v_{*\Omega_h - S_h}\right). \tag{3.106}$$

We are going to compare the variance of the simple sample mean and the variance of the estimator $\hat{y}_G$ in an asymptotic situation under the proportional allocation of the samples to the strata and under the assumption that $\alpha = \frac{m}{N}$. Let S_0 be a simple random sample of the size m+n drawn without replacement from a population U. Hence, the simple sample mean takes the form:

$$\bar{y}_{S_0} = \frac{1}{m+n} \sum_{i \in S_0} y_i . \tag{3.107}$$

It is the unbiased estimator of a population and

$$D^2\left(\bar{y}_{S_0}\right) = \frac{N-m-n}{N(m+n)} v_* \tag{3.108}$$

Wywiał (1998a) derived the following expression:

$$D^2\left(\bar{y}_G^{(P)}\right) = \frac{1}{n}(v_* - r) \tag{3.109}$$

where:

$$r = E_S\left(\bar{y}_{\Omega - S} - \bar{y}\right)^2 + E_S\left(\sum_{h=1}^{H} w_h \left(\bar{y}_{U_h - S_h} - \bar{y}_{U-S}\right)^2\right). \tag{3.110}$$

This result and the expression (3.108) under the assumptions that $N \to \infty$, $\alpha = \frac{m}{N} \to 0$ lead to the following expression:

$$D^2\left(\bar{y}_{S_0}\right) - D^2\left(\hat{y}_G^{(P)}\right) = \frac{r}{n} - \frac{m}{n(m+n)} v_* . \tag{3.111}$$

If we additionally assume that the size m is very small in comparison to the sample size n and the variance v_*, we have

$$D^2\left(\overline{y}_{S_0}\right) - D^2\left(\hat{y}_G^{(P)}\right) \approx \frac{r}{n} \geq 0 . \tag{3.112}$$

Hence, under the stated assumptions, the estimator $\hat{y}_G^{(P)}$ is not less precise than the simple sample mean $\overline{y}_{S_0}$.

The construction of the estimator $\hat{y}_G$ is strictly connected with the process of dividing a population into strata by means of a discrimination function. That is why the estimator $\hat{y}_G$ can be named a discrimination estimator of a population average.

Let us note that the statistic $N\hat{y}_G$ is the unbiased estimator of a population total.

3.6. Stratification of population after sample selection

In survey sampling, conditional methods are usually connected with post-stratification estimators for domains and with inference on the basis of regression models or contingency tables. These problems were considered e.g. by Rao (1985), Tillé (1998, 1999), Williams (1962). The problem of stratification of a population, on the basis of observations of a variable under study in a sample, was considered by e.g. Dalenius (1957).

Wywial (2001) deals with the problem of an appropriate division of a simple sample into subsamples. This partition leads to clustering a population into subpopulations. Each of these subpopulations includes one and only one previously created subsample. The linear combinations of statistics from the subsamples are used for the estimation of a population mean. The subsample means and the regression estimators from the subsamples are considered to be these statistics. The coefficients of this linear combination are proportionate to the sizes of the subpopulations. The construction of the estimators depends on the methods of clustering the sample into subsamples and the population into subpopulations. Bias and variances of a certain estimator have been derived but the precision of others should be studied by means of some simulation methods. An example of such a simulation study is presented. Moreover, some generalizations of proposed estimators have been suggested.

3.6.1 Basic notation

Let us assume that the values of an auxiliary variable are known in a fixed and finite population of size N. A k-th value of the auxiliary variable is denoted by x_k, k=1,...,N. The simple sample s of the size n is drawn without replacement from a population U. An i-th value of a variable

under study is denoted by y_i, i=1,...,N. Moreover, let us assume that the elements of the population U={1,...,N} are ordered in such a way that $x_i<x_j$ for each i<j and i=1,...,N and j=1,...,N.

Let us divide each sample $s=\{i_i,...,i_m,i_{m+1},...,i_n\}$, where $i_j < i_h$, into two following subsamples $s_1(k)=\{i_1,...,i_m\}$ and $s_2(k)=\{i_{m+1},...,i_n\}$, where $k=i_m$ and k=1,2,...,H<N. The integer $k=i_m$ is a function of observations $\{x_{i_1},...,x_{i_n}\}$ of the auxiliary variable in the sample s. Hence, $s_1(k)\cap s_2(k)=\varnothing$ and $s_1(k)\cup s_2(k)=s$. The sizes of the subsamples $s_1(k)$ and $s_2(k)$ are denoted by $n_1(k)$ and $n_2(k)$, respectively. Let $U_1(k)=\{i: x_i \le x_k\}$ and $U_2(k)=\{i: x_i > x_k\}$, k =1,..., H. Hence, $U_1(k)\cap U_2(k) = \varnothing$ and $U_1(k)\cup U_2(k)=U$, k = 1, ..., H. The sizes of the subpopulations $U_1(k)$ and $U_2(k)$ are denoted by $N_1(k)$ and $N_2(k)$, respectively. Similarly, the fractions of the elements in these subpopulations are denoted by: $w_{s_1(k)} = \frac{N_1(k)}{N}$ and $w_{s_2(k)} = \frac{N_2(k)}{N}$.

Wywiał (2001) considered the following conditional estimator of the population mean $\bar{y}$:

$$\tilde{\bar{y}}_{S/K} = w_{S_1(K)}\bar{y}_{S_1(K)} + w_{S_2(K)}\bar{y}_{S_2(K)} \tag{3.113}$$

where:

$$\bar{y}_{S_h(k)} = \frac{1}{n_h(k)}\sum_{i\in S_h(k)} y_i, \quad h = 1, 2. \tag{3.114}$$

Let s, $s_1(k)$ and $s_2(k)$ be outcomes of random samples S, $S_1(K)$ and $S_2(K)$, respectively. Let the sampling design of the sample s be denoted by P(S=s) or simply by P(s) for $s \in \mathbf{S}$, where **S** is the sampling space. The sampling design P(s) can be rewritten as follows:

$$P(s)=P(s_1(k), s_2(k), K=k)=P(s_1(k), s_2(k)|\, K=k)P(K=k)=P(s|\, K=k)P(K=k).$$

This enables us to write the expected value of the estimator in the following way:

$$E(\tilde{\bar{y}}_{S/K}) = \sum_{k=1}^{N} E_{S/k}(\bar{y}_{S/k})P(K=k) \tag{3.115}$$

where:

$$E_{S/k}(\bar{y}_{S/k}) = E_{S/k}(w_{S_1(k)}\bar{y}_{S_1(k)}) + E_{S/k}(w_{S_2(k)}\bar{y}_{S_2(k)}). \tag{3.116}$$

The conditional expected value $E_{S/K}(.)$ and the conditional variance $D^2_{S/K}(.)$ are determined on the basis of the conditional sampling design P(s|K=k). The variance is as follows:

$$D^2(\bar{y}_{S/K}) = D^2_K(E_{S/K}(\tilde{y}_{S/K})) + E_K(D^2_{S/K}(\tilde{y}_{S/K})). \tag{3.117}$$

If $E_{S/k}(\tilde{y}_{S/k}) = \bar{y}$ for each k=1,...,H the conditional statistic $\tilde{y}_{S/k}$ is the unbiased estimator of the population average $\bar{y}$ and

$$D^2_K(E_{S/K}(\tilde{y}_{S/K})) = 0. \tag{3.118}$$

We are going to show some particular forms of the estimator $\tilde{y}_{S/K}$. Each of them can be obtained through determining a function which provides the value k (of the random variable K) which constitutes the border between the subsamples $s_1(k)$ and $s_2(k)$. This function may depend on the auxiliary variable.

3.6.2. Randomly divided sample

Let us assume that all values of an auxiliary variable are distinct, so: $x_1<x_2<...x_N$. Let m = 1, ..., n-1 be a number of observations of the auxiliary variable which are less than or equal to the value x_k in sample s. Then, $m=n_1(k)$ is the size of the sub-sample $s_1(k)$. The size $n_2(k)$ of the sub-sample $s_2(k)$ is equal to n–m. Let us introduce the following notation:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad \bar{x}_{s_h(k)} = \frac{1}{n_h(k)}\sum_{i\in s_h(k)} x_i, \qquad h=1,2$$

$$v_x = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2, \quad v_{xs_h(k)} = \frac{1}{n_h(k)-1}\sum_{i\in s_h(k)}(x_i-\bar{x}_{s_h(k)})^2,$$

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i, \quad \bar{y}_{s_h(k)} = \frac{1}{n_h(k)}\sum_{i\in s_h(k)} y_i,$$

$$v_y = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})^2, \quad v_{ys_h(k)} = \frac{1}{n_h(k)-1}\sum_{i\in s_h(k)}(y_i-\bar{y}_{s_h(k)})^2,$$

$$v_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y}),$$

$$v_{xys_h(k)} = \frac{1}{n_h(k)-1} \sum_{i \in s_h(k)} \left(x_i - \bar{x}_{s_h(k)}\right)\left(y_i - \bar{y}_{s_h(k)}\right),$$

$$\bar{y}_s = \frac{1}{n} \sum_{i \in s} x_i, \quad \bar{x}_s = \frac{1}{n} \sum_{i \in s} x_i .$$

The non-decreasing sequence $\left(x_{i_1},...,x_{i_n}\right)$ is observed in the particular simple sample s. Let the value x_k be chosen randomly from the elements of the sequence $\left(x_{i_1},...,x_{i_{n-1}}\right)$. Hence, the probability of selecting the k-th value of an auxiliary variable from the sequence $\left(x_{i_1},...,x_{i_{n-1}}\right)$ is equal to $\frac{1}{n-1}$. Hence, the number of the divisions of each sample s of the size n into two subsamples $s_1(k)$ and $s_2(k)$ is equal to n–1. The number of all possible divisions of all possible simple samples s of the size n is equal to $(n-1)\binom{N}{n} = c^{-1}$.

The number of all such samples s of the size n that $s = s_1(k) \cup s_2(k)$ and $s_1(k)$ is of the fixed size m is equal to $\binom{k-1}{m-1}\binom{N-k}{n-m}$. Under the conditions: $N-k \geq n-m$, $k \leq N-n+1 \leq N-n+m$ the number of all the samples $s \in \Omega$ that $s = s_1(k) \cup s_2(k)$ is equal to

$$A_1(k) = \sum_{m=1}^{n-1} \binom{k-1}{m-1}\binom{N-k}{n-m}. \tag{3.119}$$

Hence, if $n-1 \leq k \leq N-n+1$, then

$$P_1(K=k) = cA_1(k) . \tag{3.120}$$

Let us note that

$$P_1(K=k, M=m) = c\binom{k-1}{m-1}\binom{N-k}{n-m}, \tag{3.121}$$

$$P(s_1(k), s_2(k), K=k, M=m) = \frac{1}{(n-1)\binom{N}{n}} = c,$$

$$P(s_1(k), s_2(k) \mid K = k, M = m) = \frac{1}{\binom{k-1}{m-1}\binom{N-k}{n-m}}, \tag{3.122}$$

$$\begin{cases} P(s_1(k) \mid K = k, M = m) = \dfrac{1}{\binom{k-1}{m-1}} \\ P(s_2(k) \mid K = k, M = m) = \dfrac{1}{\binom{N-k}{n-m}} \end{cases} \tag{3.123}$$

So,

$$P(s_1(k), s_2(k) \mid K = k, M = m) =$$

$$= P(s_1(k) \mid K = k, M = m) P(s_2(k) \mid K = k, M = m). \tag{3.124}$$

Using the well known Cauchy's formula (see e.g. Flachsmeyer (1977) or Lipski and Marek (1986)) we have:

$$P_1(K = k) = \frac{n}{N(n-1)}\left[1 - (N-k)\frac{\binom{k-1}{n-1}}{\binom{N-1}{n-1}}\right].$$

Similarly, we can derive that if $1 \le k \le n-2$, then

$$P_1(K = k) = cA_2(k) \tag{3.125}$$

where:

$$A_2(k) = \sum_{m=1}^{k}\binom{k-1}{m-1}\binom{N-k}{n-m}, \tag{3.126}$$

or $$A_2(k) = \sum_{r=0}^{k-1}\binom{k-1}{r}\binom{N-k}{n-1-r}.$$

If $N - n + 2 \le k \le N - 1$,

$$P_1(K = k) = cA_3(k) \tag{3.127}$$

where:

$$A_3(k) = \sum_{m=2}^{n-1} \binom{k-1}{m-1} \binom{N-k}{n-m}. \tag{3.128}$$

Finally, the expressions (3.120), (3.125) and (3.127) lead to the following one:

$$P_1(K=k) = \begin{cases} cA_2(k) & \text{for} \quad 1 \le k \le n-2 \\ cA_1(k) & \text{for} \quad n-1 \le k \le N-n+1 \\ cA_3(k) & \text{for} \quad N-n+2 \le k \le N-1 \end{cases} \tag{3.129}$$

Particularly, if n = 2,

$$P_1(K=k) = (N-k)\binom{N}{2}^{-1} = \frac{2(N-k)}{N(N-1)}, \quad k = 1,\ldots, N-1. \tag{3.130}$$

On the basis of the expression (3.113) the conditional estimator $\tilde{\bar{y}}^{(1)}_{S/K}$ takes the following particular form:

$$\tilde{\bar{y}}^{(1)}_{S/K} = \frac{K}{N}\bar{y}_{S_1(K)} + \left(1 - \frac{K}{N}\right)\bar{y}_{S_2(K)}. \tag{3.120}$$

The expressions (3.113) – (3.118) and (3.123) – (3.129) let us derive

$$E\left(\tilde{\bar{y}}^{(1)}_{S/K}\right) = \sum_{k=1}^{N-1} E_{S/k}\left(\tilde{\bar{y}}^{(1)}_{S/k}\right) P_1(K=k) =$$

$$= \sum_{k=1}^{N-1} \sum_{m=1}^{\min\{k,n-1\}} E_{S/m,k}\left(\tilde{\bar{y}}^{(1)}_{S/k}\right) P_1(K=k, M=m) \tag{3.132}$$

where $P_1(K=k,M=m)$ is given by the expression (3.121).
Let us introduce the following notation:

$$\bar{y}_{U_h(k)} = \frac{1}{N_h(k)} \sum_{i \in U_h(k)} y_i, \quad h = 1, 2. \tag{3.133}$$

Hence:

$$E_{S/m,k}\left(\tilde{\bar{y}}^{(1)}_{S/k}\right) = \frac{k}{N} E_{S_1(k)/m,k}\left(\bar{y}_{S_1(k)}\right) + \frac{N-k}{N} E_{S_2(k)/m,k}\left(\bar{y}_{S_2(k)}\right) = \bar{y}.$$

This and the result of the derivation (3.132) lead to the following expression:

$$E_{S/m.k}\left(\tilde{\bar{y}}^{(1)}_{S/k}\right)=\bar{y},\ E_{S/k}\left(\tilde{\bar{y}}^{(1)}_{S/k}\right)=\bar{y},$$

$$E\left(\tilde{\bar{y}}^{(1)}_{S/k}\right)=E_K E_{S/K}\left(\tilde{\bar{y}}^{(1)}_{S/k}\right)=\bar{y}. \qquad (3.134)$$

This means that the statistic $\tilde{\bar{y}}^{(1)}_{S/K}$ is a conditionally as well as unconditionally unbiased estimator of the mean $\bar{y}$.

The variance is as follows:

$$D^2\left(\tilde{\bar{y}}^{(1)}_{S/K}\right)=$$

$$=\sum_{k=1}^{N-1}\sum_{m=1}^{\min\{k,n-1\}}\left(\left(\frac{k}{N}\right)^2 D^2_{S/m,k}\left(\bar{y}_{S_1(k)}\right)+\left(\frac{N-k}{N}\right)^2 D^2_{S/m,k}\left(\bar{y}_{S_2(k)}\right)\right)P_1(K=k,M=m)=$$

$$=\sum_{k=1}^{N-1}\sum_{m=1}^{\min\{k,n-1\}}\left(\left(\frac{k}{N}\right)^2\frac{k-m}{km}v_{U_1(k)}+\right.$$

$$\left.+\left(\frac{N-k}{N}\right)^2 v_{U_2(k)}\frac{N-k-n+m}{(N-k)(n-m)}\right)P_1(K=k,M=m) \qquad (3.135)$$

where:

$$v_{U_h(k)}=\frac{1}{k-m-1}\sum_{k\in U_h(k)}\left(y_i-\bar{y}_{U_h(k)}\right)^2,\ h=1,2. \qquad (3.136)$$

The parameter $\bar{y}_{U_h(k)}$ is defined by the expression (3.133).

The result (3.135) can be rewritten in the following way:

$$D^2\left(\tilde{\bar{y}}^{(1)}_{S/K}\right)=\sum_{k=1}^{N-1}\left(\frac{k}{N}\right)^2 v_{U_1(k)}\sum_{m=1}^{\min\{k,n-1\}}\left(\frac{1}{m}-\frac{1}{k}\right)P_1(K=k,M=m)+$$

$$+\sum_{k=1}^{N-1}\left(1-\frac{k}{N}\right)^2 v_{U_2(k)}\sum_{m=1}^{\min\{k,n-1\}}\left(\frac{1}{n-m}-\frac{1}{N-k}\right)P_1(K=k,M=m). \qquad (3.137)$$

It is easy to prove that the unbiased estimator of the conditional variance $D^2_{S/m,k}\left(\tilde{\bar{y}}^{(1)}_{S/k}\right)$ is as follows:

$$d^2_{S/m,k}\left(\tilde{\bar{y}}^{(1)}_{S/K}\right)=\frac{1}{N^2}\left(\frac{k(k-m)}{m}v_{S_1(k)}+\frac{(N-k)(N-k-n+m)}{(n-m)}v_{S_2(k)}\right). \tag{3.138}$$

where:

$$v_{S_1(k)}=\frac{1}{m-1}\sum_{i\in S_1(k)}\left(y_i-\bar{y}_{S_1(k)}\right)^2,\quad \bar{y}_{S_1(k)}=\frac{1}{m}\sum_{i\in S_1(k)}y_i\ , \tag{3.139}$$

$$v_{S_2(k)}=\frac{1}{n-m-1}\sum_{i\in S_2(k)}\left(\bar{y}_{S_2(k)}-y_i\right)^2,\quad \bar{y}_{S_2(k)}=\frac{1}{n-m}\sum_{i\in S_2(k)}y_i\ . \tag{3.140}$$

3.6.3. Conditional weighed mean from the sample divided by median into sub-samples

Let us assume that the simple sample s is drawn without replacement and it is of the size $n < N$. Moreover, let $s = \{i_1, ..., i_n\}$ and $x_{i_j} < x_{i_e}$ and $i_j < i_e$ if and only if $j < e$. The sample s is divided into two subsamples $s_1(k) = \{i_1, ..., k\}$ and $s_2(k) = s - s_1(k)$. Let us assume that $k = i_m$ if $n = 2m$ and $k=i_{m+1}$ if $n = 2m+1$. Hence, x_k is a value of the sample median of the auxiliary variable. The number k identifies the position of the sample median in the population.

Wilks (1962), p. 243, considered the distribution of the order statistic in the simple sample drawn without replacement from a finite population. The probability distribution of the random variable K is a particular case of this distribution. If $m \geq 1$ and $n = 2m < N$:

$$P_2(K=k)=\frac{\binom{k-1}{m-1}\binom{N-k}{m}}{\binom{N}{2m}},\quad k=m,...,N-m\ ,$$

(3.141)

$$E(K)=\frac{m(N+1)}{2m+1},\qquad D^2(K)=\frac{m(N+1)(N-2m)(m+1)}{2(2m+1)^2(m+1)}.$$

If $m \geq 1$ and $n = 2m+1 < N$:

$$P_2(K=k) = \frac{\binom{k-1}{m}\binom{N-k}{m}}{\binom{N}{2m+1}}, \quad k = m+1,\ldots,N-m, \qquad (3.142)$$

$$E(K) = \frac{N+1}{2}, \qquad D^2(K) = \frac{(N+1)(N-2m+1)}{8m+12}.$$

Particularly if m=1 and n=2, the distribution is reduced to one determined by the equation (3.130). If m=1 and n=3, then

$$P_2(K=k) = \frac{6(k-1)(N-k)}{(N-2)(N-1)N}, \quad k = 2,\ldots,N-1.$$

The sampling design of the sample s can be showed in the following way:

$$P(s) = \binom{N}{n}^{-1} = P(s_1(k) \mid K=k)P(s_2(k) \mid K=k)P_2(K=k) \qquad (3.143)$$

where: $P(s_1(k) \mid K=k) = \binom{k-1}{m-1}^{-1}$ in the case when n=2m,

$P(s_1(k) \mid K=k) = \binom{k-1}{m}^{-1}$ in the case when n=2m+1,

$P(s_2(k) \mid K=k) = \binom{N-k}{m}^{-1}$.

Let us consider the following conditional estimator of the population average $\bar{y}$:

$$\tilde{\bar{y}}^{(2)}_{S/K} = \frac{K-1}{N}\bar{y}_{S_1(K-1)} + \frac{1}{N}y_K + \left(1-\frac{K}{N}\right)\bar{y}_{S_2(K)} \qquad (3.144)$$

where: $s_1(K-1) = s_1(K) - \{K\}$,

$$\begin{cases} \overline{y}_{S_1(K-1)} = \dfrac{1}{m_*} \sum\limits_{i \in S_1(K-1)} y_i \\ \overline{y}_{S_2(K)} = \dfrac{1}{m} \sum\limits_{i \in S_2(K)} y_i \end{cases} \tag{3.145}$$

where: $m_*=m-1$ if $n = 2m$ and $m_*=m$ if $n =2m+1$.

The expected value of this statistic is derived in the following way:

$$E\left(\tilde{y}^{(2)}_{S/K}\right) = E_K\left(E_{S/K}\left(\tilde{y}^{(2)}_{S/K}\right)\right) =$$

$$= E_K\left(\frac{K-1}{N} E_{S/K}\left(\overline{y}_{S_1(K-1)}\right) + \frac{1}{N} y_K + \left(1-\frac{K}{N}\right) E_{S/K}\left(\overline{y}_{S_2(K)}\right)\right) =$$

$$= E_K\left(\frac{K-1}{N} \overline{y}_{U_1(K-1)} + \frac{1}{N} y_K + \left(1-\frac{K}{N}\right) \overline{y}_{U_2(K)}\right) = E_K\left(\overline{y}\right) = \overline{y}\,.$$

where: $U_1(K-1) = U_1(K) - \{K\}$.

Hence:

$$\begin{cases} E_{S/k}\left(\tilde{y}^{(2)}_{S/k}\right) = \overline{y} \\ E\left(\tilde{y}^{(2)}_{S/K}\right) = \overline{y} \end{cases} \tag{3.146}$$

In conclusion, the statistic $\tilde{y}^{(2)}_{S/K}$ is a conditionally and unconditionally unbiased estimator of a population mean.

The property (3.143) enables us prove that

$$Cov_{S/K}\left(\overline{y}_{S_1(K-1)}, \overline{y}_{S_2(K)}\right) = 0\,.$$

This and the expressions (3.117) and (3.118) lead to the following expression:

$$D^2\left(\tilde{y}^{(2)}_{S/K}\right) =$$

$$= E_K\left(\frac{(K-1)(K-m-1)}{N^2 m} v_{U_1(K-1)} + \frac{(N-K)(N-K-m)}{N^2 m} v_{U_2(K)}\right)$$

or:

$$D^2\left(\tilde{y}^{(2)}_{S/K}\right) = \frac{1}{N^2 m} E_K\Big((K-1)(K-m-1)v_{U_1(K-1)} +$$

$$+ (N-k)(N-K-m)v_{U_2(K)}\Big). \tag{3.147}$$

The unbiased estimator of the variance $D^2\left(\tilde{\bar{y}}^{(2)}_{S/k}\right)$ is showed by the equation:

$$d_S^2\left(\tilde{\bar{y}}^{(2)}_{S/k}\right)=\frac{1}{mN^2}\left(k(k-m-1)v_{S_1(k-1)}+(N-k)(N-k-M)v_{S_2(k)}\right) \qquad (3.148)$$

where:

$$\begin{cases} v_{S_1(k-1)}=\dfrac{1}{m_*-1}\sum\limits_{i\in S_1(k-1)}\left(y_i-\bar{y}_{S_1(k-1)}\right)^2 \\ v_{S_2(k)}=\dfrac{1}{m-1}\sum\limits_{i\in S_2(k)}\left(y_i-\bar{y}_{S_2(k)}\right)^2 \end{cases} \qquad (3.149)$$

where: $m_*=m-1$ if $n=2m$ and $m_*=m$ if $n=2m+1$.

3.6.4. Conditional weighed regression estimators from the sample divided by median into sub-samples

Let us keep all the assumptions leading to the distribution P_2 derived in the previous paragraph. Instead of the estimator $\tilde{\bar{y}}^{(2)}_{S/X}$ we define the following one

$$\tilde{\bar{y}}^{(3)}_{S/K}=\frac{K-1}{N}\bar{y}^{(R)}_{S_1(K-1)}+\frac{1}{N}y_K+\left(1-\frac{K}{N}\right)\bar{y}^{(R)}_{S_2(K)} \qquad (3.150)$$

where:

$$\bar{y}^{(R)}_{S_1(K-1)}=\bar{y}_{S_1(K-1)}+b_{S_1(K-1)}\left(\bar{x}_{U_1(K-1)}-\bar{x}_{S_1(K-1)}\right), \qquad (3.151)$$

$$\bar{y}^{(R)}_{S_2(K)}=\bar{y}_{S_2(K)}+b_{S_2(K)}\left(\bar{x}_{U_2(K)}-\bar{x}_{S_2(K)}\right), \qquad (3.152)$$

$$b_{S_1(K-1)}=\frac{c_{xyS_1(K-1)}}{v_{xS_1(K-1)}},\quad b_{S_2(K)}=\frac{c_{xyS_2(K)}}{v_{xS_2(K)}}, \qquad (3.153)$$

$$c_{xyS_h(L)}=\frac{1}{m_*-1}\sum_{i\in S_h(L)}\left(x_i-\bar{x}_{S_h(L)}\right)\left(y_i-\bar{y}_{S_h(L)}\right),$$

$$v_{xS_h(L)}=\frac{1}{m_*-1}\sum_{i\in S_h(L)}\left(x_i-\bar{x}_{S_h(L)}\right)^2.$$

L=K–1 if h=1 or L=K if h=2. $m_* = m - 1$ if $n = 2m$ and $h = 1$ or $m_* = m$ if n=2m and h=2 or m_*=m if $n = 2m+1$. Similarly to the previous paragraph, we can derive the following approximate expression for the variance of the statistic $\tilde{y}^{(3)}_{S/X}$.

$$D^2\left(\tilde{y}^{(3)}_{S/K}\right) \approx \frac{1}{N^2 m} E_K\left((K-1)(K-m-1)v_{U_1(K-1)}\left(1-r^2_{U_1(K-1)}\right)+\right.$$

$$\left.+(N-K)(N-K-M)v_{U_2(K)}\left(1-r^2_{U_2(K)}\right)\right) \qquad (3.154)$$

where:

$$v_{yU_h(L)} = v_{U_h(L)} = \frac{1}{N_{U_h(L)}-1}\sum_{i\in U_h(L)}\left(y_i - \bar{y}_{U_h(L)}\right)^2,$$

$$c_{xyU_h(L)} = \frac{1}{N_{U_h(L)}-1}\sum_{i\in U_h(L)}\left(y_i - \bar{y}_{U_h(L)}\right)\left(x_i - \bar{x}_{U_h(L)}\right),$$

$$r_{U_h(L)} = \frac{c_{xyU_h(L)}}{\sqrt{v_{yU_h(L)}\ v_{xU_h(L)}}}.$$

To estimate the variance of the statistic $\tilde{y}^{(3)}_{S/k}$ we can use the following statistic:

$$d^{(3)}_S\left(\tilde{y}_{S/k}\right) = \frac{1}{mN^2}\left(k(k-m-1)v_{S_1(k-1)}\left(1-r^2_{S_1(k-1)}\right)+\right.$$

$$\left.+(N-k)(N-k-m)v_{S_2(k)}\left(1-r^2_{S_2(k)}\right)\right) \qquad (3.155)$$

where:

$$r_{S_h(L)} = \frac{c_{xyS_h(L)}}{\sqrt{v_{xS_h(L)}\ v_{yS_h(L)}}}. \qquad (3.156)$$

3.6.5. Conditional weighed sample mean obtained through stratifying the sample on the basis of an auxiliary variable

Similarly like in the previous paragraph, we assume that the sample $s = \{i_1,...,i_n\}$ and $x_{i_1} < x_{i_2} < ... < x_{i_n}$. The sub-samples are: $s_1(k)=\{i_1,...,i_m\}$ and $s_2(k)=\{i_{m+1}, ..., i_n\}$, where $i_m = k$. Additionally, we assume that $n > 4$.

Let us define the two following criteria of dividing samples into sub-samples:

$$q_1(m,k) = (m-1)v_{S_1(k)} + (n-m-1)v_{S_2(k)}, \tag{3.157}$$

$$q_2(m,k) = \frac{k(k-m)}{m} v_{S_1(k)} + \frac{(N-k)(N-k-n+m)}{(n-m)} v_{S_2(k)}. \tag{3.158}$$

The function q_1 is proportionate to the intra-subsample spread of the variable under study. The $q_2(m, k)$ criterion, considered by Wywiał (2000b), is proportional to the estimator of the variance of stratified sample mean[22]. Let $\underline{m}$ and $\underline{k} = i_{\underline{m}}$ be such parameters that $q_1(\underline{m}, \underline{k})$ = minimum and $s=s_1(\underline{k}) \cup s_2(\underline{k})$. Similarly, let $q_2\left(\hat{m}, \hat{k}\right)$ = minimum and $s = s_1(\hat{k}) \cup s_2(\hat{k})$.

On the basis of such a partition of the sample, we can define the following estimators:

$$\tilde{\bar{y}}^{(4)}_{S/\underline{K}} = \frac{\underline{K}}{N} \bar{y}_{S_1(\underline{K})} + \left(1 - \frac{\underline{K}}{N}\right) \bar{y}_{S_2(\underline{K})}, \tag{3.159}$$

$$\tilde{\bar{y}}^{(5)}_{S/\hat{K}} = \frac{\hat{K}}{N} \bar{y}_{S_1(\hat{K})} + \left(1 - \frac{\hat{K}}{N}\right) \bar{y}_{S_2(\hat{K})}. \tag{3.160}$$

Without any additional assumption we cannot state if these estimators are unbiased or consistent. We can expect that the method of dividing the sample s into sub-samples should lead to small mean square errors of the estimators. This problem will be studied by means of simulation methods.

3.6.6. Example of simulation study of the estimation efficiency

The distribution of 30 observations (x;y) of a two-dimensional variable is shown by the figure 3.1. The basic parameters of these variables in the population consisting of 30 elements are as follows: the average of auxiliary variable $\bar{x}$ =68.6824, the mean of the variable under study $\bar{y}$ =93.6536, the variances of auxiliary variable and the variable under study $v_x=(89.1094)^2$, $v_y=(17.6015)^2$, respectively and finally the correlation coefficient between these variables r=0.9940.

[22] See the idea of minimization of the sample estimator of variance considered e.g. by Lehman (1991).

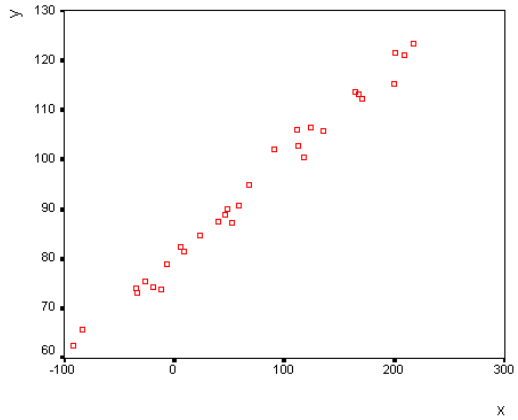


Figure 3.1. The scatter plot for variables x and y in the population

Let the population average be estimated by means of the estimators $\tilde{y}_{S/K}^{(2)}$ and $\tilde{y}_{S/\hat{K}}^{(5)}$ defined in the paragraphs 3.6.3 and 3.6.5, respectively. The simple sample drawn without replacement has 5 elements. On the basis of all these possible samples, the conditional (and the unconditional) expected values and variances of both estimators have been calculated. The variance of the simple sample mean is $D^2(\bar{y}_S)=51.6356$. The statistic $\tilde{y}_{S/K}^{(2)}$ is the unbiased estimator of the population mean but $E(\tilde{y}_{S/\hat{K}}^{(5)}) = 96.0761 \neq \bar{y} = 93.6536$. The absolute value of the bias is 2.59% of the population mean $\bar{y}$. The variances of the estimators are: $D^2(\tilde{y}_{S/K}^{(2)})=42.9823$ and $D^2(\tilde{y}_{S/\hat{K}}^{(5)})=36.4320$.

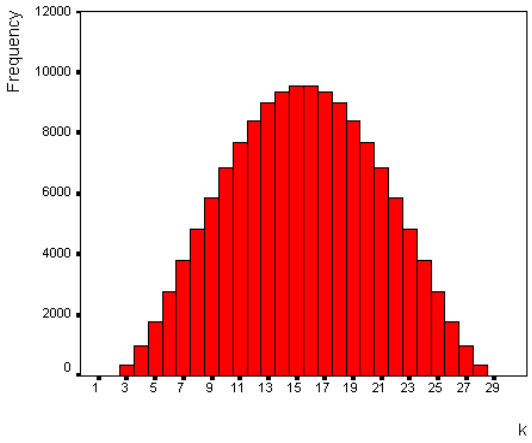


Figure 3.2. The probability distribution of the variable K in the case of the estimator $\tilde{y}_{S/K}^{(2)}$

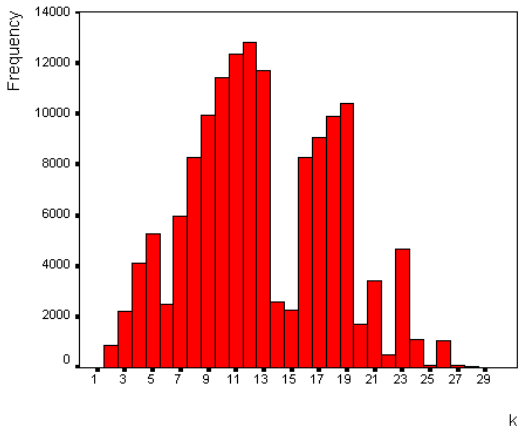


Figure 3.3. The probability distribution of the variable K in the case of the estimator $\tilde{y}^{(5)}_{S/\hat{K}}$

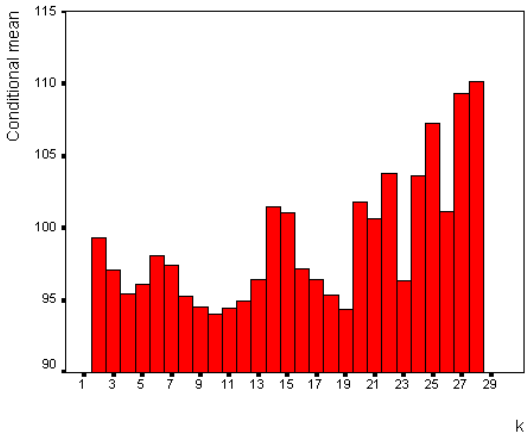


Figure 3.4. The conditional expected values of the estimator $\tilde{y}^{(5)}_{S/\hat{K}}$

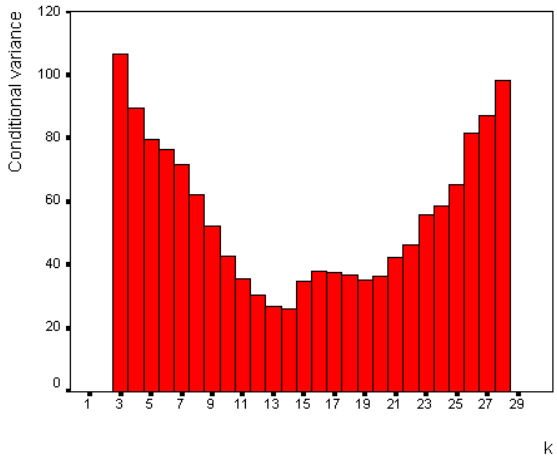


Figure 3.5. The conditional variances of the estimator $\tilde{y}^{(2)}_{S/K}$

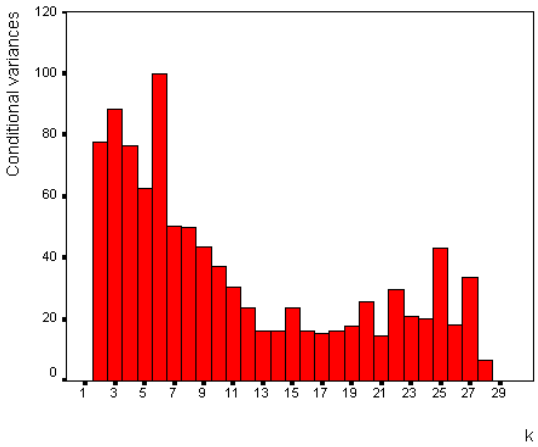


Figure 3.6. The conditional variances of the estimator $\tilde{y}^{(5)}_{S/\hat{K}}$

The relative efficiency is defined by the expression: $e_h = (100\%)D^2\left(\tilde{y}^{(h)}_{S/K}\right)/D^2\left(\bar{y}_S\right)$. In our case: e_2=83.24% and e_5=70.56%. Hence, the precision of conditional estimators $\tilde{y}^{(2)}_{S/K}$ and $\tilde{y}^{(5)}_{S/\hat{K}}$ is better than the precision of the simple sample mean.

As it was defined, the outcome k of the random variable K is the number of the population element dividing the sample into two subsamples. The probability distribution of the random variable K in the case of the estimators $\tilde{y}^{(2)}_{S/K}$ and $\tilde{y}^{(5)}_{S/\hat{K}}$ are presented by the figures 3.2 and 3.3, respectively.

The conditional values of the estimator $\tilde{y}^{(5)}_{S/\hat{K}}$ are showed by the figure 3.4. The conditional variances of the estimators $\tilde{y}^{(2)}_{S/K}$ and $\tilde{y}^{(5)}_{S/\hat{K}}$ are represented by the figures 3.5 and 3.6, respectively.

3.6.7. Some generalizations

The above considered conditional method of estimation can be generalized in several directions. Wywiał (2001, 2002) considered more than two subsamples.

Similarly to the previous paragraphs let us assume that the elements of the population $U = \{1, ..., N\}$ are ordered in such a way that $x_i < x_j$ for each $i < j = 1, ..., N$. The simple sample s of the size n is drawn without replacement from a fixed and finite population U. Let us divide each sample $s=\{i_1,..,i_n\}$, where $i_j < i_h$ if $i < h$, into H following sub-samples of size r_h-r_{h-1}: $s_1(k_1) = \{i_1,...,i_{r_1}\}$, $s_h(k_h) = \{i_{r_{h-1}+1},...,i_{r_h}\}$, where h=1,...,H–1<N, $s_H = \{i_{r_{H-1}+1},...,i_n\}$ and $k_h = i_{r_h}$. Hence, $s_h(k_h) \cap s_t(k_t) = \varnothing$ for each h≠t=1,...,H and $\bigcup_{h=1}^{H} s_h = s$. Let $U_1 = \{i : x_i \le x_{r_1}\}$, $U_h = \{i : x_{r_{h-1}} < x_i \le x_{r_h}\}$, h=1,...,H-1 and $U_H = \{i : x_i > x_{r_{H-1}}\}$. Hence $U_h \cap U_t = \varnothing$ for each h≠t=1,...,H, and $\bigcup_{h=1}^{H} U_h = U$.

Let $K_1,...,K_{H-1}$ be r_1-th,...,r_{H-1}-th order statistics, respectively, in the simple sample drawn without replacement from a finite population. The $k_h = i_{r_h}$ is a possible value of the r_h-th order statistic K_h, h=1,...,H-1. Wilks (1962), p. 252, showed that:

$$P(K_1 = k_1,...,K_{H-1} = k_{H-1}) =$$

$$= \frac{\binom{k_1 - 1}{r_1 - 1}\binom{k_1 - k_2 - 1}{r_2 - r_1 - 1}...\binom{k_{H-2} - k_{H-1} - 1}{r_{H-2} - r_{H-1} - 1}\binom{N - k_{H-1}}{n - r_{H-1}}}{\binom{N}{n}} \qquad (3.161)$$

where: $1 \le k_1 < k_2 ... k_{H-1} < N-1$ or

$$P(K_1 = k_1,...,K_{H-1} = k_{H-1}) = \frac{\binom{N - k_{H-1}}{n - r_{H-1}}\prod_{h=1}^{H-1}\binom{k_h - k_{h-1} - 1}{r_h - r_{h-1} - 1}}{\binom{N}{n}}$$

where $k_0=0$, $k_H=N$ and $r_0=0$.

Let us note (see Fisz (1963, 1967)) that the r_i-th order statistic K_i is the sample quantile of order $\lambda_i \in (0;1)$, if $r_i = [n\lambda_i] + 1$.

We can show that the unbiased estimator of the population mean is the following statistic:

$$\tilde{y}_{S/K} = \sum_{h=1}^{H-1}\left(\frac{K_h - K_{h-1} - 1}{N}\right)\bar{y}_{S_{*h}(K_h)} + \frac{1}{N}\sum_{h=1}^{H-1} y_{Kh} + \left(1 - \frac{K_{H-1}}{N}\right)\bar{y}_{S_H} \qquad (3.162)$$

where $K=[K_1...K_{H-1}]$ and $S_{*h}(K_h) = S_h(K_h) - \{K_h\}$, for h=1,...,H-1

$$\begin{cases} \bar{y}_{S_{*h}(K_h)} = \dfrac{1}{r_h - r_{h-1} - 1}\sum\limits_{i \in S_{*h}(K_h)} y_i, \quad h = 1,...., H \\ \bar{y}_{S_H} = \dfrac{1}{n - r_{h-1}}\sum\limits_{i \in S_H} y_i \end{cases} \qquad (3.163)$$

The expected value of this statistic is as follows:

$$\begin{cases} E_{S/k}\left(\tilde{y}_{S/k}\right) = \bar{y} \\ E\left(\tilde{y}_{S/K}\right) = \bar{y} \end{cases} \qquad (3.164)$$

The variance is as follows:

$$D^2\left(\tilde{y}_{S/K}\right) = E_K\left(\sum_{h=1}^{H-1}\left(\frac{K_h - K_{h-1} - 1}{N}\right)^2 \frac{K_h - K_{h-1} - r_h + r_{h-1}}{(K_h - K_{h-1})(r_h - r_{h-1})} v_{U_{*h}} + \right.$$

$$\left. + \left(1 - \frac{K_{H-1}}{N}\right)^2 \frac{N - K_{H-1} - n + r_{H-1}}{(N - K_{H-1})(n - r_{H-1})} v_{U_H}\right) \qquad (3.165)$$

where:

$$\begin{cases} v_{U_{*h}} = \dfrac{1}{K_h - K_{h-1} - 2}\sum\limits_{i \in U_{*h}} \left(y_i - \bar{y}_{U_{*h}}\right)^2, h = 1,..., H-1 \\ v_{U_H} = \dfrac{1}{N - K_{H-1} - 1}\sum\limits_{i \in U_H} \left(y_i - \bar{y}_{U_H}\right)^2 \end{cases}$$

$$\begin{cases} \bar{y}_{U_{*h}} = \dfrac{1}{K_h - K_{h-1} - 1}\sum\limits_{i \in U_{*h}} y_i, \quad h = 1,...., H-1 \\ \bar{y}_{U_H} = \dfrac{1}{N - K_{H-1}}\sum\limits_{i \in U_H} y_i \end{cases}$$

The unbiased estimator of the variance $D^2\left(\tilde{y}_{S/k}\right)$ is shown by the equation:

$$d_S^2\left(\tilde{y}_{S/k}\right)=\sum_{h=1}^{H-1}\left(\frac{k_h-k_{h-1}-1}{N}\right)^2\frac{k_h-k_{h-1}-r_h+r_{h-1}}{(k_h-k_{h-1})(r_h-r_{h-1}-1)}v_{S_{*h}(k_h)}+$$

$$+\left(1-\frac{k_{H-1}}{N}\right)^2\frac{N-k_{H-1}-n+r_{H-1}}{(N-k_{H-1})(n+r_{H-1})}v_{S_H} \qquad (3.166)$$

where:

$$\begin{cases} v_{S_{*h}(k_h)}=\dfrac{1}{r_h-r_{h-1}-2}\sum\limits_{i\in S_{*h}(k_h)}\left(y_i-\bar{y}_{S_{*h}(k_h)}\right)^2, h=1,...,H-1 \\ v_{S_H}=\dfrac{1}{n-r_{H-1}-1}\sum\limits_{i\in S_H}\left(y_i-\bar{y}_{S_H}\right)^2 \end{cases}$$

The statistic, determined by the expression (3.162), was defined by Wywiał (2000c) in the case when r_h-r_{h-1}=constance for all h-1,...,H. Particularly, if H=2, the statistic K_1 can be determined as the sample quantile of order λ and $r_1=[n\lambda]+1$. Moreover, if H=2 and r_1=m, where n=2m or n=2m+1, the statistic K_1 is the sample median while the estimator determined by the expression (3.162) is reduced to the statistic defined by the expression (3.144).

The next possible generalizations are as follows. In the case of two auxiliary variables, their sample medians let us divide the population into four non-empty and disjoint subpopulations. This lets us generalize straightforwardly the estimator considered in the paragraph 3.6.3 and 3.6.4. Secondly, instead of a one-dimensional auxiliary variable and a variable under study, the multidimensional ones can be considered because, usually, the vector of population means is estimated and the vector of auxiliary variables can be a vailable. In this case the sample s can be divided into subsamples through minimization of a criterion function dependent on auxiliary variables or a variable under study. Next, the disjoint subpopulations are selected according to the criterion function in such a way that each subpopulation includes one and only one subsample obtained previously. Hence, we have the two-stage process of clustering. At the first stage we obtain the subsamples and at the second stage - the subpopulations. The second stage of the clustering procedure can be named a conditional procedure because it provides subpopulations including appropriate subsamples (see Wywial (1998a, 2000b, 2001, 2002). The criterion function can be defined as the sum of the intra-subsample (intra-subpopulation) variances of all auxiliary variables or the spectral radius[23] of the intra-

[23] See: Rao and Scott (1981).

subsample (intra-subpopulation) matrix of variance and covariance of auxiliary variables. Next we construct coefficients (dependent on subpopulation sizes) of linear combination of estimators (the subsample means or appropriate regression or ratio estimators from the subsamples).

3.7. Classification estimator of population mean

3.7.1. Introduction

The problem of estimation of a mean value in a fixed and finite population is considered. We assume that the values of auxiliary variables are known in the population. The simple sample is selected without replacement from the population. The sample is partitioned into homogenous sub-samples using observations of a variable under study and the auxiliary variables. On the basis of this partition the whole population is clustered into strata in such a way that each strata includes one and only one sub-sample. The estimator of the population mean is the weighed average of the sub-sample means. The weights are equal to the sizes of the strata. The strata can be determined by means of a discrimination function. This function is evaluated on the basis of observations of the auxiliary variables in the sub-samples. Several criteria of clustering the sample or population are presented. The well known bootstrap or jackknife methods are suggested to estimate the variance of the estimator. The estimator can be useful when we have a census data. In this case it is possible to cluster a population in order to evaluate the weights of the estimator. The outlined problem has been considered in a similar sense by Bethlehem (1988), Huisman (2000) and Wywiał (1999, 2001).

Let $U=\{1,...,N\}$ be a fixed population of size N. The non-negative i-th observation of a k-dimensional auxiliary variable is denoted by $\mathbf{x}_{i*}=[x_{i1}\ x_{i2}\ ...\ x_{ik}]$, $i = 1, ..., N$. The matrix of observations of the auxiliary variables is denoted by $\mathbf{x}=\begin{bmatrix}\mathbf{x}_{1*}\\ ...\\ \mathbf{x}_{N*}\end{bmatrix}$. This matrix is of dimension $N\times k$. The vector of observations of a variable under study is denoted by $\mathbf{y}=\begin{bmatrix}y_1\\ ...\\ y_N\end{bmatrix}$.

The sample s of size n is drawn without replacement from the population U. The space of the samples is denoted by **S**. The sampling design is denoted by $P(s)\geq 0$ where $s\in\mathbf{S}$. The vector of observations of the variable under study in a sample s and the matrix of observations of the auxiliary variables in the sample s are denoted by $\mathbf{y}_s$ and $\mathbf{x}_s$, respectively. The vector $\mathbf{y}_s$ is of the dimensions $n\times 1$. The matrix $\mathbf{x}_s$ is of the dimensions $n\times k$. Let a sample $s\in\mathbf{S}$ be partitioned into the following set of non-empty sub-samples: $a(s)=\{s_h, h=1,...,H(s)\}$, where H(s) is the number of the sub-samples and it can depend on s. The sub-sample s_h is of size $n_h(s)$.

Let $b = b(a) = b(a(s)) = \{U_h : U_h \subset U, s_h \subseteq U_h, s_h \in a\}$ be such a set of mutually disjoint and non-empty sub-populations (strata) that $s_h \subseteq U_h(s)$ and $s_h \not\subset U_t(s)$ for $t \neq h = 1,...,H(s)$ and $\bigcup_{h=1}^{H(s)} U_h(s) = U$. Let $N_h(s)$ be the size of the stratum $U_h(s)$. The set of all possible sequences of type a(s) generated on the basis of a particular sample s is denoted by A=A(s)={a(s)}. Similarly, the set of all possible systems of type b(a(s)) generated on the basis of a sample s and the population U is denoted by B(a)=B(a(s))={b(a(s))}. Let B(A)=B(A(s)) be the set of the all sets of type B(a) generated for all $a \in A(s)$. And finally, let $\boldsymbol{B}=B(A(\boldsymbol{S}))$ be the space of all possible sequences b(a(s)) where $a \in A(\boldsymbol{S})$ and $\boldsymbol{S}$ is the sample space.

The sets a(s) and b(a) are determined on the basis of a criterion function $f(\mathbf{y}_s, \mathbf{x})$, which will be defined later. Usually, the criterion leads to simultaneous determination of the systems a(s) and b(a). We postulate that the systems a(s) and b(a) should lead to such a set of strata $b(a)=\{U_h\}$ that intra-stratum spread of observations of auxiliary variables and the variable under study is as small as possible. Moreover, the intra-sub-samples spread of these variables should be as small as possible, too.

In order to simplify the following considerations let us assume that *s* is the simple sample drawn without replacement from the population U. The following notation will be useful:

$$\bar{y} = \frac{1}{N}\sum_{i \in U} y_i, \qquad \bar{y}_h(b) = \frac{1}{N_h}\sum_{i \in U_h} y_i, \qquad \bar{y}_{S_h(a)} = \frac{1}{n_h}\sum_{i \in S_h} y_i$$

The population average $\bar{y}$ can be estimated by means of the statistic:

$$\hat{y}_{b(a(S))} = \sum_{h=1}^{H(s)} w_h(U_h)\bar{y}_{S_h} \tag{3.167}$$

where $w_h(U_h) > 0$ for each h = 1, ..., H(s) and $s \in \mathbf{S}$ and $\sum_{h=1}^{H(s)} w_h = 1$. The next two estimators are functions of the ratio or the regression estimators. The first of them is as follows:

$$\hat{y}_{Ib(a(S))} = \sum_{h=1}^{H(s)} w_h(U_h)\hat{y}_{Is_h} \tag{3.168}$$

where the ratio estimator is:

$$\hat{y}_{Is_h} = \frac{\bar{y}_{S_h}}{\bar{x}_{S_h}}\bar{x}_{U_h} \tag{3.169}$$

The next estimator is:

$$\hat{y}_{Rb(a(S))} = \sum_{h=1}^{H(s)} w_h(U_h)\hat{y}_{RS_h} \tag{3.170}$$

where

$$\hat{y}_{RS_h} = \bar{y}_{S_h} + \beta_{S_h}\left(\bar{x}_{U_h} - \bar{x}_{S_h}\right), \tag{3.171}$$

$$\beta_{S_h} = \frac{\sum_{i\in S_h}\left(x_i - \bar{x}_{S_h}\right)y_i}{\sum_{i\in S_h}\left(x_i - \bar{x}_{S_h}\right)^2} \tag{3.172}$$

The statistic $\hat{y}_{RS_h}$ is the regression estimator.

Let H be the fixed number of strata and $\{N_h,\ h=1,...,H\}$ be the sequence of stratum sizes and the sequence of the fractions is denoted by $\left\{w_h = \frac{N_h}{N}\right\}$. For a fixed partition $b=\{U_h,\ h=1,...,H\}$ let

$$\mathbf{S}_b = \left\{s : s = \bigcup_{h=1}^{H} s_h = s, \quad \text{where} \quad s_h \subseteq U_h, s_h \neq \varnothing \quad \text{for} \quad h = 1,..., H\right\}.$$

It is obvious that: $\mathbf{S}_b \subset \mathbf{S}$. Let $M=\text{Card}\{\mathbf{S}\}$ and $M_b=\text{Card}\{\mathbf{S}_b\}$, Hence, the conditional sampling design $P(s \mid s\in \mathbf{S}_b) = \frac{1}{M_b}P(s)$, for $s\in \mathbf{S}_b$. The conditional sampling design can be implemented by the simple rejective sampling scheme. Hence, we select the sample s according to sampling design P(s). If $s\in \mathbf{S}_b$, then it is the conditional sample. If $s\notin \mathbf{S}_b$, the next sample is drawn according to sampling design P(s). The probability of selecting the conditional sample s for the first draw is $p = \frac{M_b}{M}$, for the second selection p(1-p), for the third selection $p(1-p)^2$ and so on.

Under the conditional sampling design the estimator $\hat{y}_S = \hat{y}_{ba((S))}$ takes the following particular form:

$$\hat{y}_S = \sum_{h=1}^{H} w_h \bar{y}_{s_h}\ .$$

In this particular case, the conditional estimation of a population mean is equivalent to the well known problem of estimation on the basis of sample stratified after its selection, see e.g. Bracha (1996), p. 124 or Särndal, Swenson i Wretman (1992), p. 267. Hence, $E(\hat{y}_S \mid b) = \bar{y}$ which means that the statistic $\hat{y}_S$ is conditionally unbiased estimator of $\bar{y}$ and its conditional variance is as follows:

$$D^2(\hat{y}_S \mid b) = \frac{N-n}{Nn}\sum_{h=1}^{H} w_h v_{U_h}(y) + \frac{N-n}{Nn^2}\sum_{h=1}^{H}(1-w_h)v_{U_h}(y) + O(n^{-3}) \tag{3.173}$$

where

$$v_{U_h}(y) = \frac{1}{N_h - 1}\sum_{i\in U_h}\left(y_i - \bar{y}_{U_h}\right)^2, \qquad \bar{y}_{U_h} = \frac{1}{N_h}\sum_{i\in U_h} y_i \tag{3.174}$$

The approximately conditionally unbiased estimator of the variance $D^2(\hat{y}_S \mid b)$ is the statistic:

$$D_S^2(\hat{y}_S) = \frac{N-n}{Nn}\sum_{h=1}^{H} w_h v_{S_h}(y) + \frac{N-n}{Nn^2}\sum_{h=1}^{H}(1-w_h)v_{S_h}(y) + O(n^{-3}) \tag{3.175}$$

where

$$v_{S_h}(y) = \frac{1}{N_h - 1}\sum_{i\in S_h}\left(y_i - \bar{y}_{S_h}\right)^2, \qquad \bar{y}_{S_h} = \frac{1}{N_h}\sum_{i\in S_h} y_i \tag{3.176}$$

It can be shown that in the considered particular case the sub-sample size n_h is approximately proportionate to the stratum size $N_{h.}$ for each h=1,...,H.

In the general case, the systems of sub-samples a(s) and strata b(a) depend on a sample s. That is why it is difficult to derive the basic moments of the statistic $\hat{y}_{ba((S))}$, given by the expression (3.167). The variance of this estimator can be estimated by means of the well known boostrap or jackknife techniques. The properties of these methods are considered e.g. by Efron and Tibshirani (1993). In our case the bootstrap method is as follows. The simple samples Z_t (t=1,...,m) are drawn with replacement from the sample s. Usually, the size of each sample Z_t is equal to the size of the sample s. Next, each sample Z_t is partitioned into the system of sub-samples denoted by $a_t(z_t) = \{z_{t,1},...,z_{t,H(z_t)}\}$ and the population U is partitioned into the system of strata denoted by $b_t(z_t) = \{U_{t,1},...,U_{t,H(z_t)}\}$. These partitions are obtained by means of the same method as used to determine the vagues of the estimator $\hat{y}_{ba((S))}$. Next, on the basis of the expression (3.167), the following statistics are evaluated:

$$\breve{y}_{b(a(Z_t))} = \sum_{h=1}^{H(Z_t)} w_h(U_{t,h})\bar{y}_{Z_{t,h}}. \tag{3.177}$$

The estimators of the variance of the statistic $\hat{y}_{ba((S))}$ are as follows:

$$\hat{D}^2\left(\hat{y}_{b(a(s))}\right)=\frac{1}{m-1}\sum_{t=1}^{m}\left(\breve{y}_{b(a(z_t))}-\hat{y}_{b(a(s))}\right)^2 \tag{3.178}$$

or

$$\tilde{D}^2\left(\hat{y}_{b(a(s))}\right)=\frac{1}{m-1}\sum_{t=1}^{m}\left(\breve{y}_{b(a(z_t))}-\bar{y}_Z\right)^2, \qquad \bar{y}_Z=\frac{1}{m}\sum_{t=1}^{m}\breve{y}_{b(a(z_t))} \tag{3.179}$$

The suggested jacknife method can lead not only to the estimator of the variance of the statistic $\hat{y}_{ba((S))}$ as well as to a determination of an almost unbiased estimator of a population mean $\bar{y}$.

3.7.2. Classification functions

In order to determine the weights of the estimator $\hat{y}_{ba((S))}$ sizes of strata should be evaluated. We are going to show several methods of stratification of the sample and population.

Wywiał (2001b, 2001c) proposed a construction of vague estimators of the population mean in situation when non-response are present. The weights are functions of sizes of the sub-populations of non-responses and responses and they are determined on the basis of auxiliary variables observed in the whole population and by means of a classification function. This idea can be applied to evaluate of the weights of the estimator $\hat{y}_{ba((S))}$. The sample s is partitioned into system a(s) on the basis of a variable under study and auxiliary variables. This partition can be defined by means of e.g. the k-means method or the method of Ward (1963). The criterion function of the clustering method is as follows:

$$f_0(\mathbf{y}_s,\mathbf{x}_s,a(s))=\sum_{h=1}^{H(s)}\left(v_{s_h}(y)+v_{s_h}(x)\right)\left(n_{s_h}-1\right). \tag{3.180}$$

The sample s is partitioned into homogenous sub-samples $a^{(0)}(s)=\left\{s_1^{(0)},...,s_{H(s)}^{(0)}\right\}$ on the basis of the data $(\mathbf{y}_s,\mathbf{x}_s)$. This partition leads to determining the classification function but is evaluated only on the basis of the matrix $\mathbf{x}_s$ in the following way (see e.g. Krzyśko (2000), p. 255 or 266):

$$e_{s_h}(\mathbf{x}_i)=-\frac{1}{2}q_{s_h}+\frac{n_{s_h}}{n} \tag{3.181}$$

where $\mathbf{x}_i$ is the i-th row of the matrix $\mathbf{X}_U$ and

$$q_{s_h}(\mathbf{x}_i) = \left(\mathbf{x}_i - \bar{\mathbf{x}}_{s_h}\right)\mathbf{V}_{s_h}^{-1}\left(\mathbf{x}_i - \bar{\mathbf{x}}_{s_h}\right)^T .$$

where $\bar{\mathbf{x}}_{s_h}$ is the row vector of the auxiliary variables' averages from the sample s_h. The matrix of the sample variances and covariances of auxiliary variables is denoted by $\mathbf{V}_{s_h}$. Hence, the size n_{s_h} of the sample s_h has to be at least equal to number k of the auxiliary variables. The classification function divides the population U into the set of strata $U^{(0)}(a^{(0)}(s)) = \left\{U_1^{(0)},..., U_{H(s)}^{(0)}\right\}$ on the basis of the observations $\mathbf{x}_U$ of the auxiliary variables. The i-th population element is classified into the stratum $U_h^{(0)}$ if

$$e_{s_h}(\mathbf{x}_i) = \max_{t=1,...,H(s)} \left\{e_{s_t}(\mathbf{x}_i)\right\}. \tag{3.182}$$

The weights $w_h^{(0)}(a^{(0)}(s)) = \dfrac{N_h^{(0)}}{N}$ of the estimator $\hat{y}_{ba((S))}$ are evaluated on the basis of the sizes of the strata: $N^{(0)}(a^{(0)}(s)) = \left\{N_1^{(0)},..., N_{H(s)}^{(0)}\right\}$. Let us note that the elements of the subsample s_h cannot necessarily be classified to the stratum $U_h^{(0)}$. In this case we can correct the set of the subsamples $a^{(0)}(s) = \left\{s_1^{(0)},..., s_{H(s)}^{(0)}\right\}$ to such a set $a^{(*)}(s) = \left\{s_1^{(*)},..., s_{H(s)}^{(*)}\right\}$, that $s_h^{(*)} \subseteq U_h^{(0)}$ for each h=1,...,H.

3.7.3. Stratification on the basis of vector criterion function

Previously, we suggested finding such a partition of a population U into a set of homogenous strata $\underline{b}(\underline{a}(s))=\{\underline{U}_h(s)\}$ and homogenous sub-samples $\underline{a}(s)=\{\underline{s}_h(s)\}$ that the criterion function $f(\mathbf{y}_s, \mathbf{x}_U, \underline{b}(\underline{a}(s)), \underline{a}(s))$ takes a minimal value. Under this condition and the assumption that $\underline{s}_h(s) \in \underline{U}_h(s)$ for h=1,...,H, we can postulate that a distances between $\underline{s}_h(s)$ and $\underline{U}_h(s)$, h=1,...,H should be as small as possible. More precisely, we look for such set of sub-samples $\underline{a}=\{\underline{s}_h\}$ and the set of strata $\underline{b}=\{\underline{U}_h\}$ that are solution to the following optimization problem:

$$\begin{cases} f\left(\mathbf{y}_s, \mathbf{x}_U, b(a), a(s)\right) = \min \\ d\left(\mathbf{x}_U, b(a), a(s)\right) = \min \\ b(a) \in B(a), \quad a(s) = A(s) \end{cases} \tag{3.183}$$

where d(...) is the distance function between sets $a=\{s_h\}$ and $b=\{U_h\}$. A more general optimization problem is expressed by the following system:

$$\begin{cases} f(\mathbf{y}_s, \mathbf{x}_U, b(a), a(s)) = \min \\ d(\mathbf{x}_U, b(a), a(s)) = \min \\ b(a) \in B(a), \quad a(s) = A(s) \\ n_h = n w_h(b(a)), \quad h = 1,..., H(s) \\ \left| \sum_{h=1}^{H(s)} w_h(b(a)) \bar{x}_{s_h} - \bar{x} \right| \le \delta \end{cases} \tag{3.184}$$

Particularly, the criterion function f(...) can be proportionate to the variance $D_s^2(\hat{y}_S)$ shown by the expression (3.173). Here, we simplify this function to the following form:

$$f_1(\mathbf{y}_s, b(a), a(s)) = \sum_{h=1}^{H(s)} w(b(a)) v_{s_h}(y) . \tag{3.185}$$

Let us note that the function f_1 depends on observations of the auxiliary variables through the partition b of the population into strata.

Let us consider the following ratio estimator of the variance $v_{U_h}(y)$:

$$v_{Is_h}(y) = \frac{v_{s_h}(y)}{v_{s_h}(x)} v_{U_h}(x) . \tag{3.186}$$

After substituting $v_{s_h}(y)$ for $v_{Is_h}(y)$ in the expression (3.185) we obtain the following modification of the criterion f_1:

$$f_2(\mathbf{y}_s, \mathbf{x}_U, b(a), a(s)) = \sum_{h=1}^{H(s)} w(b(a)) v_{Is_h}(y) . \tag{3.187}$$

The criterion function can be defined independently from a partition of the population into strata in the following way:

$$f_3(\mathbf{y}_s, b(a), a(s)) = \sum_{h=1}^{H(s)} (n_h - 1) v_{s_h} . \tag{3.188}$$

In the case of the estimator $\hat{y}_{Ib(a(S))}$, formulated by the expressions (3.168) and (3.169), the criterion function is based on the estimator of its variance in the following way:

$$f_4(\mathbf{y}_s, \mathbf{x}_U, b(a), a(s)) = \sum_{h=1}^{H(s)} w(b(a)) \left(v_{s_h}(y) - 2 \frac{\bar{y}_{s_h}}{\bar{x}_{s_h}} v_{s_h}(x, y) + \left(\frac{\bar{y}_{s_h}}{\bar{x}_{s_h}} \right)^2 v_{s_h}(x) \right) \tag{3.189}$$

where

$$v_{s_h}(x,y) = \sum_{i \in s_h} \left(x_i - \bar{x}_{s_h}\right)\left(y_i - \bar{y}_{s_h}\right). \tag{3.190}$$

Finally, in the case of the estimator $\hat{y}_{Rb(a(S))}$, given by the expression (3.170)-(3.172), the criterion can be the following function of the estimator of variance $D^2\left(\hat{y}_{Rb(a(S))}\right)$:

$$f_5(\mathbf{y}_s, \mathbf{x}_U, b(a), a(s)) = \sum_{h=1}^{H(s)} w(b(a))\left(v_{s_h}(y) - \beta_s^2\right) \tag{3.191}$$

where β_s id determined by the form (3.172).

The particular similarity functions are as follows.

$$d_1(\mathbf{x}_U, b, a) = \sum_{h=1}^{H(s)} w_h(b(a))\left(\bar{x}_{s_h} - \bar{x}_{U_h}\right)^2, \tag{3.192}$$

$$d_2(\mathbf{x}_U, b, a) = \sum_{h=1}^{H(s)} \sum_{i \in s_h} \sum_{j \in U_h, j \neq i} \left(x_i - x_j\right)^2. \tag{3.193}$$

The third row of the system (3.184) defines the set of all possible sets of sub-samples and strata. This set is limited by the next two conditions defined in the next two rows. In the fourth row, it is assumed that the size of each sub-ample has to be proportionate to the size of the appropriate stratum. This assumption is a result of the expression (3.175) which showed the variance of the estimator in the case when strata are fixed. In this case the expected value of each sub-sample is approximately proportional to the size of the appropriate stratum. Finally, the last row of the system (3.184) shows the calibration postulate. It is formulated on the basis of the general definition of calibrated estimators proposed by Devill and Särndal (1992). Our calibration condition becomes classical when δ=0. Let C be the set of the admissible solutions. It is defined by the three last rows of the system (3.184).

Our problem is determining such sets $\underline{a}$ and $\underline{b}$ that the vector function $g(b,a) = \begin{bmatrix} f(\mathbf{y}_s, b(a), a(s)) \\ d(\mathbf{x}_U, b(a), a(s)) \end{bmatrix}$ takes a minimal value in the set C. Only in particular cases, the problem (3.184) has one such a solution $(\underline{b},\underline{a}) \in C$ that $g(\underline{b},\underline{a}) \leq g(b,a)$ for each $(b,a) \in C$. Usually, we have to look for such a set of the non dominated solutions $\underline{C}$, that $(\underline{b},\underline{a}) \in \underline{C}$ and $(b_*,a_*) \in \underline{C}$ and $(b',a') \in C\text{-}\underline{C}$ if and only if it is not true that $g(\underline{b},\underline{a}) \leq g(b_*,a_*)$ or $g(\underline{b},\underline{a}) > g(b_*,a_*)$ and $g(\underline{b},\underline{a}) \leq g(b',a')$ and $g(b_*,a_*) \leq g(b',a')$.

When the set $\underline{C}$ consists of at least two elements, the unique solution to the problem (3.184) can be determined through minimization of an additional criterion function denoted by G(b,a) in the set $\underline{C}$. Particularly, the following function can be considered:

$$G_1(b,a) = f\left(\mathbf{y}_s, \mathbf{x}_U, b(a), a(s)\right), \tag{3.194}$$

$$G_2(b,a) = d\left(\mathbf{x}_U, b(a), a(s)\right), \tag{3.195}$$

$$G_3(b,a) = \alpha f\left(\mathbf{y}_s, b(a), a(s)\right) + (1-\alpha) d\left(\mathbf{x}_U, b(a), a(s)\right),\ \alpha \in (0;1). \tag{3.196}$$

When α=1, G_3 is reduced to G_1. If α=0, G_3 is reduced to G_2.

3.7.4. Criterions based on the depth functions

Properties of the depth functions are considered e.g. by Liu (1990), Donoho an Gasko (1992), Rousseeuw and Ruts (1996), Struyf and Rousseeuw (1999), Wagner and Kobylińska (2000). For the sake of simplicity, we consider a two dimensional auxiliary variable. Let $\Delta\left(\mathbf{x}_{i*}, \mathbf{x}_{j*}, \mathbf{x}_{t*}\right)$ where i≠j=1,...,N, be the triangle whose vertexes have the coordinates determined by the vectors $\{\mathbf{x}_{i*}, \mathbf{x}_{j*}, \mathbf{x}_{t*}\}$. The number of the triangles whose vertexes are determined by the rows of the matrix **x**, is equal to $\binom{N}{3} = \frac{1}{6}N(N-1)(N-2)$. The measure of the triangle (simplicial) depth of a point $\boldsymbol{\theta}$ in R^2 is the number $l(\boldsymbol{\theta})$ of the triangles including the point $\boldsymbol{\theta}$. The normalized variant of the triangle measure is as follows

$$z(\boldsymbol{\theta}) = \binom{N}{3}^{-1} l(\boldsymbol{\theta}) \in \langle 0;1 \rangle. \tag{3.197}$$

The value of $z(\boldsymbol{\theta})$ is close to one if the point $\boldsymbol{\theta}$ is situated near the center of the considered set of points whose coordinates are determined by observations of the auxiliary variables.

In our case we consider the depth of an i-th element of a population U among the remaining elements of this population with respect to observations of the auxiliary variables represented by the elements of the matrix **x**. The i-th row $\mathbf{x}_{i*}$ of the matrix **x** is attached to the i-th element of the population U and vice-versa. So, equivalently we can consider the depth of a point with coordinates $\mathbf{x}_{i*}$ among the remaining points $\mathbf{x}_{j*}$, j=1,...,N and i≠j. Hence, the normal measure of the depth of an i-th population element with regard to the auxiliary variables is as follows (see, Wagner and Kobylińska (2000), p. 205):

$$z\left(\mathbf{x}_{i*}\right) = \binom{N-1}{3}^{-1} l\left(\mathbf{x}_i\right), \qquad z\left(\mathbf{x}_{i*}\right) \in \langle 0;1 \rangle. \tag{3.198}$$

The mean measure of the triangular (simplicial) depth of a sample s in the population U (with regard to auxiliary variables) is defined as follows:

$$z(\mathbf{x}_s) = \frac{1}{n}\sum_{i\in s} z(\mathbf{x}_{i^*}) \in \langle 0, 1\rangle . \tag{3.199}$$

Rousseeuw and Ruts (1996), Wagner and Kobylińska (2000) showed methods of checking, if a point $\mathbf{x}_{i^*}$ was included in a triangle $\Delta(\mathbf{x}_{i^*}, \mathbf{x}_{t^*}, \mathbf{x}_{b^*})$ or not.

We say that a mean depth of a sample s in the population U with regard to auxiliary variables is high when the value of $z(\mathbf{x}_s)$ is close to one. Particularly, $z(\mathbf{x}_U) = \frac{1}{N}\sum_{i\in U} z(\mathbf{x}_{i^*})$ determines the mean measure of depth of U in U. Hence, the mean depth of a sample s in U can be compared with mean depth U in U in the following way:

$$\delta(\mathbf{x}_s) = z(\mathbf{x}_U) - z(\mathbf{x}_s) \tag{3.200}$$

When $\delta(\mathbf{x}_s)$ is negative, the mean depth of s in U is bigger than U in U. In this case, the set of points $\{\mathbf{x}_{i^*}: i\in s\}$ is situated close to the center of all points from the set $\{\mathbf{x}_{i^*}: i\in U\text{-}s\}$. When $\delta(\mathbf{x}_s) > 0$, the sample s has a small mean depth in U. In this case, we can expect that the points from the set $\{\mathbf{x}_{i^*}: i\in s\}$ should be far from the center of the population U. Finally, if $\delta(\mathbf{x}_s)=\delta(\mathbf{x})$, we can expect that the points from the set $\{\mathbf{x}_{i^*}: i\in s\}$ are "uniformly" distributed (spread) among the remaining points from the set $\{\mathbf{x}_{i^*}: i\in U\text{-}s\}$.

We can treat a sample s as representative with regard to auxiliary variables if $\delta(\mathbf{x}_s)\leq 0$. Especially, it is reasonable when auxiliary variables and the variable under study are highly dependent on each others and our purpose is an estimation of the average or median of the variable under study.

The definitions introduced above lead to the following criterion function of clustering a sample into the set of strata b(a(s)):

$$f_6(\mathbf{x}_U, b, a) = -\sum_{h=1}^{H} z_h\left(\mathbf{x}_{S_h}\right) \tag{3.201}$$

where $z_h\left(\mathbf{x}_{S_h}\right)$ is the mean measure of the depth of a subsample s_h in the stratum U_h with regard to auxiliary variables. Minimization of the function f_6 leads to such partitions a(s) and b(a) that the points whose coordinates are rows of the matrix $\mathbf{x}_{S_h}$ are close to the center of the points whose coordinates are rows of the matrix $\mathbf{x}_{U_h - s_h}$.

Let us consider the following similarity function:

$$d_3(\mathbf{x}_U b, a) = \sum_{h=1}^{H} \delta_h^2\left(\mathbf{x}_{S_h}\right) \tag{3.202}$$

where

$$\delta_h\left(\mathbf{x}_{S_h}\right) = z\left(\mathbf{x}_{U_h}\right) - z\left(\mathbf{x}_{S_h}\right). \tag{3.203}$$

Minimization of the function d_3 leads to such sets of sub-samples a(s) and to the strata b(a) that the spreads of observations of auxiliary variables in the sub-sample s_h and the appropriate strata are similar, h=1,...H.

We can expect that such stratifications should lead to decreasing the mean-square error of the estimator $\hat{y}_{ba((S))}$, given by the expression (3.167).

The considered problems can be straightforwardly generalized in the case when an auxiliary variable has more than two-dimensions. When this dimension is denoted by k, the number of all k-dimensional simplexes including a point $\mathbf{x}_{i*}$ is equal to $\binom{N-1}{k+1}$. Hence, the number of these simplexes increases very quickly even in the case when k and N are not too large. This limits the applications of the considered method in practice.

3.7.5. Clustering algorithm

We are going to present a clustering algorithm which in some sense can be treated as a generalization of the well known k-means clustering algorithm. Let a fixed and finite population U be partitioned into the starting set of strata denoted by $b_o = \left\{U_1^{(o)},...,U_H^{(o)}\right\}$. Let the set of such samples $a_o(s) = \left\{s_1^{(o)},...s_H^{(o)}\right\}$ that $s_h^{(o)} \neq \emptyset$ and $s_h^{(o)} \subseteq U_h^{(o)}$, $h = 1,...,H$ and $\bigcup_{h=1}^{H} s_h^{(o)}$ be arbitrarily determined. Next, the value of a vector criterion function $g^{(0)}(b,a) = \begin{bmatrix} f^{(0)}\left(\mathbf{y}_s, b^{(0)}(a^{(0)}), a^{(0)}(s)\right) \\ d^{(0)}\left(\mathbf{x}_U, b^{(0)}(a^{(0)}), a^{(0)}(s)\right) \end{bmatrix}$ is evaluated. The size of the sub-sample $s_h^{(o)}$ is denoted by $n_h^{(o)}$. Let $N_h^{(o)}$ be the size of $U_{h1}^{(o)}$ for starting t=0: Let us note that the considered variables should be standardized if they are observed in several scales.

We say that the partition (b_*,a_*) is dominated by (b_{**},a_{**}) if and only if $g(b_*,a_*) \geq g(b_{**},a_{**})$. If this inequality as well as $g(b_*,a_*) < g(b_{**},a_{**})$ are not true, the vectors $g(b_*,a_*)$ and $g(b_{**},a_{**})$ do not dominate each other and we say that the partitions $g(b_*,a_*)$ and $g(b_{**},a_{**})$ do not dominate each other. It is obvious that the solution to our problem is such a subset of admissible partitions of the population that vectors of values of the criterion function attached to them appropriately are non dominated. In order to explain

this problem more precisely let us denote the set of non-dominated values of the vector criterion function obtained during the t-th iteration by $L_t=\{g^{(t)}(b^{(t)},a^{(t)})\}$. Let $P_t=\{(b^{(t)},a^{(t)})\}$ be the set of the non-dominated partitions obtained at the end of the t-th iteration. During the t-th iteration, elements of the population are moved from one stratum to another. This leads to new partitions of the population and values of vector of criterion function denoted as follows:

$$g^{(t+1)}\left(b_k^{(t)}\left(a_k^{(t)}\right),a_k^{(t)}\right)=\begin{bmatrix} f^{(t)}\left(\mathbf{y}_s,b_k^{(t)}\left(a_k^{(t)}\right),a_k^{(t)}\right)\\ d^{(t)}\left(\mathbf{x}_U,b_k^{(t)}\left(a_k^{(t)}\right),a_k^{(t)}\right)\end{bmatrix},k=1,...,N \qquad (3.204)$$

where: 1) if the element $\{k\}\in s_z^{(t)}$ and $n_z^{(t)}\le 2$, the partition is not admissible,

2) when $\{k\}\in s_z^{(t)}$ and $n_z^{(t)}>2$, then

$$a_k^{(t)}=\left\{s_h^{(t)},h=1,....,H;\ \ h\neq z,\ \ h\neq r;\ \ s_z^{(t)}-\{k\},s_r^{(t)}\cup\{k\}\right\} \qquad (3.205)$$

Here, other conditions of partition admissibility can be added, e.g. those defined by thew expression (3.184).

3) if $\{k\}\notin s_z^{(t)}$ and $\{k\}\in U_z^{(t)}-s_z^{(t)}$ and $N_z^{(t)}>n_z^{(t)}\ge 2$ then

$$b_k^{(t)}=\left\{U_h^{(t)},h=1,....,H;\ \ h\neq z,\ \ h\neq r;\ \ U_z^{(t)}-\{k\},\ \ U_r^{(t)}\cup\{k\}\right\} \qquad (3.206)$$

Next,

1) when for each $g(b,a)\in L_t$ the inequality $g^{(t)}\left(b_k^{(t)},a_k^{(t)}\right)>g(b,a)$ is fulfilled, the partition $\left(b_k^{(t)},a_k^{(t)}\right)$ is dominated by the set of the partitions P_t, and $\left(b_k^{(t)},a_k^{(t)}\right)$ is rejected as the solution to our problem,

2) if for each $g(b,a)\in L_t$ the inequality $g^{(t+1)}\left(b_k^{(t)},a_k^{(t)}\right)<g(b,a)$ is true, the partition $\left(b_k^{(t)},a_k^{(t)}\right)$ dominates each partition belonging to the set P_t and we state that $P_t=\left(b_k^{(t)},a_k^{(t)}\right)$,

3) when it is not true that for each $g(b,a)\in L_t$ $g^{(t)}\left(b_k^{(t)},a_k^{(t)}\right)>g(b,a)$ and $g^{(t)}\left(b_k^{(t)},a_k^{(t)}\right)<g(b,a)$, the partition $\left(b_k^{(t)},a_k^{(t)}\right)$ does not dominate any partition from the set P_t and none of the partitions from the set P_t dominates the partition $\left(b_k^{(t)},a_k^{(t)}\right)$. Hence, the partition $\left(b_k^{(t)},a_k^{(t)}\right)$ is added to the set P_t.

The (t+1)-th iteration of the clustering algorithm is terminated when k=N.

Let us underline that at the end of a t-th iteration $P_{t-1} \subset P_t$ or the one element set of non-dominated partitions can be obtained, so $P_{t-1} \neq P_t$. In this case the next iteration can be started. The clustering algorithm is terminated at the t-th iteration if $P_{t+1} = P_t$. The iteration process can be stopped when t=T where T is an admissible number of iterations, usually determined by the efficiency and speed of computer systems.

The presented algorithm leads to the set of partitions P_T optimal only in the local sense. Moreover, the obtained set P_T is not necessarily the set of non-dominated partitions. It can be only the locally an optimal set.

The set of non-dominated solutions P_t can include more than one partition but in practice we need only one of them. In order to find the unique partition, the next criterion should be defined. Sometimes such a criterion function is called the super-criterion. For instance, minimization of the function f_i or d_i in the set P_T leads to the unique partition. Of course, the clustering algorithm is simplified when instead of the vector criterion function a scalar function is considered.

Let us suppose that one of the variants of the presented clustering algorithm leads to the following partition of a population into the set of strata $P^{\#}(s) = \left\{U_1^{\#},...,U_H^{\#}\right\}$ and the partition of a sample s into the system of subsamples $a^{\#}(s) = \left\{S_1^{\#},...,S_H^{\#}\right\}$. The estimator $\hat{y}_{b(a(S))}$ defined by the expression (3.167) takes the form

$$\hat{y}_{b^{\#}(a^{\#}(S))} = \sum_{h=1}^{H} w_{\#h}\ \bar{y}_{s_h^{\#}}, \quad w_{\#h} = \frac{N_h^{\#}}{N}. \tag{3.207}$$

where: $N_h^{\#}$ is the size of the stratum $U_h^{\#}$.

The variance of the statistic $\hat{y}_{b^{\#}(a^{\#}(S))}$ can be estimated e.g. by means of the methods of bootstrap or jackknife as we mentioned earlier.

The proposed estimator can be useful in the case when values of auxiliary variables are known in all population. Such data can derived from administrative registers or a census. Moreover, the estimators can be applied in the second phase of two-phase survey sampling when in the first phase sample values of auxiliary variables are observed. A similar situation is in the case of some rotation sampling designs when surveys are represented on at least two occasions.

Let us note that some other methods of clustering a population can be applied in order to make up reasonable partitions of a sample as well as of the population.

IV. CLUSTER SAMPLING

4.1. Basic definitions and notation

A fixed population of the size N is denoted by $\Omega=\{1,2,...,N\}$. Let us assume that the population Ω is divided into G such mutually disjoint clusters Ω_p (p=1,...,G) that $\bigcup_{p=1}^{G}\Omega_p = \Omega$. If each cluster is of the same size denoted by M, the population Ω is of the size N = GM. Let S be the cluster sample of the size g. The random sample S is drawn according to the following design:

$$P_g(s) = \binom{G}{g}^{-1}.$$

A k-th outcome of an i-th variable is denoted by y_{ik}. The sum of observations of an i-th variable in a p-th cluster is as follows:

$$z_{ip} = \sum_{k\in\Omega_p} y_{ik} .$$

The mean value of an i-th variable in a p-th cluster is:

$$\bar{y}_{ip} = \frac{1}{M} z_{ip} .$$

The mean value of an i-th variable per cluster is:

$$\bar{z}_i = \frac{1}{G}\sum_{p=1}^{G} z_{ip} .$$

The population mean of an i-th variable takes the following form:

$$\bar{y}_i = \frac{1}{N}\sum_{p=1}^{G} z_{ip} .$$

The variance-covariance matrix is denoted by: $C_* = [c_*(\mathbf{y}_i, \mathbf{y}_j)]$, where:

$$c_*(y_i, y_j) = \frac{1}{N-1}\sum_{p=1}^{G}\sum_{k\in\Omega_p}(y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j).$$

The variance-covariance matrix of cluster sums is denoted by: $\mathbf{C}_*(z) = [c_*(\boldsymbol{z}_i, \boldsymbol{z}_j)]$, where:

$$c_*(z_i, z_j) = \frac{1}{G-1}\sum_{p=1}^{G}(z_{ip} - \bar{z}_i)(z_{jp} - \bar{z}_j).$$

The estimator of the vector $\bar{\mathbf{y}} = [\bar{y}_1 \ldots \bar{y}_m]$ is defined as the vector $\bar{\mathbf{y}}_{gS} = [\bar{y}_{1gS} \ldots \bar{y}_{mgS}]$, where:

$$\bar{y}_{igS} = \frac{1}{gM}\sum_{p\in S}\sum_{k\in\Omega_p} y_{ik} = \frac{1}{gM}\sum_{p\in S} z_{ip} \ . \tag{4.1}$$

The vector $\bar{\mathbf{y}}_{gS}$ is the unbiased estimator of the mean vector $\bar{\mathbf{y}}$.

A covariance of the estimators $\bar{y}_{igS}, \bar{y}_{jgS}$ (i≠j=1,...,m) can be derived similarly to a variance of the statistic $\bar{y}_{igS}$ (i=1,...,m)[24]:

$$\mathrm{Cov}(\bar{y}_{igS}, \bar{y}_{jgS}) = \frac{G-g}{GgM^2} c_*(z_i, z_j) \tag{4.2}$$

The variance-covariance matrix of the $\bar{\mathbf{y}}_{gS}$ can be written down in the following way:

$$\mathbf{V}(\bar{y}_{gS}, P_g) = \frac{G-g}{GgM^2}\mathbf{C}_*(z) \tag{4.3}$$

where: $\mathbf{C}_*(z) = [c_*(z_i, z_j)]$.

The unbiased estimator of the covariance is obtained through substitution of the following statistic for the parameter $c_*(z_i, z_j)$:

$$c_{*S}(z_i, z_j) = \frac{1}{g-1}\sum_{p\in S}(z_{ip} - \bar{z}_i)(z_{jp} - \bar{z}_j).$$

4.2. The matrix of coefficients of within-cluster correlation

The covariance $c_*(\boldsymbol{z}_i, \boldsymbol{z}_j)$ can be decomposed in the following way[25]:

$$c_*(z_i, z_j) = \frac{N-1}{G-1}\sqrt{v_*(y_i)v_*(y_j)}\left[r_{ij} + (M-1)r_{ij}^{(w)}\right] \tag{4.4}$$

[24] See e.g. Cochran (1963), Konijn (1973).

[25] See e.g. Cochran (1963), Konijn (1973), Zasępa(1972).

where:

$$r_{ij} = \frac{c_*(y_i, y_j)}{\sqrt{v_*(y_i)v_*(y_j)}},$$

$$r_{ij}^{(w)} = \frac{c^{(w)}(y_i, y_j)}{\sqrt{v_*(y_i)v_*(y_j)}}, \tag{4.5}$$

$$c^{(w)}(y_i, y_j) = \frac{2}{(N-1)(M-1)} \sum_{p=1}^{G} \sum_{k \neq l \in \Omega_p} \sum (y_{ik} - \bar{y}_i)(y_{jl} - \bar{y}_j) \tag{4.6}$$

The parameter $r_{ii}^{(w)}$ is the coefficient of the within-cluster correlation of the i-th variable. Similarly, the parameter $r_{ij}^{(w)}$ can be named the coefficient of the within-cluster correlation of the i-th and j-th variables.

The coefficient $r_{ii}^{(w)}$ takes values from the interval[26]: $\left\langle -\frac{1}{M-1}; 1 \right\rangle$.

On the basis of the expressions (4.4) and (4.2), we have:

$$\mathrm{Cov}(\bar{y}_{igS}, \bar{y}_{jgS}) = \frac{(N-1)(G-g)}{(G-1)M^2Gg} \sqrt{v_*(y_i)v_*(y_j)} \left[r_{ij} + (M-1)r_{ij}^{(w)} \right]. \tag{4.7}$$

Let $\mathbf{R}^{(w)} = \left[r_{ij}^{(w)} \right]$ be the matrix of the coefficients of within-cluster correlation and let $\mathbf{R} = \left[r_{ij} \right]$ be the matrix of the correlation coefficients. The diagonal matrix of the variances of the variables is: $\mathbf{D}_*$=diag$\mathbf{C}_*$. Hence, the expression (4.7) leads to the following covariance matrix of the vector $\bar{\mathbf{y}}_{gS}$:

$$\mathbf{V}(\bar{\mathbf{y}}_{gS}, P_g) = \frac{(N-1)(G-g)}{(G-1)M^2Gg} \mathbf{D}_*^{1/2} \left(\mathbf{R} + (M-1)\mathbf{R}^{(w)} \right) \mathbf{D}_*^{1/2}. \tag{4.8}$$

If N and G are large, then:

$$\mathbf{V}(\bar{\mathbf{y}}_{gS}, P_g) \approx \frac{1}{Mg} \mathbf{D}^{1/2} \left(\mathbf{R} + (M-1)\mathbf{R}^{(w)} \right) \mathbf{D}^{1/2} \tag{4.9}$$

where: $\mathbf{D} = \frac{N-1}{N} \mathbf{D}_*$.

[26] See e.g. Cochran (1963), Konijn (1973).

The standardized value of the outcome y_{ik} is as follows:

$$h_{ik} = \frac{y_{ik} - \overline{y}_i}{\sqrt{v_*(y)}} . \tag{4.10}$$

Hence, on the basis of the expressions (4.5) and (4.6), the coefficient of within-cluster correlation can be rewritten in the following way:

$$r_{ij}^{(w)} = \frac{2}{(N-1)(M-1)} \sum_{p=1}^{G} \sum_{k \neq l \in \Omega_p} \sum h_{ik} h_{jl} \tag{4.11}$$

Let $\mathbf{h}=[h_{ij}]$ be an N×m matrix of data, where h_{ij} is an i-th observation of a j-th variable. The first M rows of the matrix **h** consist of observations of the standardized variables in the first cluster, the next M rows of this matrix consist of observations of variables in the second cluster and so on. Let $\mathbf{J}_b$ be the column b×1 vector, consisting of b elements, all equal to one. The Kronecker matrices' product is denoted by ⊗. Hence, the expressions (4.10) and (4.11) lead to the following formula:

$$\mathbf{R}^{(w)} = \frac{1}{(N-1)(M-1)} \mathbf{h}^T \left(\mathbf{B}\mathbf{B}^T - \mathbf{I}_N \right) \mathbf{h} \tag{4.12}$$

where:

$$\mathbf{h} = \mathbf{A}\mathbf{y}\mathbf{D}^{-1/2} , \tag{4.13}$$

$$\mathbf{A} = \mathbf{I}_N - \frac{1}{N} \mathbf{J}_N \mathbf{J}_N^T , \qquad \mathbf{B} = \mathbf{I}_N \otimes \mathbf{J}_M^T .$$

Let **P** be such an orthogonal matrix that $\mathbf{P}^T \mathbf{R}^{(w)} \mathbf{P} = \mathbf{D}_R$, where $\mathbf{D}_R$ is the diagonal matrix which consists of the eigenvalues of the matrix $\mathbf{R}^{(w)}$. They are denoted by d_{Ri} (i=1,...,m). Hence, on the basis of the expression (4.12), we have:

$$\mathbf{D}_R = \frac{1}{(N-1)(M-1)} \mathbf{u}^T \left(\mathbf{B}^T \mathbf{B} - \mathbf{I}_N \right) \mathbf{u} \tag{4.14}$$

where **u** = **hP**. The expression (4.13) leads to the following one:

$$\mathbf{u} = \mathbf{A}\mathbf{y}\mathbf{D}^{-1/2}\mathbf{P} . \tag{4.15}$$

Hence, the diagonal element d_{Ri} of the matrix $\mathbf{D}_R$ is the coefficient of within-cluster correlation of an i-th variable. Their observations create the i-th column of the matrix **u**. The matrix **u** is the linear transformation (expressed

by the formula (4.15)) of the data matrix **y**. The well known properties of the coefficient of within-cluster correlation (see. e.g. Konijn (1973)) lead to the conclusion:

$$-\frac{1}{M-1} \le d_{Ri} \le 1, \qquad i = 1,\ldots,m \tag{4.16}$$

where:

$$d_{Ri} = 1 - \frac{v_w(u_i)}{v(u_i)}, \tag{4.17}$$

$$v(u_i) = \frac{1}{N}\sum_{p=1}^{G} \sum_{k\in\Omega_p} (u_{ik} - \bar{u}_i)^2, \qquad \bar{u}_i = \frac{1}{N}\sum_{p=1}^{G} \sum_{k\in\Omega_p} u_{ik}\ , \tag{4.18}$$

$$v_w(u_i) = \frac{1}{G(M-1)}\sum_{p=1}^{G} \sum_{k\in\Omega_p} (u_{ik} - \bar{u}_{ip})^2, \qquad \bar{u}_{ip} = \frac{1}{M}\sum_{k\in\Omega_p} u_{ik}\ . \tag{4.19}$$

The within-cluster variance is denoted by $v_w(u_i)$. Hence, the coefficient d_{Ri} is the ratio of the within-cluster variance and the order variance. Similarly to the one-dimensional case (see e.g. Konijn (1973), p. 225-227), the following properties can be derived:

$$\mathbf{R}^{(w)} = \mathbf{D}^{-1/2}\left(\mathbf{C}_m - \frac{1}{M}\mathbf{C}_w\right)\mathbf{D}^{-1/2} \tag{4.20}$$

or:

$$\mathbf{R}^{(w)} = \mathbf{D}^{-1/2}(\mathbf{C} - \mathbf{C}_w)\mathbf{D}^{-1/2} \tag{4.21}$$

or:

$$\mathbf{R}^{(w)} = \frac{1}{M-1}\mathbf{D}^{-1/2}(M\mathbf{C}_m - \mathbf{C})\mathbf{D}^{-1/2} \tag{4.22}$$

where $\mathbf{C} = \frac{N-1}{N}\mathbf{C}_*$ and $\mathbf{C}_m = [c_m(y_i, y_j)]$ is the between-cluster matrix of the covariances, where:

$$c_m(y_i, y_j) = \frac{1}{G}\sum_{p=1}^{G}(\bar{y}_{ip} - \bar{y}_i)(\bar{y}_{jp} - \bar{y}_j). \tag{4.23}$$

The within-cluster matrix of the covariances is denoted by $\mathbf{C}_w = [c_w(y_i, y_j)]$, where:

$$c_w(y_i, y_j) = \frac{1}{G(M-1)}\sum_{p=1}^{G} \sum_{k\in\Omega_p} (y_{ik} - \bar{y}_{ip})(y_{jk} - \bar{y}_{jp}).$$

The well known properties of the matrix determinant and the expressions (4.21), (4.22), (4.14) let us prove the following theorem:

Theorem 4.1: Let the matrices **C** and $\mathbf{C}_w$ be positive semi-definite. Hence, a) if the matrix $\mathbf{R}^{(w)}$ is positive semi-definite, then $\det\mathbf{C} \geq \det\mathbf{C}_w$ and $d_{Ri} \geq 0$, for i=1,...,m; b) if $\mathbf{R}^{(w)}$ is negative semi-definite, then $\det\mathbf{C} \leq \det\mathbf{C}_w$ and $d_{Ri} \leq 0$ for i=1,...,m. These inequalities become sharp if the matrix $\mathbf{R}^{(w)}$ in the case a) is positive definite and in the case b) is negative definite.

We can say that the within-cluster spread of observations of a multidimensional variable is less than their population spread if the matrix $\mathbf{R}^{(w)}$ is positive definite. When $\mathbf{R}^{(w)}$ is negative definite, then we say that the population spread of values of a multidimensional variable is less than their within-cluster spread.

4.3. Homogeneity coefficient of a multidimensional variable

Similarly to the one-dimensional case[27], the variance-covariance matrix $\mathbf{C}_*$ can be decomposed in the following way:

$$(N-1)\mathbf{C}_* = G\mathbf{C}_m + (N-G)\mathbf{C}_w. \tag{4.24}$$

The matrix $\mathbf{C}_*(z)$ can be rewritten as follows:

$$\mathbf{C}_*(z) = \frac{MG}{G-1}\mathbf{C}_m. \tag{4.25}$$

This expression and the equation (4.24) lead to the following result:

$$\mathbf{C}_*(z) = \frac{M}{G-1}\left((N-1)\mathbf{C}_* - (N-G)\mathbf{C}_w\right).$$

Provided the matrix **C** is nonsingular, we have:

$$\mathbf{C}_*(z) = M\mathbf{C}_*\left(\mathbf{I} + \frac{N-G}{G-1}\boldsymbol{\Delta}\right) \tag{4.26}$$

where:

$$\boldsymbol{\Delta} = \mathbf{I} - \mathbf{C}_*^{-1}\mathbf{C}_w. \tag{4.27}$$

In the case of a one-dimensional variable y_i, when $\mathbf{C}_*$ is reduced to the variance v and $\mathbf{C}_w$ is the within-cluster variance v_w, the matrix $\boldsymbol{\Delta}$ is reduced to the homogeneity coefficient[28]:

---

[27] See e.g. Cochran (1963), p. 240.

[28] See Särndal, Swenson, Wretman (1992), p. 130.

$$\delta(y_i) = 1 - \frac{v_w(y_i)}{v_*(y_i)}, \quad -\frac{G-1}{N-G} \le \delta(y_i) \le 1 \tag{4.28}$$

where:

$$v_*(y_i) = \frac{1}{N-1}\sum_{p=1}^{G} \sum_{k\in\Omega_p} (y_{ik} - \bar{y}_i)^2, \quad \bar{y}_i = \frac{1}{N}\sum_{p=1}^{G} \sum_{k\in\Omega_p} y_{ik}, \tag{4.29}$$

$$v_w(y_i) = \frac{1}{G(M-1)}\sum_{p=1}^{G} \sum_{k\in\Omega_p} (y_{ik} - \bar{y}_{ip})^2, \qquad \bar{y}_{ip} = \frac{1}{M}\sum_{k\in\Omega_p} y_{ik}. \tag{4.30}$$

Then, the matrix $\mathbf{\Delta}$ can be treated as a generalization of the homogeneity coefficient δ. That is why the matrix $\mathbf{\Delta}$ can be named homogeneity matrix of a multidimensional variable.

Theorem 4.2.: If the variance-covariance matrix $\mathbf{C}_*$ is non-singular, the eigenvalues λ_i (i=1,...,m) of the matrix $\mathbf{\Delta}$ fulfill the following inequalities:

$$-\frac{G-1}{N-G} \le \lambda_i \le 1, \qquad \text{for each i=1,...,m} \tag{4.31}$$

Proof: The characteristic equation for the matrix $\mathbf{\Delta}$ can be transformed as follows:

$$|\mathbf{\Delta} - \lambda\mathbf{I}| = 0, \tag{4.32}$$

$$|\mathbf{I} - \mathbf{C}_*^{-1}\mathbf{C}_w - \lambda\mathbf{I}| = 0,$$

$$|\mathbf{C}_*^{-1}\mathbf{C}_w - \kappa\mathbf{I}| = 0 \tag{4.33}$$

where $\kappa=(1-\lambda)$. Since the matrix $\mathbf{C}_*^{-1}\mathbf{C}_w$ is positive semi-definite, its eigenvalues $\kappa_i \ge 0$ for each i=1,...,m. Hence, the eigenvalues of the matrix $\mathbf{\Delta}$ are: $\lambda_i \le 1$ for each i=1,...,m.

Since the matrix $\mathbf{C}_m$ is positive semi-definite, the equation (4.24) leads to the matrix

$$\mathbf{A}_1 = (N-1)\mathbf{C}_* - (N-G)\mathbf{C}_w$$

which is positive semi-definite, too. Because the matrix $\mathbf{C}_*$ is positive definite the following matrix is positive semi-definite:

$$\mathbf{A}_2 = \frac{1}{N-G}\mathbf{C}^{-1}\mathbf{A}_1 = \frac{N-1}{N-G}\mathbf{I} - \mathbf{C}^{-1}\mathbf{C}_w.$$

After several simple algebraic transformations we have:

$$\mathbf{A}_2 = \boldsymbol{\Delta} + \frac{G-1}{N-G}\mathbf{I}. \tag{4.34}$$

Let us do the following transformations:

$$|\boldsymbol{\Delta}-\lambda\mathbf{I}|=0,$$

$$\left|\boldsymbol{\Delta} + \frac{G-1}{N-G}\mathbf{I} - \frac{G-1}{N-G}\mathbf{I} - \lambda\mathbf{I}\right| =0,$$

$$|\mathbf{A}_2-\varsigma\mathbf{I}|=0 \tag{4.35}$$

where:

$$\varsigma = \frac{G-1}{N-G} + \lambda. \tag{4.36}$$

Since the matrix $\mathbf{A}_2$ is positive semi-definite, the eigenvalue $\varsigma_i \geq 0$ for each i=1,...,m. Hence, on the basis of the expression (4.36) we have: $\lambda_i \geq -\frac{G-1}{N-G}$ for i=1,...,m. This completes the proof.

We can say that the within-cluster spread of observations of a multidimensional variable is less than their population spread if the matrix $\boldsymbol{\Delta}$ is positive definite. When $\boldsymbol{\Delta}$ is negative definite, then we say that the population spread of values of a multidimensional variable is less than the within-cluster spread.

4.4. Accuracy of cluster sample mean vector in relation to simple sample mean vector

Let $\bar{\mathbf{y}}_S$ be the vector of the mean from the simple random sample of the size n, selected without replacement from a population of the size N. Its variance-covariance matrix is of the following form:

$$\mathbf{V}(\bar{\mathbf{y}}_S, P_3) = \frac{N-n}{Nn}\mathbf{C}_* = \frac{N-n}{Nn}\mathbf{D}_*^{1/2}\mathbf{R}\mathbf{D}_*^{1/2} \tag{4.37}$$

where P_3 is defined by the expression (1.28).

If the number of clusters $G \to \infty$ and n=gM, then, on the basis of the expression (4.9), we have:

$$\mathbf{V}\left(\bar{\mathbf{y}}_{gS}, P_g\right) - \mathbf{V}\left(\bar{\mathbf{y}}_S, P_3\right) = \frac{M-1}{gM}\mathbf{D}_*^{1/2}\mathbf{R}^{(w)}\mathbf{D}_*^{1/2}. \tag{4.38}$$

This result leads to the following property.

Theorem 4.3: Let a fixed population be divided into mutually disjoint clusters and let each cluster be of the same size. Hence, if $G \to \infty$, $N \to \infty$, $\frac{N}{G}$ =const and the matrix $\mathbf{R}^{(w)}$ is negative semi-definite (positive semi-definite), the strategy $\left(\bar{\mathbf{y}}_{gS}, P_g\right)$ is not worse (not better) than the strategy $\left(\bar{\mathbf{y}}_S, P_3\right)$.

This theorem as well as the expressions (4.38) and (4.8) lead to the following conclusions:

$$e_{0i} = \frac{D^2\left(\bar{y}_{gSi}, P_g\right)}{D^2\left(\bar{y}_{Si}, P_3\right)} = v\left(1 + (M-1)r_{ii}^{(w)}\right), \quad i=1,...,m, \tag{4.39}$$

$$e_1 = \frac{\det \mathbf{V}\left(\bar{\mathbf{y}}_{gS}, P_g\right)}{\det \mathbf{V}\left(\bar{\mathbf{y}}_S, P_3\right)} = \frac{\det\left(\mathbf{R} + (M-1)\mathbf{R}^{(w)}\right)}{\det \mathbf{R}}, \tag{4.40}$$

$$e_2 = \frac{q^2\left(\bar{\mathbf{y}}_{gS}, P_g\right)}{q^2\left(\bar{\mathbf{y}}_S, P_3\right)} = \sum_{i=1}^{m} r_{ii}^{(w)} a_i = \tilde{r} \tag{4.41}$$

where:

$$a_i = \frac{v(y_i)}{\sum_{i=a}^{m} v(y_i)},$$

$$e_3 = \frac{\lambda_1\left(\bar{\mathbf{y}}_{gS}, P_g\right)}{\lambda_1\left(\bar{\mathbf{y}}_S, P_3\right)}. \tag{4.42}$$

where $\lambda_1(.,.)$ is the maximal eigenvalue of a variance-covariance matrix of a strategy. Hence, on the basis of the theorem 4.3, we conclude that if the matrix $\mathbf{R}^{(w)}$ is positive definite (negative definite), the strategy $\left(\bar{\mathbf{y}}_{gS}, P_g\right)$ is less (more) accurate than the strategy $\left(\bar{\mathbf{y}}_S, P_3\right)$ in the sense of the above defined coefficients of relative efficiency. Particularly, if $\mathbf{R}^{(w)}$ is negative definite, then $e_k<1$ for k=1,2,3 and $e_{0i}\leq 1$ for i=1,...,m and $e_{0j}<1$ for at least one index j=1,...,m. The strategy $\left(\bar{\mathbf{y}}_{gS}, P_g\right)$ can be better than the strategy $\left(\bar{\mathbf{y}}_S, P_3\right)$ if it is possible to cluster a population in such a way that the within-cluster spread of values of the multidimensional variable (under research) is bigger than the population spread of observations of these variables.

Let $\bar{\mathbf{y}}_S$ be the vector of the mean from the simple random sample of the size n, selected without replacement from a population of the size N. Its covariance matrix is of the following form:

$$\mathbf{V}(\bar{\mathbf{y}}_S, P_3) = \frac{N-n}{Nn}\mathbf{C}_* .$$

On the basis of the equations (4.3) and (4.26) we have:

$$\mathbf{V}(\bar{\mathbf{y}}_{gS}, P_g) = \frac{G-g}{GgM}\mathbf{C}_*\left(\mathbf{I} + \frac{N-G}{G-1}\boldsymbol{\Delta}\right). \tag{4.43}$$

Under the assumption that GM=N and gM=n:

$$\mathbf{V}(\bar{\mathbf{y}}_{gS}, P_g) = \frac{N-n}{Nn}\mathbf{C}_*\left(\mathbf{I} + \frac{N-G}{G-1}\boldsymbol{\Delta}\right).$$

Hence:

$$\mathbf{V}(\bar{\mathbf{y}}_S, P_3) - \mathbf{V}(\bar{\mathbf{y}}_{gS}, P_g) = -\frac{N-n}{Nn}\frac{N-G}{G-1}\mathbf{C}_*\boldsymbol{\Delta} .$$

or

$$\mathbf{V}(\bar{\mathbf{y}}_S, P_3) - \mathbf{V}(\bar{\mathbf{y}}_{gS}, P_g) = -\frac{N-n}{Nn}\frac{N-G}{G-1}(\mathbf{C}_* - \mathbf{C}_w)$$

This leads to the following theorem:

Theorem 4.4. If the matrix ($\mathbf{C}_*$-$\mathbf{C}_w$) is negative semi-definite (positive semi-definite) then the strategy $\mathbf{V}(\bar{\mathbf{y}}_{gS}, P_g)$ is not worse (better) than the strategy $\mathbf{V}(\bar{\mathbf{y}}_S, P_3)$. Particularly, if the matrix $\mathbf{C}_*$ is nonsingular and the $\boldsymbol{\Delta}$ is negative semi-definite (positive semi-definite) then the strategy $\mathbf{V}(\bar{\mathbf{y}}_{gS}, P_g)$ is not worse (better) than the strategy $\mathbf{V}(\bar{\mathbf{y}}_S, P_3)$.

Hence, the strategy $\mathbf{V}(\bar{\mathbf{y}}_{gS}, P_g)$ is not worse than the strategy $\mathbf{V}(\bar{\mathbf{y}}_S, P_3)$, if the within-cluster spread of a multidimensional variable represented by the matrix $\mathbf{C}_w$ is larger than its population spread represented by the matrix $\mathbf{C}$.

The relative efficiency coefficients are as follows:

$$e_{0i} = \frac{D^2(\bar{y}_{gSi}, P_g)}{D^2(\bar{y}_{Si}, P_3)} = 1 + \frac{N-G}{G-1}\delta(y_i), \ i=1,...,m. \tag{4.44}$$

where $\delta(y_i)$ expresses the formulas (4.28)-(4.30).

$$e_1 = \frac{\det \mathbf{V}(\bar{\mathbf{y}}_{gS}, P_g)}{\det \mathbf{V}(\bar{\mathbf{y}}_S, P_3)} = \det\left(\mathbf{I} + \frac{N-G}{G-1}\boldsymbol{\Delta}\right). \tag{4.45}$$

$$e_2 = \frac{tr\mathbf{V}(\bar{\mathbf{y}}_{gS}, P_g)}{tr\mathbf{V}(\bar{\mathbf{y}}_S, P_3)} = 1 + \frac{N-G}{G-1}\bar{\delta}\,. \tag{4.46}$$

where:

$$\bar{\delta} = \sum_{i=1}^{m} \delta(y_i)a_i\,, \qquad a_i = \frac{v_*(y_i)}{\sum_{i=1}^{m} v_*(y_i)}\,,$$

$$e_3 = \frac{\lambda_1(\bar{\mathbf{y}}_{gS}, P_g)}{\lambda_1(\bar{\mathbf{y}}_S, P_3)}\,. \tag{4.47}$$

Hence, on the basis of the theorem 4.4 and the well known matrix properties, we conclude that if if the matrix $\mathbf{C}_*$ is nonsingular and the matrix $\boldsymbol{\Delta}$ is negative definite, then $e_k<1$ for k=1,2,3 and $e_{0i}\leq 1$ for i=1,...,m and $e_{0j}<1$ for at least one index j=1,...,m.

On the basis of the expression (1.57) we evaluate the deff-coefficient:

$$deff(\bar{\mathbf{y}}_{gS}, P_g) = \mathbf{V}(\bar{\mathbf{y}}_{gS}, P_g)\mathbf{V}^{-1}(\bar{\mathbf{y}}_S, P_1)$$

$$deff(\bar{\mathbf{y}}_{gS}, P_g) = \rho\left(\frac{N-n}{N-1}\left(\mathbf{I} + \frac{N-G}{G-1}\boldsymbol{\Delta}\right)\right) \tag{4.48}$$

Hence, the $deff(\bar{\mathbf{y}}_{gS}, P_g)$ is equal to the maximal eigenvalue of the matrix $\frac{N-n}{N-1}\left(\mathbf{I} + \frac{N-G}{G-1}\boldsymbol{\Delta}\right)$. The eigenvalues of this matrix are evaluated in the following way:

$$\left|\frac{N-n}{N-1}\left(\mathbf{I} + \frac{N-G}{G-1}\boldsymbol{\Delta}\right) - \mathbf{I}\kappa\right| = 0$$

This equation is equivalent to the following

$$|\boldsymbol{\Delta} - \mathbf{I}\lambda| = 0$$

where:

$$\lambda = \frac{G-1}{N-G}\left(\frac{N-1}{N-n}\kappa - 1\right)$$

This and the theorem 4.2 lead to the following inequalities: $0 \le \kappa \le \frac{N-1}{G-1}$. Hence:

$$0 \le \text{deff}\left(\bar{\mathbf{y}}_{gS}, P_g\right) \le \frac{N-1}{G-1} \tag{4.49}$$

The strategy $\left(\bar{\mathbf{y}}_{gS}, P_g\right)$ can be better than the strategy $\left(\bar{\mathbf{y}}_S, P_3\right)$ if it is possible to cluster a population in such a way that the matrix ($\mathbf{C}_*$-$\mathbf{C}_w$) is negative definite. It means that the within-cluster spread of values of the multidimensional variable (under research) should be bigger than the population spread of observations of those variables. Hence, if it is possible a population should be clustered into such clusters that intra-cluster spread of variables under study is large as possible. The clustering algorithm proposed by Wywiał (2002a) can lead to realisation of this postulate.

4.5. Prediction of population average under the regression superpopulation model

Let us consider the regression superpopulation model defined by the equations (1.19)-(1.21). In the case of a one-dimensional variable under study this model is defined by the vector: $\mathbf{Y}=[Y_1...Y_N]$, where Y_k is attached to the k-th element of a population $\Omega=\{1,...,N\}$. The probability distribution of the vector $\mathbf{Y}$ has the following properties :

$$Y_k=\beta_o+\mathbf{x}_k\boldsymbol{\beta}+U_k,\ k=1,...,N \tag{4.50}$$

where: $\mathbf{x}_k=[x_{k1}...x_{km}]$ is a vector of observations of auxiliary variables. It is attached to the k-th element of a population. Let $\boldsymbol{\beta}^T=[\beta_1...\beta_m]$ be the vector of regression parameters. Let us assume that:

$$E(Y_k)=\mu_k=\beta_o+\mathbf{x}_k\boldsymbol{\beta}, \qquad E(U_k)=0,$$

$$D^2(Y_k)=D^2(U_k)=\sigma^2, \qquad \text{Cov}(Y_k,Y_l)=\text{Cov}(U_k,U_l)=0$$

where: $k \ne l=1,...,N$.

The problem is the prediction of the population mean defined by the following expression:

$$\bar{Y}=\frac{1}{N}\sum_{i=1}^{N} Y_i .$$

It is predicted by means of the strategy $\left(\bar{Y}_{gS}, P_g\right)$, where:

$$\overline{Y}_{gS} = \frac{1}{Mg}\sum_{p\in S} Z_p \tag{4.51}$$

where:

$$Z_p = \sum_{k\in\Omega_p} Y_k \ .$$

Wywial (1993) proved that $\left(\overline{Y}_{gS}, P_g\right)$ is p-ξ unbiased predictor of the mean $\overline{Y}$ and

$$EE\left(\overline{Y}_{gS} - \overline{Y}\right)^2 = \frac{G-g}{gN}\left(\frac{1}{M}\boldsymbol{\beta}^T\mathbf{C}_*(\mathbf{a})\boldsymbol{\beta} + \sigma^2\right) \tag{4.52}$$

where:

$$\mathbf{C}_*(\mathbf{a}) = \left[c_*(a_i, a_j)\right], \qquad c_*(a_i, a_j) = \frac{1}{G-1}\sum_{k=1}^{G}\left(a_{ik} - \bar{a}_i\right)\left(a_{jk} - \bar{a}_j\right),$$

$$a_{ik} = \sum_{t\in\Omega_k} x_{ti}\,, \qquad \bar{a}_i = \frac{1}{M}\sum_{k=1}^{G} a_{ik}\ .$$

E(.) is the expected value evaluated on the basis of probability distribution determining the superpopulation model and E(.) is determined by sampling design.

If $\mathbf{e} = \dfrac{\boldsymbol{\beta}}{\sqrt{\boldsymbol{\beta}^T\boldsymbol{\beta}}}$, $EE\left(\overline{Y}_{gS} - \overline{Y}\right)^2 = \dfrac{G-g}{gN}\left(\dfrac{\boldsymbol{\beta}^T\boldsymbol{\beta}}{M}\mathbf{e}^T\mathbf{C}_*(\mathbf{a})\mathbf{e} + \sigma^2\right)$.

This and the property:

$$\lambda_1 = \mathbf{e}_1^T\mathbf{C}_*(\mathbf{a})\mathbf{e}_1 = \max_{\mathbf{e}^T\mathbf{e}=1}\left\{\mathbf{e}^T\mathbf{C}_*(\mathbf{a})\mathbf{e}\right\}$$

lead to the following inequality:

$$EE\left(\overline{Y}_{gS} - \overline{Y}\right)^2 \le \frac{G-g}{gN}\left(\frac{\boldsymbol{\beta}^T\boldsymbol{\beta}}{M}\,\lambda_1 + \sigma^2\right). \tag{4.53}$$

Let us consider the regression model for h-variables under study. It is described by the expression (1.20)-(1.21) for ρ=0. Let us additionally assume that $Cov(U_{ij},U_{tk})=0$ for $i\neq t=1,...,N$ and $j,k=1,...,h$ and $D^2(U_{ij})=\sigma_j^2$ for i=1,...,N. Our purpose is prediction of the average vector $\overline{\mathbf{Y}} = \left[\overline{Y}_1 ... \overline{Y}_m\right]$ by means of the vector $\overline{\mathbf{Y}}_{gS} = \left[\overline{Y}_{1gS} ... \overline{Y}_{mgS}\right]$, where $\overline{Y}_{jgS}$, j=1,...,h, is determined by the expression (4.51) for the j-th variable under study. The strategy $\left(\overline{\mathbf{Y}}_{gS}, P_g\right)$ is p-ξ unbiased for the vector $\overline{\mathbf{Y}}$ and

$$E\mathrm{E}\left(\overline{\mathrm{Y}}_{\mathrm{jgS}}-\overline{\mathrm{Y}}_{\mathrm{j}}\right)^2=\frac{\mathrm{G}-\mathrm{g}}{\mathrm{gN}}\left(\frac{1}{\mathrm{M}}\boldsymbol{\beta}_{*\mathrm{j}}^{\mathrm{T}}\mathbf{C}_{*}(\mathbf{a})\boldsymbol{\beta}_{*\mathrm{j}}+\sigma_{\mathrm{j}}^2\right) \tag{4.54}$$

and

$$E\mathrm{E}\left(\overline{\mathrm{Y}}_{\mathrm{gS}}-\overline{\mathrm{Y}}\right)^2\le\frac{\mathrm{G}-\mathrm{g}}{\mathrm{gN}}\left(\frac{\boldsymbol{\beta}_{*\mathrm{j}}^{\mathrm{T}}\boldsymbol{\beta}_{*\mathrm{j}}}{\mathrm{M}}\ \lambda_1+\sigma_{\mathrm{j}}^2\right) \tag{4.55}$$

for j=1,...,h. Hence a population should be clustered into G clusters (each of the size M) in such a way that the maximal eigenvalue of the variance-covariance matrix $\mathbf{C}_*(\mathbf{a})$ is minimal. Let us note that Wywiał (1991) considered the problem of the prediction in the particular case when an auxiliary variable is one-dimensional.

4.6. Clustering algorithm

Let $W=\{\Omega_p\}$ be a set of such G disjoint clusters of a fixed population that $\bigcup_{p=1}^{G}\Omega_p=\Omega$. Each cluster consists of the same number M of elements. Our purpose is to find such a set $\underline{W}$ where a criterion function $f(\underline{W})$=minimum. Particularly:

$$f_1(W)=\mathrm{tr}\mathbf{C}_*(z),\ f_2(W)=\det\mathbf{C}_*(z)\text{ or }f_3(W)=\lambda_1(\mathbf{C}_*(z)) \tag{4.56}$$

where $\mathbf{C}_*(z)$ is expressed by the beginning of the paragraph 4.1. It is easy to demonstrate that if $f_i(\underline{W})$=minimum, the relative efficiency coefficient e_i=minimum, i=1,2,3, too. To determine the set $\underline{W}$, we can construct the following iterative algorithm. Let $W_0=\{\Omega_{10},...,\Omega_{G0}\}$ be an arbitrary start of divisions of the population. Let $\underline{W}_t=\{\underline{\Omega}_{1t},...,\underline{\Omega}_{Gt}\}$ be the set of clusters resulting from the t-th iteration of the clustering algorithm. Let h-th and k-th population elements belong to the clusters $\underline{\Omega}_{i,t}$ and $\underline{\Omega}_{j,t}$, respectively. So, $h\in\underline{\Omega}_{i,t}$ and $k\in\underline{\Omega}_{j,t}$. Cluster created during the (t+1)-th iteration are denoted by $\Omega_{i,t+1}(h,k)$ and $\Omega_{j,t+1}(k,h)$. They are obtained in the following way:

$$\Omega_{i,t+1}(h,k)=\underline{\Omega}_{i,t}-\{h\}\cup\{k\},\qquad \Omega_{j,t+1}(k,h)=\underline{\Omega}_{j,t}-\{k\}\cup\{h\}. \tag{4.57}$$

Hence, the h-th population element is moved from the cluster $\underline{\Omega}_{i,t}$ to the cluster $\underline{\Omega}_{j,t}$ and the k-th ($k\neq h$) element is moved from the cluster $\underline{\Omega}_{j,t}$ to the cluster $\underline{\Omega}_{i,t}$. Let us introduce the following set:

$$W_{k+1}(h,k)=\{\underline{W}_t-\underline{\Omega}_{i,t}-\underline{\Omega}_{j,t},\ \Omega_{i,t+1}(h,k),\ \Omega_{j,t+1}(h,k)\}. \tag{4.58}$$

Hence, at the end of the (t+1)-th iteration, the optimal set of clusters is obtained through the minimization of the criterion function in the following way:

$$\underline{W}_{k+1}= W_{k+1}(\underline{h},\underline{k}): f_a(W_{k+1}(\underline{h},\underline{k}))= \min_{k=1,\dots,N;k\neq h} \operatorname{imum} \min_{h=1,\dots,N} \operatorname{imum}\{f_a(W_{t+1}(h,k))\} \tag{4.59}$$

where the criterion function f_a can be chosen according to one of the functions defined in the expression (4.56). The iterative clustering algorithm should be continued until the time when any two elements of a population are not moved from one cluster to another or the number of the iterations reaches the admissible level which is usually assigned arbitrarily.

Sarndal, Swenson and Wretman (1992) considered the population which consists of 284 municipalities in Sweden. They considered the data on three variables observed in this population: revenues from the 1985 municipal taxation (in millions of kronor), number of municipal employees in 1984, real estate values according to 1984 assessment (in millions of kronor). We denote them by y_1, y_2, y_3, respectively.

The mean value of these variables can be estimated on the basis of the vector of a simple sample mean or the vector of the simple cluster sample. Both samples are drawn without replacement. We are going to compare the accuracy of these strategies on the basis of the above results.

The population means of the variables y_1, y_2, y_3 are $\bar{\mathbf{y}} = [\bar{y}_1 \quad \bar{y}_2 \quad \bar{y}_3] = [245.3415 \quad 1774.1585 \quad 3073.6585]$. The correlation matrix of these variables is as follows:

$$\mathbf{R} = \begin{bmatrix} 1.0000 & .9988 & .9356 \\ .9988 & 1.0000 & .9395 \\ .9356 & .9395 & 1.0000 \end{bmatrix}$$

The population was divided into 71 clusters. Each cluster consists of 4 elements. Let us consider two divisions of this population into clusters. The first division is quite arbitrary because the clusters consist of elements which are usually neighbours (see: Särndal at al (1992)). This division is denoted by W_0. The second division will be called an optimal one because it was obtained in the following way: the population is divided according to the criterion expressed by the equation (4.59) where $f_a=f_3$ is given by the expression (4.56). Hence, clusters are selected in such a way that the maximal eigenvalue of the variance-covariance matrix $\mathbf{C}(z)$ approaches the minimum. The optimal division of the population is denoted by $\underline{W}$. The table 4.1 represents the results of the estimation of the mean value on the basis of the strategies $(\bar{\mathbf{y}}_S, P_3)$ and $(\bar{\mathbf{y}}_{gS}, P_g)$. The second strategy is considered for two divisions of the population into clusters. They have been denoted by W_0 and $\underline{W}$, respectively. In the last two columns of the table 1 there are values of the relative efficiency coefficients defined by the expressions (4.44)-(4.47) in the cases of the divisions W_0 and $\underline{W}$, respectively. The simple sample strategy without replacement $(\bar{\mathbf{y}}_S, P_3)$ is considered for the sample size n=40. The strategy $(\bar{\mathbf{y}}_{gS}, P_g)$ is studied under the sample size g=10 and the size of clusters M=4. Then n=Mg.

Table 4.1

accuracy parameter	$\mathbf{T} = \bar{\mathbf{y}}_S, P_3$	$\mathbf{T} = \bar{\mathbf{y}}_{gS}, P_g, W_0$	$\mathbf{T} = \bar{\mathbf{y}}_{gS}, P_g, \underline{W}$	$e_i(W_0)$	$e_i(\underline{W})$
$D(T_1)$	87.4	91.8	76.7	90.6	129.9
$D(T_2)$	623.5	654.7	544.1	90.7	131.3
$D(T_3)$	695.8	766.4	533.9	82.4	169.8
q(V(T))	938.3	1012.2	766.1	92.7	122.5
det(V(T))	3.83E+11	6.072E+11	2.522E+11	63.1	152.0
$\lambda_1(V(T))$	854056.8	992954.4	563846.8	86.0	151.5

The analysis of accuracy of estimation results, recorded in the table 4.1, leads to the following conclusions. In the case of the division W_0 of the population, the strategy $(\bar{\mathbf{y}}_S, P_3)$ is better than the strategy $(\bar{\mathbf{y}}_{gS}, P_g)$. In the case of the optimal division $\underline{W}$, the strategy $(\bar{\mathbf{y}}_{gS}, P_g)$ is better than the strategy $(\bar{\mathbf{y}}_S, P_3)$. Then, the special clustering method, introduced here, leads to an increase in the relative efficiency of the simple cluster sampling strategy.

4.7. Two phase sampling for clustering

A simple sample w of the size n is selected without replacement from a population Ω of size N. In this sample, auxiliary variables are observed. The space of these samples is denoted by **W**. On the basis of the auxiliary variables, the sample w is divided into G clusters: $\Omega_h(w)$, h=1,...,G. The clusters are disjoint and each of them is of the same size m. The problem of a clustering criterion will be considered later. Next, the simple cluster sample s of the size g is selected from the set of clusters $\{\Omega_h(w), h=1,...,G\}$ Finally, values of the variable under study are observed in the sample s. Let us note that s,w are outcomes of random samples S and W, respectively.

The population mean $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ is estimated by means of the statistic:

$$\tilde{y}_S = \frac{1}{gm}\sum_{p\in S}\sum_{k\in\Omega_p(w)} y_k = \frac{1}{gm}\sum_{p\in S} z_p \tag{4.60}$$

where:

$$z_p = \sum_{k\in\Omega_p(w)} y_k \,.$$

The sample mean is denoted by:

$$\bar{y}_w = \frac{1}{n}\sum_{k\in w} y_k \,. \tag{4.61}$$

The expected value of the statistic $\tilde{y}_S$ is derived as follows:

$$E(\tilde{y}_S) = E_W E_{S/W}(\tilde{y}_S) = E_W(\bar{y}_W) = \bar{y} \,. \tag{4.62}$$

Hence, the statistic $\tilde{y}_S$ is a p-unbiased estimator of the average $\bar{y}$. Its mean square error is derived in the following way:

$$E(\tilde{y}_S - \bar{y})^2 = E_W E_{S/W}\left[(\tilde{y}_S - \bar{y}) + (\bar{y}_W - \bar{y})\right]^2 =$$

$$= E_W D^2_{S/W}(\tilde{y}_S \mid W) + D^2_W(\bar{y}_W) \,. \tag{4.63}$$

The conditional variance and its expected value can be expressed as follows[29]:

$$D^2_{S/w}(\tilde{y}_S \mid w) = D^2_{S/w}(\tilde{y}_S \mid W = w) = \frac{G-g}{Ggm} v_*(z \mid w) \tag{4.64}$$

where:

$$v_*(z \mid w) = \frac{1}{g-1}\sum_{k\in S}(z_k - \bar{z}_S)^2, \qquad \bar{z}_S = \frac{1}{g}\sum_{k\in S} z_k \,, \tag{4.65}$$

$$E_W D^2_{S/W}(\tilde{y}_S \mid W) = \binom{N}{n}^{-1} \frac{G-g}{Ggm} \sum_{w\in \mathbf{W}} v_*(z \mid w) \,. \tag{4.66}$$

It is well known that the variance of the simple sample mean is as follows:

$$D^2_W(\bar{y}_W) = \frac{N-n}{Nn} v_*(y) \tag{4.67}$$

where:

$$v_*(y) = \frac{1}{N-1}\sum_{k\in\Omega}(y_k - \bar{y})^2, \qquad \bar{y} = \frac{1}{N}\sum_{k\in\Omega} y_k \,. \tag{4.68}$$

Then on the basis of the expression (4.63), we have:

$$D^2(\tilde{y}_S) = \frac{G-g}{Ggm} E_W(v_*(z \mid w)) + \frac{N-n}{Nn} v_*(y) \,. \tag{4.69}$$

[29] See e.g. Särndal, Swenson, Wretman (1992).

The following statistic is an unbiased estimator of the variance $D^2(\tilde{y}_S)$:

$$\hat{D}^2(\tilde{y}_S) = \frac{G-g}{Ggm} v_*(z \mid W) + \frac{N-n}{Nn} v_*(y \mid W) \tag{4.70}$$

where:

$$v_*(y \mid w) = \frac{1}{n-1} \sum_{k \in w} (y_k - \bar{y}_w)^2 . \tag{4.71}$$

Let us denote a simple sample of the size mg by A. The sample mean is as follows:

$$\bar{y}_A = \frac{1}{mg} \sum_{k \in A} y_k . \tag{4.72}$$

Its variance expressed by the following formula:

$$D_A^2(\bar{y}_A) = \frac{N-mg}{Nmg} v_*(y) . \tag{4.73}$$

Then, the expressions (4.69), (4.73) and the assumption that N=mH lead to the result:

$$D^2(\tilde{y}_S) - D^2(\bar{y}_A) = \frac{G-g}{Ggm} E_W\left(v_*(z \mid w)\right) - \frac{mH-mg}{mHmg} v_*(y) ,$$

$$D^2(\tilde{y}_S) - D^2(\bar{y}_A) = \frac{1}{n}\left[\left(1-\frac{g}{G}\right) E_W\left(v_*(z \mid w)\right) - \left(1-\frac{g}{H}\right) v_*(y)\right]. \tag{4.74}$$

Hence, the estimator $\tilde{y}_S$ is more precise than the estimator $\bar{y}_A$ if

$$E_W\left(v_*(z \mid w)\right) < \kappa v_*(y) , \text{ where: } \kappa = \left(1-\frac{g}{G}\right)^{-1}\left(1-\frac{g}{H}\right) > 1 .$$

In conclusion, each sample w should be divided into such clusters of the same size m that $v_*(z \mid w)$ takes a minimal value.

The well known regression superpopulation model is being considered. This model is defined by the expressions (4.50). It is determined as vector $\mathbf{Y}=[Y_1...Y_N]$, where Y_k is attached to the k-th element of a population $\Omega=\{1,...,N\}$.

The problem is the prediction of the population average $\overline{Y} = \sum_{i=1}^{N} Y_i$ by means of the statistic:

$$\tilde{Y}_{gS} = \frac{1}{gm} \sum_{p \in S} \sum_{k \in \Omega_p} Y_k = \frac{1}{gm} \sum_{p \in S} Z_p \ . \tag{4.75}$$

The expected value of the statistic $\tilde{Y}_{gS}$ is derived as follows:

$$E(\tilde{Y}_{gS}) = E_W E_{S/W}(\tilde{Y}_{gS}) = E_W(\overline{Y}_W) = \overline{Y} \ ,$$

$$E E(\tilde{Y}_{gS}) = E(\overline{Y}) = \frac{1}{N} \sum_{k=1}^{N} \mu_k = \mu. \tag{4.76}$$

where $E(.)$ is the expected value evaluated on the basis of probability distribution determining the superpopulation. Hence, the statistic $\tilde{Y}_{gS}$ is a p-unbiased and p-ξ unbiased predictor of the average $\overline{Y}$. Its mean square error is derived on the basis of the expressions (4.50) and (4.63) in the following way:

$$E E(\tilde{Y}_{gS} - \overline{Y})^2 = E E_W E_{S/W}[(\tilde{Y}_{gS} - \overline{Y}_W) + (\overline{Y}_W - \overline{Y})]^2$$

$$E E(\tilde{Y}_{gS} - \overline{Y})^2 = E E_W D^2_{S/W}(\tilde{Y}_{gS} | W) + E D^2_W(\overline{Y}_W) = E D^2(\tilde{y}_S), \tag{4.77}$$

$$E D^2(\tilde{y}_S) = \frac{G - g}{Ggm} E E_W(v_*(z \mid w)) + \frac{N - n}{Nn} E(v_*(y)) \ . \tag{4.78}$$

The expected value of the conditional variance $v_*(z|w)$ (computed on the basis of the superpopulation distribution) can be expressed in the following way[30]:

$$E(v_*(Z \mid w)) = \frac{1}{m} \boldsymbol{\beta}^T \mathbf{C}_*(a|w) \boldsymbol{\beta} + \sigma^2 \tag{4.79}$$

where a is an m-dimensional auxiliary variable whose a k-th observation is as follows:

$$a_k = \sum_{i \in U_k} x_i \ ,$$

$$E(v_*(Y)) = \boldsymbol{\beta}^T \mathbf{C}_*(X) \boldsymbol{\beta} + \sigma^2 \tag{4.80}$$

[30] See Wywiał (1993).

The variance-covariance matrix of the variable *a* in the sample w is as follows:

$$\mathbf{C}_*(a \mid w) = [c_*(a_i a_j \mid w)]$$

where:

$$c_*(a_i, a_j \mid w) = \frac{1}{g-1}\sum_{p\in S}\left(a_{ip} - \bar{a}_{wi}\right)\left(a_{jp} - \bar{a}_{wj}\right),$$

$$\bar{a}_{wi} = \frac{1}{g}\sum_{p\in S} a_{ip}\ .$$

The population variance-covariance matrix of auxiliary variables is defined by the expression:

$$\mathbf{C}_*(\mathbf{X}) = [c_*(x_i x_j)]$$

where:

$$c_*(x_i x_j) = \frac{1}{N-1}\sum_{p\in\Omega}\left(x_{ip} - \bar{x}_i\right)\left(x_{ip} - \bar{x}_i\right), \qquad \bar{x}_i = \frac{1}{N}\sum_{p\in\Omega} x_{ip}\ .$$

The formulas (4.63)-(4.69), (4.79) and (4.80) lead to the following mean square prediction error:

$$E\mathrm{E}(\tilde{\bar{Y}}_{gS} - \bar{Y})^2 = \binom{N}{n}^{-1}\sum_{w\in W}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{K}(w)\boldsymbol{\beta} + \left(\frac{G-g}{Ggm} + \frac{N-n}{Nn}\right)\sigma^2 \tag{4.81}$$

where:

$$\mathbf{K}(w) = \frac{G-g}{Ggm^2}\mathbf{C}_*(a|w) + \frac{N-n}{Nn}\mathbf{C}_*(\mathbf{X}). \tag{4.82}$$

Hence:

$$E\mathrm{E}(\tilde{\bar{Y}}_{gS} - \bar{Y})^2 = \binom{N}{n}^{-1}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta}\sum_{w\in \mathbf{W}}\boldsymbol{\varepsilon}\mathbf{K}(w)\boldsymbol{\varepsilon}^{\mathrm{T}} + \left(\frac{G-g}{Ggm} + \frac{N-n}{Nn}\right)\sigma^2$$

where:

$$\boldsymbol{\varepsilon} = \frac{\boldsymbol{\beta}}{\sqrt{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta}}}\ .$$

Let $\kappa_1(w)$ be the spectral radius (the maximal eigenvalue) of the matrix $\mathbf{K}(w)$. Hence, the well known properties of matrix eigenvalues lead to the following inequality:

$$E\mathrm{E}(\tilde{\mathrm{Y}}_{\mathrm{gS}} - \overline{\mathrm{Y}})^2 \le \binom{\mathrm{N}}{\mathrm{n}}^{-1} \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta} \sum_{\mathrm{w}\in\mathbf{W}} \boldsymbol{\kappa}_1(\mathrm{w}) + \left(\frac{\mathrm{G}-\mathrm{g}}{\mathrm{Ggm}} + \frac{\mathrm{N}-\mathrm{n}}{\mathrm{Nn}}\right)\sigma^2 \qquad (4.83)$$

The conclusion is: the sample w should be clustered in such a way that $\boldsymbol{\kappa}_1(\mathrm{w})$ takes the minimal value. In the previous paragraph we considered the clustering algorithm which, in our case, can be implemented to find the optimal division of the sample w into disjoint and of the same size clusters.

4.8. Conditional intra-cluster coefficient of correlation

In the paragraph 4.2, the matrix of coefficients of within-cluster correlation of a multidimensional variable $X=[x_1...x_\mathrm{m}]$ is considered. In our case, when auxiliary variables are observed in the sample w, it is denoted by: $\mathbf{R}^{(\mathrm{w})}(X \mid \mathrm{w}) = \left[\mathrm{r}_{\mathrm{ij}}^{(\mathrm{w})}(X \mid \mathrm{w})\right]$ where the coefficient of within-cluster correlation between variables x_i and x_j (i,j=1,...,m) is defined by the following expression:

$$\mathrm{r}_{\mathrm{ij}}^{(\mathrm{w})}(X \mid \mathrm{w}) = \frac{\mathrm{c}_{*\mathrm{ij}}^{(\mathrm{w})}(X \mid \mathrm{w})}{\mathrm{d}_{*\mathrm{i}}(X \mid \mathrm{w})\mathrm{d}_{*\mathrm{j}}(X \mid \mathrm{w})} \qquad (4.84)$$

where the within-cluster covariance of the variables in the sample w has the form:

$$\mathrm{c}_{*\mathrm{ij}}^{(\mathrm{w})}(X \mid \mathrm{w}) = \frac{1}{(\mathrm{n}-1)(\mathrm{m}-1)} \sum_{\mathrm{k}=1}^{\mathrm{G}} \sum_{\mathrm{p}\neq\mathrm{h}\in\mathrm{U}_\mathrm{k}} \left(\mathrm{x}_{\mathrm{ip}} - \overline{\mathrm{x}}_{\mathrm{wi}}\right)\left(\mathrm{x}_{\mathrm{jh}} - \overline{\mathrm{x}}_{\mathrm{wj}}\right). \qquad (4.85)$$

The variance of the variable x_i is defined by the expression:

$$\mathrm{d}_{*\mathrm{i}}^2(X \mid \mathrm{w}) = \frac{1}{\mathrm{n}-1} \sum_{\mathrm{p}\in\mathrm{w}} (\mathrm{x}_{\mathrm{ip}} - \overline{\mathrm{x}}_{\mathrm{iw}})^2 .$$

Let $\mathbf{D}_*(X \mid \mathrm{w}) = \left[\mathrm{d}_{*\mathrm{i}}^2(X \mid \mathrm{w})\right]$ be the diagonal matrix of variances of auxiliary variables from the sample w and let $\mathbf{R}(X|\mathrm{w})=[\mathrm{r}_{\mathrm{ij}}(X|\mathrm{w})]$ be the correlation matrix of these variables, where:

$$\mathrm{r}_{\mathrm{ij}} = \frac{\mathrm{d}_{*\mathrm{ij}}(X \mid \mathrm{w})}{\mathrm{d}_{*\mathrm{i}}(X \mid \mathrm{w})\mathrm{d}_{*\mathrm{j}}(X \mid \mathrm{w})},$$

$$\mathrm{d}_{*\mathrm{ij}}(X \mid \mathrm{w}) = \frac{1}{\mathrm{n}-1} \sum_{\mathrm{p}\in\mathrm{w}} \left(\mathrm{x}_{\mathrm{ip}} - \overline{\mathrm{x}}_{\mathrm{wi}}\right)\left(\mathrm{x}_{\mathrm{jp}} - \overline{\mathrm{x}}_{\mathrm{wj}}\right), \quad \mathrm{d}_{*\mathrm{i}}^2 = \mathrm{d}_{*\mathrm{ii}} .$$

Under the assumption (see the paragraph 4.2) that

$$h_{ik} = \frac{x_{ik} - \bar{x}_{wi}}{d_{*i}(X \mid w)}, \tag{4.86}$$

we have:

$$\mathbf{R}^{(w)}(X \mid w) = \frac{1}{(n-1)(m-1)} \mathbf{h}^T (\mathbf{B}\mathbf{B}^T - \mathbf{I}_n)\mathbf{h}, \tag{4.87}$$

where:

$$\mathbf{h} = [h_{ij}] = \mathbf{A}\mathbf{x}\mathbf{D}_*^{-1/2}(X \mid w), \tag{4.88}$$

$$\mathbf{A} = \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\mathbf{J}_n^T, \qquad \mathbf{B} = \mathbf{I}_n \otimes \mathbf{J}_n^T.$$

Let **P** be such an orthogonal matrix that

$$\mathbf{P}^T\mathbf{R}^{(w)}(X \mid w)\mathbf{P} = \mathbf{D}^{(w)}(X \mid w)$$

where $\mathbf{D}^{(w)}(X \mid w)$ is the diagonal matrix which consists of the eigenvalues of the matrix $\mathbf{R}^{(w)}(X|w)$. They are denoted by $d_i^{(w)}(X \mid w)$ (i=1,...,m). Hence, on the basis of the expression (4.87), we have:

$$\mathbf{D}^{(w)}(X \mid w) = \frac{1}{(n-1)(m-1)} \mathbf{t}^T (\mathbf{B}^T\mathbf{B} - \mathbf{I}_n)\mathbf{t} \tag{4.89}$$

where:

$$\mathbf{t} = \mathbf{A}\mathbf{x}\mathbf{D}^{-½}\mathbf{P}. \tag{4.90}$$

Hence, the diagonal element $d_i^{(w)}(X \mid w)$ of the matrix $\mathbf{D}^{(w)}(X \mid w)$ is the coefficient of within-cluster correlation of an i-th variable, whose observations create the i-th column of the matrix **t**. Then, $d_i^{(w)}(X \mid w) = r_{ii}^{(w)}(X \mid w)$. The matrix **t** is the linear transformation (expressed by the formula (4.90)) of the data matrix **x**. The well known properties of the coefficient of within-cluster correlation (see. e.g. Konijn (1973)) lead to the conclusion:

$$-\frac{1}{m-1} \le d_i^{(w)} \le 1, \qquad i = 1, \ldots, m \tag{4.91}$$

where:

$$d_i^{(w)} = 1 - \frac{v_w(t_i)}{v(t_i)}, \tag{4.92}$$

$$v(t_i) = \frac{1}{n}\sum_{p=1}^{G} \sum_{k\in\Omega_p} (t_{ik} - \bar{t}_i)^2, \qquad \bar{t}_i = \frac{1}{n}\sum_{p=1}^{G} \sum_{k\in\Omega_p} t_{ik},$$

$$v_w(t_i) = \frac{1}{G(m-1)}\sum_{p=1}^{G} \sum_{k\in\Omega_p} (t_{ik} - \bar{t}_{ip})^2, \qquad \bar{t}_{ip} = \frac{1}{m}\sum_{k\in\Omega_p} t_{ik}.$$

The within-cluster variance is denoted by $v_w(t_i)$. The coefficient $d_i^{(w)}$ is the function of the within-cluster variance and the ordinary variance. Hence, it assesses the degree of within-cluster homogeneity of observation in the sample w.

Similarly to the one-dimensional case (see e.g. Konijn (1973), p. 225-227), the following properties can be derived:

$$\mathbf{R}^{(w)}(X \mid w) = \mathbf{D}^{-1/2}(X \mid w)\left(\mathbf{C}_m(X \mid w) - \frac{1}{M}\mathbf{C}_w(X \mid w)\right)\mathbf{D}^{-1/2}(X \mid w) \tag{4.93}$$

or:

$$\mathbf{R}^{(w)}(X \mid w) = \mathbf{D}^{-1/2}(X \mid w)\left(\mathbf{C}(X \mid w) - \mathbf{C}_w(X \mid w)\right)\mathbf{D}^{-1/2}(X \mid w) \tag{4.94}$$

or:

$$\mathbf{R}^{(w)}(X \mid w) = \frac{1}{M-1}\mathbf{D}^{-1/2}(X \mid w)\left(M\mathbf{C}_m(X \mid w) - \mathbf{C}(X \mid w)\right)\mathbf{D}^{-1/2}(X \mid w) \tag{4.95}$$

where $\mathbf{C}(X \mid w) = \frac{n-1}{n}\mathbf{C}_*(X \mid w)$, $\mathbf{D}(X \mid w) = \frac{n-1}{n}\mathbf{D}_*(X \mid w)$ and $\mathbf{C}_m(X \mid w) = [c_m(x_i, x_j \mid w)]$ is the between-cluster matrix of the covariances, where:

$$c_m(x_i, x_j \mid w) = \frac{1}{G}\sum_{p=1}^{G}(\bar{x}_{ip} - \bar{x}_i)(\bar{x}_{jp} - \bar{x}_j), \quad \bar{x}_{wi} = \frac{1}{n}\sum_{p=1}^{G} \sum_{k\in\Omega_p(w)} x_{ik}.$$

The within-cluster matrix of the covariances is denoted by $\mathbf{C}_w(X \mid w) = [c_w(x_i, x_j \mid w)]$, where:

$$c_w(x_i, x_j) = \frac{1}{G(m-1)}\sum_{p=1}^{G} \sum_{k\in\Omega_p} (x_{ik} - \bar{x}_{ip})(x_{jk} - \bar{x}_{jp}).$$

The well known properties of the matrix determinant and the expression (4.94) allow us to derive the following conclusion. Let the matrices $\mathbf{C}(X|w)$ and $\mathbf{C}_w(X|w)$ be nonnegative definite. Hence, if the matrix $\mathbf{R}^{(w)}(X|w)$ is nonnegative definite, then $\mathbf{C}(X|w)$ $-\mathbf{C}_w(X|w)$ is nonnegative definite and $d_i^{(w)}(X \mid w) \geq 0$, for i=1,...,m. If $\mathbf{R}^{(w)}(X|w)$ is nonpositive definite, then $\mathbf{C}(X|w)$ $-\mathbf{C}_w(X|w)$ is nonpositive and $d_i^{(w)}(X \mid w) \leq 0$ for i=1,...,m. If the matrix $\mathbf{R}^{(w)}(X|w)$ is positive (negative) definite, then $d_i^{(w)}(X \mid w) \geq 0$ ($d_i^{(w)}(X \mid w) \leq 0$) for all i=1,...,m and there exists at least one index k=1,...,m that $d_i^{(w)}(X \mid w) > 0$ ($d_i^{(w)}(X \mid w) < 0$).

We can say that the within-cluster spread of observations of a multidimensional variable is less than their spread in the sample w if the matrix $\mathbf{R}^{(w)}(X|w)$ is positive definite. When $\mathbf{R}^{(w)}(X|w)$ is negative definite, we say that the spread of values of a multidimensional variable in the sample w is less than their within-cluster spread.

The following expressions can be derived[31]:

$$\mathbf{C}_*(a \mid w) = \frac{n-1}{G-1}\left[\mathbf{C}_*(X) + (m-1)\mathbf{C}_*^{(w)}(X \mid w)\right], \tag{4.96}$$

$$\mathbf{C}_*^{(w)}(a \mid w) = \mathbf{D}_*^{1/2}(X \mid w)\mathbf{R}^{(w)}(X \mid w)\mathbf{D}_*^{1/2}(X \mid w), \tag{4.97}$$

$$\mathbf{C}_*(a \mid w) = \frac{n-1}{G-1}\mathbf{D}_*^{1/2}(X \mid w)\left[\mathbf{R}(X \mid w) + (m-1)\mathbf{R}^{(w)}(X \mid w)\right]\mathbf{D}_*^{1/2}(X \mid w). \tag{4.98}$$

This result and the expressions (4.81) and (4.82) lead to the following formula:

$$E\mathrm{E}(\tilde{\bar{Y}}_{gS} - \bar{Y})^2 = \left(\frac{n-1}{G-1}\frac{G-g}{Ggm^2} + \frac{N-n}{Nn}\right)\boldsymbol{\beta}^T\mathbf{C}_*(X)\boldsymbol{\beta} +$$

$$+ \frac{(n-1)(m-1)(G-g)}{(G-1)Ggm^2}\binom{N}{n}^{-1}\sum_{w\in\mathbf{W}}\boldsymbol{\beta}^T\mathbf{C}^{(w)}(X \mid w)\boldsymbol{\beta} +$$

$$+ \frac{N-mg}{Nmg}\sigma^2 \tag{4.99}$$

Hence, each sample w has to be clustered in such a way that the quadratic form $\boldsymbol{\beta}^T\mathbf{C}^{(w)}(X \mid w)\boldsymbol{\beta}$ is negative definite and its value is as short as possible. It is similar to dividing the sample w into such clusters that all eigenvalues of the matrix $\mathbf{C}^{(w)}(X \mid w)$ are negative and they take values as short as possible.

[31] See Wywial (1992, 1995).

Let u be a simple random sample selected without replacement from a fixed population. The statistic

$$\overline{\mathrm{Y}}_{\mathrm{a}} = \frac{1}{\mathrm{mg}} \sum_{\mathrm{k}\in\mathrm{a}} \mathrm{Y}_{\mathrm{k}} \tag{4.100}$$

is a p-unbiased and p-ξ-unbiased predictor of the average $\overline{\mathrm{Y}}$ and

$$E\,\mathrm{D}^2\left(\overline{\mathrm{Y}}_{\mathrm{a}}\right) = \frac{\mathrm{N}-\mathrm{mg}}{\mathrm{Nmg}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{C}(X)\boldsymbol{\beta}+\sigma^2) \tag{4.101}$$

This formula and expressions (4.97) and (4.99) lead to the following:

$$E\mathrm{E}(\tilde{\mathrm{Y}}_{\mathrm{gS}} - \overline{\mathrm{Y}})^2 - E\,\mathrm{D}^2\left(\overline{\mathrm{Y}}_{\mathrm{a}}\right) = \frac{1}{\mathrm{m}}\left(\frac{\mathrm{n}-1}{\mathrm{G}-1}\frac{\mathrm{G}-\mathrm{g}}{\mathrm{Ggm}} + \frac{1}{\mathrm{G}} - \frac{1}{\mathrm{g}}\right)\boldsymbol{\beta}^{\mathrm{T}}\mathbf{C}_{*}(X)\boldsymbol{\beta} +$$

$$+ \frac{(\mathrm{n}-1)(\mathrm{m}-1)(\mathrm{G}-\mathrm{g})}{(\mathrm{G}-1)\mathrm{Ggm}^2}\,\boldsymbol{\beta}^{\mathrm{T}}\mathrm{E}\left(\mathbf{C}_{*}^{(\mathrm{w})}(\mathrm{X}\,|\,\mathrm{w})\right)\boldsymbol{\beta} \tag{4.102}$$

where:

$$\mathrm{E}\left(\mathbf{C}_{*}^{(\mathrm{w})}(\mathrm{X}\,|\,\mathrm{w})\right) = \binom{\mathrm{N}}{\mathrm{n}}^{-1} \sum_{\mathrm{w}\in\mathbf{W}} \mathbf{D}^{1/2}(X\,|\,\mathrm{w})\mathbf{R}^{(\mathrm{w})}(X\,|\,\mathrm{w})\mathbf{D}^{1/2}(X\,|\,\mathrm{w}).$$

If n and G are sufficiently large, $\frac{\mathrm{n}-1}{\mathrm{G}-1} \approx \frac{\mathrm{n}}{\mathrm{G}}$. Under this assumption, we have:

$$E\mathrm{E}(\tilde{\mathrm{Y}}_{\mathrm{gS}} - \overline{\mathrm{Y}})^2 - \mathrm{D}^2\left(\overline{\mathrm{Y}}_{\mathrm{A}}\right) \approx \frac{(\mathrm{m}-1)(\mathrm{G}-\mathrm{g})}{\mathrm{Ggm}}\,\boldsymbol{\beta}^{\mathrm{T}}\mathrm{E}\left(\mathbf{C}_{*}^{(\mathrm{w})}(\mathrm{X}\,|\,\mathrm{w})\right)\boldsymbol{\beta}. \tag{4.103}$$

Hence, if $\frac{\mathrm{n}-1}{\mathrm{G}-1} \approx \frac{\mathrm{n}}{\mathrm{G}}$ and the expected value of the matrix of within-cluster covariances is negative definite, then the predictor $\tilde{\mathrm{Y}}_{\mathrm{gS}}$ is more precise than the predictor $\overline{\mathrm{Y}}_{\mathrm{A}}$. Then, each sample should be divided into such clusters of the same size that all eigenvalues of the matrix $\mathbf{C}^{(\mathrm{w})}(X|\mathrm{w})$ are as short as possible.

V. TWO-STAGE SAMPLING

5.1. Basic properties

A fixed population of the size N is denoted by $\Omega=\{1,2,...,N\}$. Let us assume that the population Ω is divided into G such mutually disjoint clusters Ω_h (p=1,...,G) that $\bigcup_{h=1}^{G}\Omega_h=\Omega$. Let N_h be the size of the cluster Ω_h. The mean size of the clusters is denoted by: $\overline{N}=\frac{1}{G}\sum_{h=1}^{G}N_h$.

Let S be the two-stage sample. At the first stage, g clusters are drawn without replacement with a constant probability of inclusion of the first and second degrees. Next, the simple sample S_h of size n_h is drawn from a selected cluster Ω_h, where $h\in S$. The sampling design is as follows:

$$P_d(s)=\binom{G}{g}^{-1}\prod_{h=1}^{g}\binom{N_h}{n_h}^{-1} .$$

Let us consider the strategy $\left(\tilde{\mathbf{y}}_{gS},P_d\right)$, where $\tilde{\mathbf{y}}_{gS}=[\tilde{y}_{g1S}...\tilde{y}_{gmS}]$

$$\tilde{y}_{giS}=\frac{1}{g\overline{N}}\sum_{h\in S}N_h\bar{y}_{S_h} \tag{5.1}$$

where:

$$\bar{y}_{S_h}=\frac{1}{n_h}\sum_{k\in S_h}y_k .$$

It is well known that the strategy $\left(\tilde{\mathbf{y}}_{gS},P_d\right)$ is unbiased for the population vector of averages $\bar{\mathbf{y}}$. The variance-covariance matrix is as follows:

$$\mathbf{V}\left(\tilde{\mathbf{y}}_{gS},P_d\right)=\frac{G-g}{Gg}\mathbf{C}_*(z)+\frac{1}{gG}\sum_{h=1}^{G}\frac{N_h\left(N_h-n_h\right)}{n_h}\mathbf{C}_{*h} \tag{5.2}$$

where the elements of the matrix $\mathbf{C}_*(z)=[c_*(z_i,z_j)]$, are determined on the page 138 and the elements of the variance-covariance matrix $\mathbf{C}_{*h}=[c_{*h}(y_i,y_j)]$ are defined by the expression:

$$c_{*h}(y_i, y_j) = \frac{1}{N_h - 1} \sum_{k \in \Omega_h} (y_{ik} - \bar{y}_{ih})(y_{jk} - \bar{y}_{jh}), \quad \bar{y}_{ih} = \frac{1}{N_h} \sum_{k \in \Omega_h} y_{ik} \tag{5.3}$$

The estimator of the matrix $\mathbf{V}(\tilde{\mathbf{y}}_{gS}, P_d)$ is as follows:

$$\mathbf{V}_S(\tilde{y}_{gS}, P_d) = \frac{G-g}{Gg} \mathbf{C}_{*S} + \frac{1}{Gg} \sum_{h=1}^{G} \frac{N_h(N_h - n_h)}{n_h} \mathbf{C}_{S*h} \tag{5.4}$$

where the elements of the matrices $\mathbf{C}_{*S} = [c_{*S}(z_i, z_j)]$ and $\mathbf{C}_{*hS} = [c_{*hS}(y_i, y_j)]$ are defined by the expressions:

$$c_{*S}(z_i, z_j) = \frac{1}{g-1} \sum_{h=1}^{g} (N_n \bar{y}_{iS_h} - \bar{y}_{iS})(N_n \bar{y}_{jS_h} - \bar{y}_{jS}), \qquad \bar{y}_{iS} = \frac{1}{g} \sum_{h=1}^{g} N_h \bar{y}_{iS_h},$$

$$c_{*hS}(y_i, y_j) = \frac{1}{n_h - 1} \sum_{k \in S_h} (y_{ik} - \bar{y}_{iS_h})(y_{jk} - \bar{y}_{jS_h}).$$

Let the two-stage sample S be automatically balanced. So, it means that for each h=1,...,G:

$$n_h = fN_h \tag{5.5}$$

where 0<f<1

The estimator of the population mean vector is the vector $\tilde{\mathbf{y}}_{gS} = [\tilde{y}_{g1S} \ldots \tilde{y}_{gmS}]$ where:

$$\tilde{y}_{giS} = \frac{1}{g\bar{N}f} \sum_{h \in S} \sum_{k \in S_h} y_{ik} \tag{5.6}$$

Let the sampling design of the two-stage balanced sample be denoted by P'_d. The expressions (5.2) and (5.5) lead to the following one:

$$\mathbf{V}(\tilde{\mathbf{y}}_{gS}, P'_d) = \frac{G-g}{Gg} \mathbf{C}_*(z) + \frac{1-f}{gf} \bar{N} \mathbf{C}_{*w} \tag{5.7}$$

where:

$$\mathbf{C}_{*w} = \sum_{h=1}^{G} w_h \mathbf{C}_{*h}, \qquad w_h = \frac{N_h}{N}. \tag{5.8}$$

The matrix $\mathbf{C}_{*w}$ will be named the intra-cluster variance-covariance matrix.

Let us introduce the following notation $\mathbf{y}=[\mathbf{y}_{1\#\#}...\mathbf{y}_{m\#\#}]$ is the matrix of dimensions N×m of variable observations, where $\mathbf{y}_{i\#\#}=\left[\mathbf{y}_{i\#1}^{T}...\mathbf{y}_{i\#G}^{T}\right]$ is the column vector of dimensions N×1 of observations of an i-th variable in the population, where the vector $\mathbf{y}_{i\#h}=\left[y_{i1h}...y_{iN_hh}\right]^{T}$ is the sub-vector of dimensions N_h×1 of observations of an i-th variable in an h-th cluster. Let $\mathbf{J}_a$ be a unit vector of dimensions a×1. The sum of the observations of an i-th variable in an h-th cluster will be denoted by $z_{hi}=\mathbf{y}_{i\#h}^{T}\mathbf{J}_{N_h}$. Let $\mathbf{z}=[\mathbf{z}_{\#1}...\mathbf{z}_{\#m}]$ be the matrix of dimensions G×m, where: $\mathbf{z}_{\#i}^{T}=\left[z_{1i}...z_{Gi}\right]$. Finally, let $\mathbf{D}_J$ be the following matrix of dimensions G×N:

$$\mathbf{D}_J=\begin{bmatrix}\mathbf{J}_{N_1}^{T} & 0 & 0 & ... & 0\\ 0 & \mathbf{J}_{N_2}^{T} & 0 & ... & 0\\ ... & ... & ... & ... & ...\\ 0 & 0 & 0 & ... & \mathbf{J}_{N_G}^{T}\end{bmatrix}. \tag{5.9}$$

Hence:

$$\mathbf{z}=\mathbf{D}_J\mathbf{y}. \tag{5.10}$$

The vector $\bar{\mathbf{z}}=[\bar{z}_1...\bar{z}_G]$ is defined as the following function of the matrix $\mathbf{z}$:

$$\bar{\mathbf{z}}=\frac{1}{G}\mathbf{J}_G^{T}\mathbf{z}. \tag{5.11}$$

This leads to the following expression

$$\mathbf{C}_*(z)=\frac{1}{G-1}\left(\mathbf{z}-\mathbf{J}_G\bar{z}\right)^{T}\left(\mathbf{z}-\mathbf{J}_G\bar{z}\right). \tag{5.12}$$

On the basis of (5.10) and (5.11), we have:

$$\mathbf{z}-\mathbf{J}_G\bar{\mathbf{z}}=\mathbf{A}_G\mathbf{D}_J\mathbf{y} \tag{5.13}$$

where:

$$\mathbf{A}_G=\mathbf{I}_G-\frac{1}{G}\mathbf{J}_G\mathbf{J}_G^{T},\qquad \mathbf{A}_G^{2}=\mathbf{A}_G. \tag{5.14}$$

The matrix $\mathbf{A}_G$ is idempotent. This and the expressions (5.12), (5.13) lead to the following one:

$$\mathbf{C}_*(z)=\mathbf{y}^{T}\mathbf{A}\mathbf{y} \tag{5.15}$$

where:

$$\mathbf{A}=\frac{1}{G-1}\mathbf{D}_J^{T}\mathbf{A}_G\mathbf{D}_J. \tag{5.16}$$

The covariances defined by (5.3) can be expressed in the following way:

$$c_{*h}(y_i y_j) = \frac{1}{N_h - 1}\left(\mathbf{y}_{i\#h} - \mathbf{J}_{N_h}\bar{y}_{ih}\right)^T\left(\mathbf{y}_{j\#h} - \mathbf{J}_{N_h}\bar{y}_{jh}\right)$$

where:

$$\bar{y}_{ih} = \frac{1}{N_h}\mathbf{J}_{N_h}^T\mathbf{y}_{i\#h}.$$

This leads to the following expression:

$$\mathbf{y}_{i\#h} - \mathbf{J}_{N_h}\bar{y}_{ih} = \mathbf{B}_h\mathbf{y}_{i\#h}$$

where:

$$\mathbf{B}_h = \mathbf{I}_{N_h} - \frac{1}{N_h}\mathbf{J}_{N_h}\mathbf{J}_{N_h}^T, \qquad \mathbf{B}_h^2 = \mathbf{B}_h. \tag{5.17}$$

This derivation can simplify the expression (5.3) in the following way:

$$c_{*h}(y_j y_j) = \frac{1}{N_h - 1}\mathbf{y}_{i\#h}^T\mathbf{B}_h\mathbf{y}_{j\#h} \tag{5.18}$$

This lets us rewrite an element of the variance-covariance matrix shown by the expression (5.8) in the following way:

$$c_*(y_i y_j) = \frac{1}{N}\sum_{h=1}^{G}\frac{N_h}{N_h - 1}\mathbf{y}_{i\#h}^T\mathbf{B}_h\mathbf{y}_{j\#h}. \tag{5.19}$$

Let us define the following matrix:

$$\mathbf{B} = \frac{1}{N}\begin{bmatrix} q_1\mathbf{B}_1 & 0 & 0 & \ldots & 0 \\ 0 & q_2\mathbf{B}_2 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & q_G\mathbf{B}_G \end{bmatrix} \tag{5.20}$$

where:

$$q_h = \frac{N_h}{N_h - 1}.$$

This lets us rewrite the expression (5.19) as follows:

$$c_*(y_i y_j) = \mathbf{y}_{i\#\#}^T\mathbf{B}\mathbf{y}_{j\#\#}.$$

Hence, the matrix $\mathbf{C}_{*w}$ defined by (5.8) can be determined by the following expression:

$$\mathbf{C}_{*w} = \mathbf{y}^T\mathbf{B}\mathbf{y} . \tag{5.21}$$

This and (5.15) and (5.7) let us derive the following expression

$$\mathbf{V}\left(\tilde{\mathbf{y}}_{wS}, P_d^{'}\right) = a\mathbf{y}^T\mathbf{A}\mathbf{y} + b\mathbf{y}^T\mathbf{B}\mathbf{y} \tag{5.22}$$

where:

$$a = \frac{G-g}{Gg}, \qquad b = \frac{1-f}{gf}\overline{N} . \tag{5.23}$$

Let the matrix $\mathbf{C}_*(z)$ be positive definite. Let $\mathbf{H}$ be such an orthogonal matrix of degree m that $\mathbf{H}^T\mathbf{H} = \mathbf{I}_m$ and

$$\mathbf{H}^T\mathbf{C}_*(z)\,\mathbf{H}=\mathbf{D}_{m.} \tag{5.24}$$

where: $\mathbf{D}_m$ is the diagonal matrix of degree m and consists of the eigenvalues $d_{mi}>0$ (i=1,...,m) of the matrix $\mathbf{C}_*(z)$. Let $\mathbf{F}$ be such an orthogonal matrix of degree m that $\mathbf{F}^T\mathbf{F} = \mathbf{I}_m$ and:

$$\mathbf{F}^T\mathbf{D}_m^{1/2}\,\mathbf{H}\mathbf{C}_{*w}\mathbf{H}\mathbf{D}_m^{1/2}\mathbf{F} = \mathbf{D}_w \tag{5.25}$$

where: $\mathbf{D}_w=[d_i]$ is the diagonal matrix of degree m consisting of eigenvalues of the matrix $\mathbf{D}_m^{1/2}\mathbf{H}\mathbf{C}_{*w}\mathbf{H}\mathbf{D}_m^{1/2}$. Let

$$\mathbf{G} = \mathbf{H}\mathbf{D}_m^{1/2}\mathbf{F} . \tag{5.26}$$

This and the equations (5.24), (5.25) lead to conclusion that [see e.g. Rao (1982)]:

$$\mathbf{G}^T\mathbf{C}_*(z)\,\mathbf{G} = \mathbf{I}_m , \qquad \mathbf{G}^T\mathbf{C}_{*w}\mathbf{G} = \mathbf{D}_w . \tag{5.27}$$

This leads to the following expression:

$$\mathbf{C}_*(z) = \left(\mathbf{G}^T\right)^{-1}\mathbf{G}^{-1}, \qquad \mathbf{C}_{*w} = \left(\mathbf{G}^T\right)^{-1}\mathbf{D}_w\mathbf{G}^{-1}. \tag{5.28}$$

This result lets us rewrite the expression (5.7) in the following way:

$$\mathbf{V}\left(\tilde{\mathbf{y}}_{wS}, P_d^{'}\right) = \left(\mathbf{G}^T\right)^{-1}\left(a\mathbf{I}_m + b\mathbf{D}_w\right)\mathbf{G}^{-1} \tag{5.29}$$

where a and b are defined by the expression (5.23).

Let us transform the data matrix **y** into the matrix **u** of dimensions N×m in the following way:

$$\mathbf{u} = \mathbf{yG}. \tag{5.30}$$

The columns of the matrix **u** can be treated as outcomes of new variables. This and the expressions (5.15), (5.21), (5.27), (5.30) lead to the equations: $\mathbf{I}_m = \mathbf{u}^T\mathbf{Au}$, $\mathbf{D}_w = \mathbf{u}^T\mathbf{Bu}$. This lets us rewrite the expression (5.29) in the following way:

$$\mathbf{V}\left(\tilde{\mathbf{y}}_{wS}, P_d^{'}\right) = \left(\mathbf{G}^T\right)^{-1}\left(a\mathbf{u}^T\mathbf{Au} + b\mathbf{u}^T\mathbf{Bu}\right)\mathbf{G}^{-1}. \tag{5.31}$$

The diagonal matrix $\mathbf{D}_w = \mathbf{u}^T\mathbf{Bu}$ is the intra-cluster variance-covariance matrix of the variables. Then the diagonal element d_i of the matrix $\mathbf{D}_w$ is the intra-cluster variance of the i-th variable whose outcomes are elements of the i-the column of the matrix **u**.

When $\mathbf{C}_*(z)$ is non-singular, it is possible to make such a decomposition of the matrix $\mathbf{V}\left(\tilde{\mathbf{y}}_{wS}, P_d^{'}\right)$ that the matrix $\mathbf{C}_*(z)$ is transformed into a diagonal matrix and the matrix $\mathbf{C}_w$ into the unit one.

The well known properties of the matrix determinant and the expressions (5.23), (5.28), (5.29) lead to the following expressions:

$$\det \mathbf{V}\left(\tilde{\mathbf{y}}_{wS}, P_d^{'}\right) = \det \mathbf{C}_*(z) \prod_{i=1}^{m}\left(a + bd_i\right),$$

$$\det \mathbf{V}\left(\tilde{\mathbf{y}}_{wS}, P_d^{'}\right) = \det \mathbf{C}_*(z) \prod_{i=1}^{m}\left(\frac{1-\overline{N}d_i}{g} + \frac{\overline{N}d_i}{gf} - \frac{1}{G}\right). \tag{5.32}$$

Let us assume that elements of the diagonal matrix $\mathbf{D}_w$ fulfill the inequalities: $0 \le d_i \le \frac{1}{\overline{N}}$. Let $\mathbf{v}_*(z_i)$ and $v_{*w}(y_i)$ be the diagonal elements of $\mathbf{C}_*(z)$ and C_{*w}, respectively. Then, the expression (5.7) leads to the following one:

$$q^2\left(\tilde{\mathbf{y}}_{wS}, P_d^{'}\right) = \mathrm{tr}\mathbf{V}\left(\tilde{\mathbf{y}}_{wS}, P_d^{'}\right) = \frac{G-g}{Gg}\sum_{h=1}^{m} v_*\left(z_i\right) + \frac{1-f}{gf}\overline{N}\sum_{h=1}^{m} v_{*w}\left(y_i\right). \tag{5.33}$$

Let us introduce the following notation:

$$d_{\#i} = \frac{v_{*w}(y_i)}{v_*(z_i)}. \tag{5.34}$$

If $\mathbf{C}_*(z)$ is positive definite, then

$$q^2\left(\tilde{\bar{y}}_{wS}, P_d'\right) = \frac{1}{g}\operatorname{tr}\mathbf{C}_*(z)\left(1+\bar{N}\frac{1-f}{f}\bar{d}_{\#}\right) - \frac{1}{G}\operatorname{tr}\mathbf{C}_*(z) \tag{5.35}$$

where:

$$\bar{d}_{\#} = \sum_{i=1}^{m} w_i \bar{d}_{\#i}, \qquad w_i = \frac{v_*(z_i)}{\operatorname{tr}\mathbf{C}_*(z)}. \tag{5.36}$$

Hence, we can say that $\bar{d}_{\#}$ determines the mean degree of intra-cluster spread of an m-dimensional variable. This spread is larger and larger when the value of the coefficient $\bar{d}_{\#}$ increases. Moreover, the mean radius of the strategy $\left(\tilde{\bar{y}}_{wS}, P_d'\right)$ is the increasing function of the coefficient $\bar{d}_{\#}$.

5.2. Minimization of the expected costs under fixed accuracy of estimation

Let k_1 be the unit cost of preparing the first stage sampling design. The unit cost of observations of variables in the second stage sample is denoted by k_2. In the case of the automatically balanced two stage design, the expected total cost function is as follows (see e.g. Konijn (1973), p. 322):

$$k(g,f) = k_1 g + k_2 \bar{N} g f. \tag{5.37}$$

5.2.1. Fixed level of risk function

Let the risk function be the following linear combination of the variances of the elements of the vector estimator $\tilde{\bar{y}}_{igS}$:

$$h(g,f) = \sum_{i=1}^{m} a_i D^2\left(\tilde{\bar{y}}_{igS}\right) = \frac{G-g}{Gg}\sum_{i=1}^{m} a_i v_*(z_i) + \frac{1-f}{gf}\bar{N}\sum_{i=1}^{m} a_i v_{*w}(y_i) \tag{5.38}$$

or if $v_{*m}(y_i) > 0$ for each i=1,...,m:

$$h(g,f) = q_m\left\{\frac{1}{g}\left(1+\frac{1-f}{f}\overline{N}\tilde{d}\right)-\frac{1}{G}\right\} \tag{5.39}$$

where:

$$q_m = \sum_{i=1}^{m} a_i v_*(z_i), \tag{5.40}$$

$$\tilde{d} = \sum_{i=1}^{m} d_{\#i} w_i', \qquad w_i' = \frac{a_i v_*(z_i)}{q_m}. \tag{5.41}$$

The coefficient $d_{\#i}$ is explained by the expression (5.34).

The problem is such a determination of g (the size of the first stage simple sample) and of the fraction f that the expected cost function takes the minimal value under the fixed risk function. So:

$$\begin{cases} k(g,f) = \text{minimum} \\ h(g,f) \le h_d \\ 0 < f \le 1, \qquad 1 < g \le G \end{cases} \tag{5.42}$$

It is easy to prove the following lemmas.

Lemma 5.1 [Wywiał (1992)]: The cost function k(g,f) explained by the expression (5.37) has the positive derivative in the direction of each vector attached to the point (g=1,f=0) and having the end point (g^o,f^o), where $g^o > 1$ and $f^o > 0$.

Lemma 5.2 [Wywiał (1992)]: If $\tilde{d} < \overline{N}^{-1}$, the function h(g,f) is strictly convex for f >0 and g>0.

Proof: Let Q(t,z) be the quadratic form of the Hessian of the function h(g,f). After some transformation we have:

$$Q(t,z) = \frac{1}{g}q_m\left\{\frac{2(1-\overline{N}\tilde{d}}{g^2}t^2+\frac{\overline{N}\tilde{d}}{f}\left[\left(\frac{t}{g}\right)^2+\left(\frac{z}{f}\right)^2\right]+\frac{\overline{N}\tilde{d}}{f}\left(\frac{t}{g}+\frac{z}{f}\right)^2\right\} > 0$$

for t≠0 and z≠0

Hence, if $\tilde{d} < \overline{N}^{-1}$ and f>0 and g>0, the function h(g,f) is strictly convex.

Let us introduce the following notation:

$$g_* = \frac{q_m}{h_d}\left\{1-\overline{N}\tilde{d}+\overline{N}\sqrt{\frac{k_2(1-\tilde{d})\tilde{d}}{k_1}}\right\}, \tag{5.43}$$

$$g_{**} = q_m h_d^{-1}, \tag{5.44}$$

$$f_* = \sqrt{\frac{k_1}{k_2}\frac{\tilde{d}}{1-\overline{N}\tilde{d}}}, \tag{5.45}$$

$$f_{**} = \frac{\overline{N}\tilde{d}}{Gg_{**}^{-1} + \overline{N}\tilde{d} - 1}, \tag{5.46}$$

$$f_{***} = \frac{\overline{N}\tilde{d}}{g_{**}^{-1} + \overline{N}\tilde{d} - 1}. \tag{5.47}$$

The optimal solution to the problem is denoted by ($\underline{g}$,$\underline{f}$).

When $\tilde{d} < \frac{1}{N}$ and

a) if $g_* \le G$ and $f_* \le 1$, then ($\underline{g}$,$\underline{f}$) = (g_*,f_*),
b) if $f_*>1$ and $g<G$, then ($\underline{g}$,$\underline{f}$) = (g_{**},1),
c) if $g_*>G$ and $f_*<1$, then ($\underline{g}$,$\underline{f}$) = (G,f_{**}),
d) if $g_* <1$ and $f_* <1$, then ($\underline{g}$,$\underline{f}$) = (1,f_{***}).

5.2.2. Fixed level of the generalized variance

Let us assume that the sample size G is so large that the quantity G^{-1} can be neglected in the expression (5.32). In this case the generalized variance of the vector $\bar{\mathbf{y}}_{gS}$ is as follows:

$$u(g,f) = \frac{1}{g^m}\det \mathbf{C}_*(z)\prod_{i=1}^{m}\left(1-\overline{N}d_i + \frac{1}{f}\overline{N}d_i\right) \tag{5.48}$$

where d_i is the diagonal element of the matrix $\mathbf{D}_w$ defined by the expressions (5.25)-(5.27).

The problem is such a determination of $\underline{g}$ and $\underline{f}$ that the expected total cost takes the minimal value under the fixed level of the generalized variance. The m-th root of the generalized variance is as follows:

$$u_{\#}(g,f) = \sqrt[m]{\det \mathbf{V}(\bar{\mathbf{y}}_{gS})} = \frac{1}{g}c_o\sqrt[m]{\prod_{i=1}^{m}\left(b_i + \frac{1}{f}\right)} \tag{5.49}$$

where:

$$c_o = \sqrt[m]{\det \mathbf{C}_*(z)}\,, \tag{5.50}$$

$$b_i = \frac{1}{\overline{N}d_i} - 1\,. \tag{5.51}$$

The solution to the problem does not change when the m-th root of the generalized variance is substituted for the generalized variance. Hence, the problem can be specified as follows:

$$\begin{cases} k(g,f) = \text{minimum} \\ u_{\#}(g,f) \le u_d \\ g \ge 1, \qquad 0 < f \le 1 \end{cases} \tag{5.52}$$

Lemma 5.3 [Wywiał (1992)]: The function $u_{\#}$ defined by the expression (5.49) is strictly convex for $g\neq 0$ and $f\neq 0$.

Proof: The quadratic form of the Hessian of the function $u_{\#}(g,f)$ is as follows:

$$Q_{\#}(t,v) = u_{\#}(g,f)\left\{\frac{1}{m}\sum_{i=1}^{m}\left(\frac{t}{g}+\frac{v}{z_i}\right)+\left(\frac{t}{g}\right)^2+\frac{2v^2}{m}\sum_{i=1}^{m}\frac{b_i}{z_i^2}+\left(\frac{\partial u_{\#}}{\partial z}\right)^2 v^2\right\} > 0$$

for $t\neq 0$ and $v\neq 0$ where $z_i=b_if^2+f>0$ because $f>0$ and $b_i\ge 0$. Hence, the function $u_{\#}(g,f)$ is strictly convex for $g\neq 0$ and $f\neq 0$.

Let f_* be the root of the following equation:

$$\frac{1}{k_0+f} - \frac{1}{mf}\sum_{i=1}^{m}\frac{1}{b_if+1} = 0 \tag{5.53}$$

where:

$$k_0 = \frac{k_1}{\overline{N}k_2}\,, \tag{5.54}$$

$$g_* = \frac{c_o}{u_d}\sqrt[m]{\bigcap_{i=1}^{m}\left(b_i - \frac{1}{f_*}\right)}\,. \tag{5.55}$$

Let f' be the root of the equation $u_{\#}(1,f)=u_d$ and let g' be the root of the equation $u_{\#}(g,1)=u_d$.

Under the asumption:

$$\bar{d} > \frac{k_2}{k_1 + \bar{N}k_2} \tag{5.56}$$

where:

$$\bar{d} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} d_i} \ . \tag{5.57}$$

the lemma 5.3 leads to the following optimal solution $(\underline{g},\underline{f})$ to the problem (5.52):
a) if $f_* \in (0,1>$ and $g_*>1$, $(\underline{g},\underline{f})=(g_*,f_*)$,
b) if $f_*>1$, $(\underline{g},\underline{f})=(g',1)$,
c) if $g_*<1$, $(\underline{g},\underline{f})=(1,f')$.

The root f_* of the equation (5.53) should be obtained by means of the appropriate method of solving a nonlinear equation.

5.3. Maximization of estimation accuracy under fixed expected total costs

5.3.1. Minimization of squared risk function

Our problem is determining such a size $\underline{g}$ of clusters and such a fraction $\underline{f}$ that the risk function defined by the expression (5.39) takes the minimal value under the fixed expected value of the total costs defined by the expression (5.37). Hence:

$$\begin{cases} h(g,f) = \text{minimum} \\ k(g,f) \le K \\ f_d \le f \le 1, \ \ 1 \le g \le G \end{cases} \tag{5.58}$$

The following transformation is considered:

$$x=gf, \qquad g=g. \tag{5.59}$$

This lets us form the following problem equivalent to the problem (5.58):

$$\begin{cases} h_{\$}(g,x) = \text{minimum} \\ k_{\$}(g,x) \le K \\ 0 < x \le g \le G, \ \ g \ge 1 \end{cases} \tag{5.60}$$

where:

$$h_{\$}(g,x)=\frac{1}{g}\left(1-\overline{N}\tilde{d}\right)+\frac{1}{x}\overline{N}\tilde{d}-\frac{1}{G}, \tag{5.61}$$

$$k_{\$}(g,x)=k_1 g+k_2\overline{N}x\,. \tag{5.62}$$

Lemma 5.4: If $\tilde{d}\overline{N}<1$, the function $h_{\$}$ is strictly convex for x>0 and g>0.

Let us introduce the following notation:

$$g'=x'=\frac{f_d K}{k_1+k_2\overline{N}f_d}, \tag{5.63}$$

$$g''=\frac{K-k_2\overline{N}f_d}{k_1}, \tag{5.64}$$

$$x''=\frac{K-k_1 G}{\overline{N}k_2}, \tag{5.65}$$

$$g_*=e\sqrt{\frac{1-\overline{N}\tilde{d}}{k_1}},\qquad x_*=e\sqrt{\frac{\tilde{d}}{k_2}} \tag{5.66}$$

where:

$$e=K\left(\sqrt{\left(1-\overline{N}\tilde{d}\right)k_1}+\overline{N}\sqrt{k_2\tilde{d}}\right)^{-1}.$$

On the basis of the lemma 5.4 and under the assumptions that: $\tilde{d}\overline{N}<1$ and $G>g'$, the following solution $(\underline{g},\underline{f})$ can be derived:

a) if $P_*\in\overline{AC}$, $\left(\underline{g},\underline{f}\right)=\left(g_*,\frac{x_*}{g_*}\right)$,

b) if $g_*<g'$, $(\underline{g},\underline{f})=(g',1)$,

c) if $G<g''$ and $g_*>G$, $\left(\underline{g},\underline{f}\right)=\left(G,\frac{x''}{G}\right)$,

where the coordinates of the ends of the segment $\overline{AC}$ are as follows A(g',g') and B(G,x").

5.3.2. Minimization of generalized variance

Let us assume that the number of cluster $G \to \infty$. The problem is to determine such a size of the selected clusters $\underline{g}$ and such a fraction $\underline{f}$ that the generalized variance of the strategy $(\bar{\mathbf{y}}_{gS}, P_d')$ takes the minimal value under the fixed expected total costs of observation of population elements selected to the sample. This problem is equivalent to the following problem of minimization of the root of the m-th degree of the generalized variance:

$$\begin{cases} u_{\#}(g,f) = \text{minimum} \\ k(g,f) \le K \\ g \ge 1, \quad f_d \le f \le 1 \end{cases} \tag{5.67}$$

where the functions $u_{\#}$, and k are defined by the expressions (5.49) and (5.37), respectively. Let us introduce the notation:

$$g_* = \frac{K}{k_1 + \bar{N} k_2 f_*} . \tag{5.68}$$

When $f_* > 1$, then $\underline{f}=1$ and the above equation leads to the following:

$$g' = \frac{K}{k_1 + \bar{N} k_2} . \tag{5.69}$$

When $k_1 + k_2 \bar{N} < K$ and the inequality (5.56) is fulfilled, the optimal solution $(\underline{g},\underline{f})$ to the problem (5.67) is as follows:
a) if $f_* \le 1$, $(\underline{g},\underline{f}) = (g_*, f_*)$,
b) if $f_* > 1$, $(\underline{g},\underline{f}) = (g', 1)$.

The root f_* of the equation (5.53) should be obtained by means of an appropriate numerical method.

VI. VECTOR OF REGRESSION ESTIMATORS

6.1. Basic properties

In the one-dimensional case, properties of regression estimator are considered e.g. by: Bracha (1978, 1982, 1983, 1987), Cochran (1963), Greń (1969, 1970), Konijn (1962, 1973), Murthy (1977), Sarndal, Swensson, Wretman (1992), Tripathi (1973) and Wywiał (1992,1995). We are going to generalize their results on a multidimensional case.

Let **A** be a real matrix of dimensions z×m, where: m is the number of estimated population averages: $\bar{\mathbf{y}} = [\bar{y}_1 ... \bar{y}_m]$. The number of elements of the vector of auxiliary means $\bar{\mathbf{x}} = [\bar{x}_1 ... \bar{x}_z]$ is denoted by z. Let S be the simple sample drawn without replacement. The vectors of the sample means of variables under study and auxiliary variables are denoted by $\bar{\mathbf{y}}_S$ and $\bar{\mathbf{x}}_S$, respectively. The vector of difference estimators is defined by the equation:

$$\mathbf{t}_{AS} = \bar{\mathbf{y}}_S + (\bar{\mathbf{x}}_S - \bar{\mathbf{x}})\mathbf{A}. \tag{6.1}$$

The strategy $(\mathbf{t}_{AS}, P_3)$ is the unbiased estimator of the mean vector $\bar{y}$ and its variance-covariance matrix is as follows:

$$\mathbf{V}(\mathbf{t}_{AS}, P_3) = \frac{N-n}{Nn}(\mathbf{C}_{*yy} + \mathbf{C}_{*yx}\mathbf{A} + \mathbf{A}^T\mathbf{C}_{*xy} + \mathbf{A}^T\mathbf{C}_{*xx}\mathbf{A}). \tag{6.2}$$

where $\mathbf{C}_{*yy}$, $\mathbf{C}_{***}$ and $\mathbf{C}_{*xy} = \mathbf{C}_{*yx}$ are the covariance matrices of the vector of variables.

Let us substitute the following matrix for the matrix **A** in the equations (6.1) and (6.2):

$$\mathbf{B} = -\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy,} \qquad \det(\mathbf{C}_{xx}) > 0. \tag{6.3}$$

This determines the vector of regression estimators:

$$\mathbf{t}_{BS} = \bar{\mathbf{y}}_S + (\bar{\mathbf{x}}_S - \bar{\mathbf{x}})\mathbf{B} \tag{6.4}$$

and the variance-covariance matrix:

$$\mathbf{V}(\mathbf{t}_{BS}, P_3) = \frac{N-n}{Nn}\left(\mathbf{C}_{*yy} - \mathbf{C}_{*yx}\mathbf{C}_{*xx}^{-1}\mathbf{C}_{*xy}\right). \tag{6.5}$$

The vector $(\mathbf{t}_{BS}, P_3)$ is the unbiased estimator of the mean vector $\bar{\mathbf{y}}$.

Let $\boldsymbol{A}$ be the set of vectors of difference estimators $\mathbf{t}_{AS}$ where $\boldsymbol{A} \in R^{m+z}$. Hence:

Theorem 6.1 [Wywiał (1988)]: If S is a simple sample drawn without replacement, $\mathbf{t}_{BS}$ is the effective estimator of the mean vector $\overline{\mathbf{y}}$ in the class $\boldsymbol{A}$.

Proof: The matrix

$$\mathbf{V}(\mathbf{t}_{AS},P_3)\text{-}\mathbf{V}(\mathbf{t}_{BS},P_3)=\frac{N-n}{Nn}(\mathbf{A}-\mathbf{B})^{T}\mathbf{C}_{*xx}(\mathbf{A}-\mathbf{B})$$

is the Gramma matrix. Hence, it is non-negative definite.

The relative efficiency index of the estimator $\mathbf{t}_{BS}$ to the simple sample $\overline{\mathbf{y}}_S$ is as follows:

$$e_w(\mathbf{t}_{BS}/\overline{\mathbf{y}}_S)=\det\mathbf{V}(\mathbf{t}_{BS},P_3)\det\mathbf{V}^{-1}(\overline{\mathbf{y}}_S,P_3)=\frac{\det(\mathbf{C}_{yy}-\mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy})}{\det(\mathbf{C}_{yy})}. \tag{6.6}$$

Let r_{qi}, i=1,...,p=min(m,z), be the coefficients of the canonical correlation. They measure the linear dependence between the vector of the variables under study $y=[y_1...y_m]$ and the vector of auxiliary variables $x=[x_1...x_z]$. The coefficient $r_{qi}^2 \le 1$ (i=1,...,p) is obtained as the eigenvalue of the matrix $\mathbf{C}_{yy}^{-1}\mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}$ [see e.g. Anderson (1958) or C.R.Rao (1982)]. This leads to the following equations:

$$\det\mathbf{V}(\mathbf{t}_{BS},P_3) = \det(\mathbf{C}_{*yy})\prod_{i=1}^{p}(1-r_{qi}^2),$$

$$e_w(\mathbf{t}_{BS}/\overline{\mathbf{y}}_S)=\prod_{i=1}^{p}(1-r_{qi}^2). \tag{6.7}$$

The precision of the vector of regression estimators $\mathbf{t}_{BS}$ is high in relation to the simple sample mean $\overline{\mathbf{y}}_S$ when the variables under study are highly correlated with the auxiliary variables in the sense of canonical correlation.

The theorems 1.7 and 6.1 lead to the inequalities: $q(\mathbf{t}_{BS},P_3)\le q(\overline{\mathbf{y}}_S,P_3)$ and $\rho(\mathbf{t}_{BS},P_3)\le\boldsymbol{\rho}(\overline{\mathbf{y}}_S,P_3)$.

Let $v_{*i}\,r_i^2$ (i=1,..,m) be the i-th diagonal element of the matrix $\mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}$ where $0\le r_i\le 1$ is the well known multiple correlation coefficient between the i-th variable y_i under study and auxiliary variables $x=[x_1...x_z]$. The parameter r_i^2 is named a determination coefficient. This and the expression (6.5) lead to the following results:

$$q(\mathbf{t}_{BS}) = q(\bar{\mathbf{y}}_S)\sqrt{1-\hat{r}^2} \tag{6.8}$$

where:

$$\hat{r}^2 = \sum_{i=1}^{m} g_i r_i^2 , \tag{6.9}$$

$$g_i = v_i \left(\sum_{j=1}^{m} v_j \right)^{-1} , \; i=1,...,m.$$

The mean determination coefficient will be denoted by $\hat{r}^2$ and $0 \le \hat{r}^2 \le 1$. The inequality $0 \le q(\mathbf{t}_{BS},P_3) \le q(\bar{\mathbf{y}}_S,P_3)$ and the expression (6.9) lead to the following conclusion. The mean radius of the estimator $\mathbf{t}_{BS}$ decreases when the mean determination coefficient increases.

On the basis of the expressions (1.57) and (6.5) we evaluate the deff-coefficient:

$$\operatorname{deff}\left(\mathbf{t}_{BS}, P_3\right) = \rho \mathbf{V}\left(\left(\mathbf{t}_{BS}, P_3\right) \mathbf{V}^{-1} \left(\bar{\mathbf{y}}_S, P_1\right) \right)$$

$$\operatorname{deff}\left(\mathbf{t}_{BS}, P_3\right) = \frac{N-n}{N-1} \rho \left(I - \mathbf{C}_{*yx} \mathbf{C}_{*xx}^{-1} \mathbf{C}_{*xy} \mathbf{C}_{*yy}^{-1} \right)$$

or

$$\operatorname{deff}\left(\mathbf{t}_{BS}, P_3\right) = \frac{N-n}{N-1} \left(1 - r_{qp}^2\right) \tag{6.10}$$

where r_{qp}^2 is the lowest coefficient of canonical correlation between the auxiliary variables and variables under study. More precisely, the coefficient r_{qp}^2 is equal to the lowest eigenvalue of the matrix $\mathbf{Q} = \mathbf{C}_{*yx} \mathbf{C}_{*xx}^{-1} \mathbf{C}_{*xy} \mathbf{C}_{*yy}^{-1}$. The expression (6.10) is derived on the basis of the following expressions:

$$\left| \mathbf{I} - \mathbf{Q} - \lambda \mathbf{I} \right| = 0 ,$$

which is equivalent to the following:

$$\left| \mathbf{Q} - \gamma \mathbf{I} \right| = 0$$

where $\gamma = 1-\lambda$ and eigenvalues $1 \ge r_{q1}^2 \ge r_{q2}^2 \ge ... \ge r_{qp}^2 \ge 0$ of the matrix $\mathbf{Q}$ are the coefficients of canonical correlation.

Let us assume that the vector of auxiliary variables $\bar{\mathbf{x}}$ is known. The matrix of the regression parameters **B** can be estimated on the basis of the simple sample drawn without replacement by means of the following matrix:

$$\mathbf{B}_S = -\mathbf{C}_{*xxS}^{-1}\mathbf{C}_{*xyS} \tag{6.11}$$

where: $\mathbf{C}_{*xyS}=[c_{*S}(x_i,y_k)]$ and $\mathbf{C}_{*xxS}=[c_{*S}(x_i,x_j)]$, k=1,...,m; i,j=1,...,z:

$$c_{*S}(x_i,y_u)=\frac{1}{n-1}\sum_{k\in S}(x_{ik}-\bar{x}_{iS})(y_{uk}-\bar{y}_{uS}),\ \bar{x}_{iS}=\frac{1}{n}\sum_{k\in S}x_{ik},$$

$$c_{*S}(x_i,x_j)=\frac{1}{n-1}\sum_{k\in S}(x_{ik}-\bar{x}_{iS})(x_{jk}-\bar{x}_{jS}),\qquad \bar{y}_{uS}=\sum_{k\in S}y_{uk}\ .$$

Substituting the matrix $\mathbf{B}_S$ for **B** in the expression (6.4), we have:

$$\mathbf{t}_{BS}=\bar{\mathbf{y}}_S+(\bar{\mathbf{x}}_S-\bar{\mathbf{x}})\mathbf{B}_S\ . \tag{6.12}$$

The statistic $\mathbf{t}_{BS}$ is an asymptotically unbiased estimator of the vector of population means $\bar{\mathbf{y}}$ and its variance covariance matrix is approximately determined by the expression (6.5). The estimator of this matrix is as follows:

$$\mathbf{V}(\mathbf{t}_{BS},P_3)=\frac{N-n}{Nn}\left(\mathbf{C}_{*yyS}-\mathbf{C}_{*yxS}\mathbf{C}_{*xxS}^{-1}\mathbf{C}_{*xyS}\right). \tag{6.13}$$

Let us note that the problems of estimation on the basis of a vector of ratio or product estimators are considered e.g. by: John (1969), Lynch (1978), Olkin (1958), Tripathi (1976) and Wywiał (1992).

6.2. Vector of regression estimators from double sample

Usually the matrix **B** as well as the vector $\bar{\mathbf{x}}$ are not known. In this situation, the two-phase simple sample $S=\{S_1, S_2\}$ can be selected from a population in order to determine those parameters. The simple sample S_1 is drawn without replacement from a population. In this sample, values of auxiliary variables are observed. The simple sample S_2 is selected from the outcome of the sample S_1. Additionally, in this sample the values of variables under study y are observed. The samples S_1 and S_2 are of sizes n_1 and n_2, respectively and $n_1 > n_2$. The sampling design of the double sample is as follows:

$$P_{3d}(s)=\binom{N}{n_1}^{-1}\binom{n_1}{n_2}^{-1}.$$

The matrices $\mathbf{C}_{*yx}$, $\mathbf{C}_{*xx}$ are estimated by the following unbiased estimators $\overline{\mathbf{C}}_{*yx}=\left[\overline{c}_*(y_i,x_j)\right]$, $\overline{\mathbf{C}}_{*xx}=\left[\overline{c}_*(x_i,x_j)\right]$, respectively, where:

$$\overline{c}_*(x_i,x_u)=\frac{1}{n_2-1}\sum_{k\in S_2}(x_{ik}-\overline{x}_{iS_2})(x_{uk}-\overline{x}_{uS_2}),\qquad \overline{x}_{iS_2}=\frac{1}{n_2}\sum_{k\in S_2}x_{ik},$$

$$\overline{c}_*(y_i,x_u)=\frac{1}{n_2-1}\sum_{k\in S_2}(y_{ik}-\overline{y}_{iS_2})(x_{uk}-\overline{x}_{uS_2}),\qquad \overline{y}_{iS_2}=\frac{1}{n_2}\sum_{k\in S_2}y_{ik}.$$

Under the assumption that $\det(\overline{\mathbf{C}}_{*xx})>0$, the estimator of the matrix **B** is as follows:

$$\overline{\mathbf{B}}=-\overline{\mathbf{C}}_{*xx}^{-1}\overline{\mathbf{C}}_{*xy} \tag{6.14}$$

Let $\overline{\mathbf{x}}_{S_2}=[\overline{x}_{1S_2}...\overline{x}_{zS_2}]$, $\overline{\mathbf{x}}_{S_1}=[\overline{x}_{1S_1}...\overline{x}_{zS_1}]$, $\overline{\mathbf{y}}_{S_2}=[\overline{y}_{1S_2}...\overline{y}_{zS_2}]$ where:

$$\overline{x}_{iS_2}=\frac{1}{n_2}\sum_{k\in S_2}x_{ik},\qquad \overline{x}_{iS_1}=\frac{1}{n_1}\sum_{k\in S_1}x_{ik},\qquad \overline{y}_{iS_2}=\frac{1}{n_1}\sum_{k\in S_2}y_{ik}.$$

Substituting the matrix $\overline{\mathbf{B}}$ and the vectors $\overline{\mathbf{x}}_{S_1}$, $\overline{\mathbf{y}}_{S_2}$, $\overline{\mathbf{x}}_{S_2}$ for **B** and $\overline{\mathbf{x}}$, $\overline{\mathbf{y}}$, $\overline{\mathbf{x}}_S$, respectively, in the expression (6.4) we have the following estimator of the vector $\overline{\mathbf{y}}$:

$$\overline{\mathbf{t}}_{BS}=\overline{\mathbf{y}}_{S_2}+(\overline{\mathbf{x}}_{S_2}-\overline{\mathbf{x}}_{S_1})\overline{\mathbf{B}}. \tag{6.15}$$

The parameters of the vector $\overline{\mathbf{t}}_{BS}$ are as follows (see Wywiał (1988 and 1992)):

$$E(\overline{\mathbf{t}}_{BS},P_{3d})=\overline{y}+0(n_1^{-1})+0(n_2^{-1}), \tag{6.16}$$

$$\mathbf{V}(\overline{\mathbf{t}}_{BS},P_{3d})=\frac{N-n_1}{Nn_1}\mathbf{C}_{*yy}+\frac{n_1-n_2}{n_1n_2}(\mathbf{C}_{*yy}-\mathbf{C}_{*yx}\mathbf{C}_{*xx}^{-1}\mathbf{C}_{*xy})+$$

$$+0(n_1^{-1})+0(n_2^{-2})+0(n_1^{-1}n_2^{-1}). \tag{6.17}$$

This and the definition of the mean radius of a vector estimator lead to following expression:

$$q^2\left(\bar{\mathbf{t}}_{BS}, P_{3d}\right) = \mathrm{tr}\mathbf{V}\left(\bar{\mathbf{t}}_{BS}, P_{3d}\right) = \mathrm{tr}\left(\mathbf{C}_{*yy}\right)\left(\frac{\hat{r}^2}{n_1} + \frac{1-\hat{r}^2}{n_2} - \frac{1}{N}\right) \tag{6.18}$$

where $\bar{r}$ is defined by the expression (6.9).

6.3. Optimization of sample sizes

Let k_1 be the per observation cost of auxiliary variables and let k_2 be the per observation cost of variables under study. The admissible total cost will be denoted by K. Usually, the following linear cost function is considered:

$$k(n_1,n_2) = k_1n_1 + k_2n_2. \tag{6.19}$$

This function is involved in all optimization problems formed below.

6.3.1. Minimization of square risk function

The variance of the i-th element of the vector $\bar{\mathbf{t}}_{BS}$ can be obtained on the basis of the expression (6.17) and it is as follows:

$$D^2\left(\bar{\mathbf{t}}_{BiS}, P_{3d}\right) = v_*\left(y_i\right)\left(\frac{r_i^2}{n_1} + \frac{1-r_i^2}{n_2} - \frac{1}{N}\right),\ i=1,...,m. \tag{6.20}$$

Let us consider the following risk function:

$$u_1(n_1,n_2) = \sum_{i=1}^{m} a_i D^2(\bar{\mathbf{t}}_{BiS}, P_5) \tag{6.21}$$

where $\mathbf{a} = [a_1...a_m] \in R^m - \{\mathbf{o}_m\}$. This and the expression (6.20) let us rewrite the risk function in the following way:

$$u_1(n_1,n_2) = q^2\left(\frac{\bar{r}^2}{n_1} + \frac{1-\bar{r}^2}{n_2} - \frac{1}{N}\right) \tag{6.22}$$

where:

$$q^2 = \sum_{i=1}^{m} a_i v_*(y_i), \tag{6.23}$$

$$\bar{r}^2 = \sum_{i=1}^{m} r_i^2 w_i, \qquad w_i = q^{-2} a_i v_*(y_i). \tag{6.24}$$

The parameter $\bar{r}^2$ is the weighed mean of the squared multiple correlation coefficients r_i (i=1,...,m). The coefficient r_i measures a linear dependence between the i-th variable under study y_i and auxiliary variables x. The determination coefficients r_i^2 is equal to the i-th diagonal element of the matrix $\mathbf{C}_{*yx}\mathbf{C}_{*xx}^{-1}\mathbf{C}_{*xy}$. The inequalities $0 \le \bar{r} \le 1$ result from the fact that $0 \le r_i \le 1$ and $0 \le w_i \le 1$ for each i=1,...,m. Let us note that the parameter $\bar{r}^2$ can be treated as the weighed mean of the determination coefficients r_i^2 (i=1,...,m).

Let us determine such optimal sizes of sample S_1 and S_2 that the risk function takes a minimal value under a fixed total cost of variable observation:

$$\begin{cases} u_1(n_1, n_2) = \min imum \\ k(n_1, n_2) \le K, \quad 2 \le n_2 < n_1 \le N \end{cases} \tag{6.25}$$

Wywiał (1992) showed that if $3k_1+2k_2 \le K$ and $\bar{r} > r_* = \dfrac{k_1}{k_1 + k_2}$, the optimal solution is as follows:

$$\begin{cases} \underline{n}_1 = \min imum\left\{N, n_1', n_1^*\right\} \\ \underline{n}_2 = \max imum\left\{n_2', 2, n_2^*\right\} \end{cases} \tag{6.26}$$

where:

$$n_1' = \frac{K - 2k_1}{k_1}, \qquad n_2' = \frac{K - Nk_1}{k_2}, \qquad n' = \frac{K}{k_1 + k_2},$$

$$n_1^* = g\frac{\bar{r}}{\sqrt{k_1}}, \qquad n_2^* = g\sqrt{\frac{1-\bar{r}^2}{k_2}}, \qquad g = \frac{K}{\bar{r}\sqrt{k_1} + \sqrt{\left(1-\bar{r}^2\right)k_2}}.$$

The vector of the regression estimators $\bar{\mathbf{t}}_{BS}$ from the double simple sample is not less precise than the vector of means $\bar{\mathbf{y}}_S$ from a simple sample drawn without replacement if:

$$u_1(\underline{n}_1, \underline{n}_2) \le u_2(\underline{n}) = \left(k_2^{-1}K - N^{-1}\right)q^2 . \tag{6.27}$$

This inequality is true

a) for the pair $(\underline{n}_1, \underline{n}_2) = (n_1^*, n_2^*)$, if:

$$\bar{r} \ge r_o = 2\frac{\sqrt{k_1 k_2}}{k_1 + k_2} \ge r_*^2 \tag{6.28}$$

where $r_* = \dfrac{k_1}{k_1 + k_2}$

b) for the pair $(\underline{n}_1, \underline{n}_2) = (N, n_2')$, if:

$$\bar{r} \ge r_{**} = N\sqrt{\frac{k_1 k_2}{K\left[N(k_1 + k_2) - K\right]}} \ge r_o \tag{6.29}$$

c) for the pair $(\underline{n}_1, \underline{n}_2) = (n_1', 2)$, if:

$$\bar{r} \ge r_{***} = \frac{K - 2k_2}{\sqrt{K(K - 2k_1 - 2k_2)}} \ge r_o \tag{6.30}$$

The necessary condition for $D^2(\bar{\mathbf{t}}_{BS}) \le D^2(\bar{y}_S)$ is that $k_2 > k_1$ and $\bar{r} \le r_0$, where r_o is determined by (6.28).

Let us note that if all the elements of the vector $\mathbf{a}=[a_1 \ldots a_m]$ are equal to one, the risk function is equal to the trace of the variance-covariance matrix of the estimator $\bar{\mathbf{t}}_{BS}$. Moreover, if $a_i = \bar{y}_i^{-2}$ $(i = 1, \ldots, m)$, the risk function is equal to the sum of variation coefficients of the elements of the vector $\bar{\mathbf{t}}_{BS}$.

6.3.2. Minimization of generalized variance

Let us denote the generalized variance of the estimator $\bar{\mathbf{t}}_{BS}$ as a function of the sample sizes in the following way $u_5(n_1,n_2)=\det\mathbf{V}\left(\bar{\mathbf{t}}_{BS}\right)$ where the matrix $\mathbf{V}\left(\bar{\mathbf{t}}_{BS}, P_{3d}\right)$ is defined by the equation (6.17). Our problem is to determine such sample sizes that the generalized variance of the estimator $\bar{\mathbf{t}}_{BS}$ takes the minimal value under fixed total costs of observation of variables. More precisely this problem is explained by the equation

$$\begin{cases} u_5(n_1,n_2) = \text{minimum} \\ k(n_1,n_2) \le K, \quad 2 \le n_2 < n_1 \le N \end{cases} \tag{6.31}$$

where the cost function $k(n_1,n_2)$ is defined by the expression (6.19).

Let us assume that the covariance matrix $\mathbf{C}_{*yy}$ is positively definite The well known theorems of linear algebra, see e.g. Rao (1982), let us find such a matrix $\mathbf{Q}$ of degree m that $\mathbf{Q}^T\mathbf{Q}=\mathbf{C}_{*yy}$ and $\mathbf{Q}^T\mathbf{R}_Q\mathbf{Q}=\mathbf{C}_{*yx}\mathbf{C}_{*xx}^{-1}\mathbf{C}_{*xy}$, where $\mathbf{R}_Q$ is the diagonal matrix with diagonal elements denoted by r_{Qi}^2, $i=1,\ldots,m$. The determinant $u_5(n_1,n_2)=\det\mathbf{V}\left(\bar{\mathbf{t}}_{BS}\right)$ can be written as follows:

$$u_5(n_1,n_2) = \det\left\{\mathbf{Q}^T\left(\frac{1}{n_1}\mathbf{R}_Q + \frac{1}{n_2}\left(\mathbf{I}_m - \mathbf{R}_Q\right) - \frac{1}{N}\mathbf{I}_m\right)\mathbf{Q}\right\}.$$

The well known properties of the matrix determinant let us rewrite the determinant in the following way:

$$u_5(n_1,n_2) = \det\left(\mathbf{C}_{*yy}\prod_{i=1}^{m} f_{*i}\right) \tag{6.32}$$

where:

$$f_{*i} = f_i - \frac{1}{N}, \qquad f_i = \frac{r_{Qi}^2}{n_i} + \frac{1-r_{Qi}^2}{n_2}. \tag{6.33}$$

The diagonal elements of the matrix $\mathbf{R}_Q$ can be determined as the solution to the following equation, see e.g. Rao (1982):

$$\det(\mathbf{C}_{*yx}\mathbf{C}_{*xx}^{-1}\mathbf{C}_{*xy} - r_Q^2\mathbf{C}_{*yy}) = 0.$$

The parameters $0 \le r_{Qi}^2 \le 1 (i=1,..,m)$ are the squared coefficients of canonical correlation. They measure the degree of linear dependence between the vectors *y* and *x*.

Lemma 6.1 [Wywiał (1992)]: The function $u_5(n_1,n_2)$ is strictly convex in the field $D=\{(n_1,n_2):0<n_2<n_1\le N\}$.

Proof: Let $a(w,z)=[w \quad z]\mathbf{H}(n_1,n_2)\begin{bmatrix} w \\ z \end{bmatrix}$ be the quadratic form, where $\mathbf{H}(n_1,n_2)$ is the hessian of the function u_5. After appropriate operations we have:

$$a(w,z)=2\sum_{i>j}^{m}\sum_{j=1}^{m}\frac{a_{ij}}{f_{*i}f_{*j}}+2\frac{z^2}{n_1^3}\sum_{i=1}^{m}\frac{r_{Qi}^2}{f_{*i}}+2\frac{w^2}{n_2^3}\sum_{i=1}^{m}\frac{1-r_{Qi}^2}{f_{*i}}$$

where:

$$a_{ij}=[wz]\begin{bmatrix} r_{Qi}^2 r_{Qj}^2 & r_{Qj}^2\left(1-r_{Qi}^2\right) \\ r_{Qi}^2\left(1-r_{Qj}^2\right) & \left(1-r_{Qi}^2\right)\left(1-r_{Qj}^2\right) \end{bmatrix}\begin{bmatrix} w \\ z \end{bmatrix}.$$

The function $u_5(w,z)$ is convex in *D* because $a\ge 0$ for all $z,w\in R-\{0\}$.

Let us introduce the following notation:

$$\begin{cases} n_1' = \min\left\{\dfrac{K-2k_2}{k_1}, N\right\} \\ n_2' = \max\left\{2, \dfrac{K-Nk_1}{k_2}\right\} \\ n'' = \dfrac{K}{k_1+k_2} \end{cases} \tag{6.34}$$

Let (n_1^*,n_2^*) be the solution to the following system of two equations:

$$n_1 = c\sqrt{\frac{1}{k_1}\sum_{i=1}^{m}\frac{r_{Qi}^2}{f_{*i}}}, \qquad n_2 = c\sqrt{\frac{1}{k_2}\sum_{i=1}^{m}\frac{1-r_{Qi}^2}{f_{*i}}} \tag{6.35}$$

where:

$$\frac{1}{c}=\frac{1}{k}\left\{\sqrt{k_1\sum_{i=1}^{m}\frac{r_{Qi}^2}{f_{*i}}}+\sqrt{k_2\sum_{i=1}^{m}\frac{1-r_{Qi}^2}{f_{*i}}}\right\}$$

Wywiał (1992, 1995) showed that the solution to the problem (6.31) is as follows:

$$\begin{cases}\underline{n}_1 &= \min\left\{n_1', n_1^*\right\}\\ \underline{n}_2 &= \max\left\{n_2', n_2^*\right\}\end{cases} \tag{6.36}$$

The solution (n_1^*, n_2^*) of the set of two nonlinear equations, given by the expression (6.35), can be obtained only approximately by means of an appropriate numerical method like Newton, gradient or iteration method. We are using the last one but under some additional assumptions.

If $N \to \infty$, $f_{*i} \to f_i$ for i=1,...,m, where the functions f_{*i} and f_i are defined by the expression (6.33). Next, in the first equation of the set (6.35), the transformation $n_2 = wn_1$ is implemented. After dividing the second equation of the set by the first one we have:

$$w = p(w) \tag{6.37}$$

where:

$$p(w)=\sqrt{\frac{k_1}{k_2}\frac{\sum_{i=1}^{m}\frac{1-r_{Qi}^2}{h_i}}{\sum_{i=1}^{m}\frac{r_{Qi}^2}{h_i}}}=\sqrt{\frac{k_1}{k_2}\left(\frac{\sum_{i=1}^{m}\frac{1}{h_i}}{\sum_{i=1}^{m}\frac{r_{Qi}^2}{h_i}}-1\right)}, \tag{6.38}$$

$$h_i = wr_{Qi}^2 + 1 - r_{Qi}^2, \qquad i=1,...m. \tag{6.39}$$

Theil (1979) proposed the average of squared coefficients of canonical correlation to asses the degree of linear dependence between two vectors of variables. In our case we have:

$$\bar{r}_Q^2 = \frac{1}{m'}\sum_{i=1}^{m} r_{Qi}^2 = \frac{1}{m'}\mathrm{tr}(\mathbf{C}_{*yx}\mathbf{C}_{*xx}\mathbf{C}_{*xy}) \tag{6.40}$$

where: m' = min{m,z}. Hence, $0 \le \bar{r}_Q^2 \le 1$ and we say that the degree of linear dependence between the vector of variables under study *y* and the auxiliary vector *x* increases when the value of the coefficient $\bar{r}_Q^2$ increases.

Lemma 6.2 [Wywiał (1992)]: If $\frac{m'}{m}\bar{r}_Q^2 > \frac{k_1}{k_2}$, then $0<w<1$ and $n_1>n_2$.

The following iteration process leads to the solution to the equation set (6.37):

$$w_{j+1} = p(w_j), \qquad j = 0,1,2,... \tag{6.41}$$

Wywiał (1992) proved that this iteration process is convergent on the true solution denoted by w_* and it is not dependent on the start solution $0<w_0<1$.

Demidowicz and Maron (1965) explain how to estimate the accuracy of an approximated solution from the iteration process:

$$| w_* - w_j | \le 2^{1-j} | w_1 - w_0 |.$$

This lets us determine the necessary number of iteration steps. The sufficiently accurate solution w_* leads to the optimal sample sizes (n_1^*, n_2^*). On the basis of the equations: $n_2 = wn_1$ and $k_1n_1+k_2n_2=K$ we have:

$$n_1^* = \frac{K}{k_1 + w_* k_2}, \qquad n_2^* = w_* n_1^*. \tag{6.42}$$

6.3.3. Goal optimization of sample sizes

Let us consider two types of problems. The first of them leads to finding a compromising solution which is close (in some sense) to all the optimal sample sizes obtained as a solution to particular problems. The second problem: for particular problems, the optimal values of the criterion functions are determined. Next, a purpose function is formed in such a way that it measures the distance between particular criterion functions and its optimal values, respectively. The solutions to these problems are treated as compromising sample sizes.

Let $(\underline{n}_{1i}, \underline{n}_{2i})$, i=1,...,m, be optimal sizes of the samples obtained as a solution to the following problem:

$$\begin{cases} u_{6i}(n_1, n_2) = \min \\ k_1 n_1 + k_{2i} n_2 \le K, \quad 2 \le n_2 < n_1 \le N \end{cases} \tag{6.43}$$

where $u_{6i}(n_1,n_2)=D^2(\bar{t}_{BiS},P_5)$ are defined by the expression (6.20). This optimal solution is a particular case of the solution to the problem (6.25) for m=1, and $k_2=k_{2i}$, given by the expression (6.26) where k_{2i} and r_i should be substituted for k_2 and $\bar{r}$, respectively. The parameter r_i is the multiple correlation coefficient between the variable under study y_i and auxiliary variables *x*.

In order to find the compromising solution to the problem, the following additional criterion can be introduced:

$$u_6(n_1,n_2) = \sum_{i=1}^{m}(\underline{n}_{1i}-n_1)^2 w_{1i} + \sum_{i=1}^{m}(\underline{n}_{2i}-n_2)^2 w_{2i} \qquad (6.44)$$

where $\sum_{i=1}^{m} w_{hi}=1$ and $w_{hi} > 0$ for each i=1,..,m, h=1,2. This leads to the following optimization problem:

$$\begin{cases} u_6(n_1,n_2)=\text{minimum} \\ k_1n_1+k_2n_2 \le K, \quad 2 \le n_2 < n_1 \le N \end{cases} \qquad (6.45)$$

where: $k_2 = \sum_{i=1}^{m} k_{2i}$. Hence, we are looking for such pair of sample sizes $(\underline{n}_1,\underline{n}_2)$ that it is close to all the particular solutions $(\underline{n}_{1i}, \underline{n}_{2i})$, i=1,..,m.

Wywiał (1992) showed the solution to the problem (6.45), when $n_1^* < n'$ and $n_2^* > n'$, where:

$$n' = \frac{K}{k_1+k_2}, \quad n_1^* = \bar{\underline{n}}_1 + ck_1, \qquad n_2^* = \bar{\underline{n}}_2 + ck_2 \qquad (6.46)$$

where:

$$c = \frac{K-k_1\bar{\underline{n}}_1-k_2\bar{\underline{n}}_2}{k_1^2+k_2^2}, \qquad \bar{\underline{n}}_h = \sum_{i=1}^{m}\underline{n}_{hi}w_{hi}. \qquad \text{for h=1,2}$$

In this case the solution to the problem (6.45) is determined by the expression (6.26) and the parameters (n_1^*, n_2^*).

Let $\underline{u}_{6i}(\underline{n}_{1i},\underline{n}_{2i})=\underline{u}_6$, i=1,...,m, be the minimal value of the purpose function defined in the problem (6.43). Let the coefficient of the relative estimation efficiency be defined by the expression:

$$e_i = \frac{\underline{u}_{6i}}{u_{6i}(n_1,n_2)}, \qquad i=1,...,m. \qquad (6.47)$$

The optimization problem is shown by the expression:

$$\begin{cases} u_7(n_1,n_2) = \min imum \\ k_1n_1 + k_2n_2 \le K, \qquad 2 \le n_2 < n_1 \le N \end{cases} \tag{6.48}$$

where:

$$u_7(n_1,n_2) = \sum_{i=1}^{m} \frac{1}{e_i}. \tag{6.49}$$

This problem is a particular case of the problem (6.25) for $\mathbf{a}=\left[\frac{1}{\underline{u}_{61}}...\frac{1}{\underline{u}_{6m}}\right]$. Hence, the expression (6.26) lets derive the solution to this problem.

6.4. Minimization of observation costs under fixed risk function

6.4.1. Fixed squared risk function

Let u_o be an admissible level of the risk function $u_1(n_1,n_2)$ defined by the expressions (6.20)-(6.24). Hence: $u_1(n_1,n_2) \le u_o$. This inequality is equivalent to the following one:

$$u_*(n_1,n_2) = \frac{\bar{r}^2}{n_1} + \frac{1-\bar{r}^2}{n_2} \le u_*$$

where: $\bar{r}$ is defined by the equation (6.24) and $u_* = \frac{1}{N} + \frac{u_o}{q^2}$ where: q^2 is determined by the formula (6.23). Let us consider the following optimization problem:

$$\begin{cases} k(n_1,n_2) = \min \\ u_{*1}(n_1,n_2) \le u_* \\ 2 \le n_2 < n_1 \le N \end{cases} \tag{6.50}$$

Let:

$$n_1'' = \frac{2\bar{r}^2}{2u_* + \bar{r}^2 - 1}, \qquad n_2'' = \frac{N(1-\bar{r}^2)}{Nu_* - \bar{r}^2}, \qquad n'' = \frac{1}{u_*}, \tag{6.51}$$

$$n_1^{**} = c\frac{\bar{r}}{\sqrt{k_1}}, \qquad n_2^{**} = c\sqrt{\frac{1-\bar{r}^2}{k_2}}. \tag{6.52}$$

where:

$$c = \frac{\bar{r}\sqrt{k_1} + \sqrt{(1-\bar{r}^2)k_2}}{u_*}. \tag{6.53}$$

Wywiał (1992) showed that if $q^2\left(\frac{1}{2}+\frac{1}{N}\right) > u_{o2}$ and $\bar{r} > r_* = \frac{k_1}{k_1+k_2}$, the solution to the problem is as follows:

$$\begin{cases} \underline{n}_1 = \min\left\{N, n_1'', n_1^{**}\right\} \\ \underline{n}_2 = \max\left\{n_2'', 2, n_2^{**}\right\} \end{cases} \tag{6.54}$$

6.4.2. Fixed standard errors of estimation of averages in a population

The problem is such a determination of samples sizes that the cost function $k(n_1,n_2)$, given by the expression (6.19), takes the minimal value under the restriction that the variances $D^2\left(\bar{t}_{iBS}, P_{3d}\right) \le d_i$, i=1,...,m and $1<n_2<n_1\le N$. On the basis of the expression (6.20) we have: $u_{1i}(n_1,n_2)\le d_{*i}$, i=1,...,m, where:

$$u_{1i}(n_1,n_2) = \frac{r_i^2}{n_1} + \frac{1-r_i^2}{n_2}, \qquad d_{*i} = \frac{d_i}{v_*(y_i)} + \frac{1}{N}. \tag{6.55}$$

Hence, the optimization problem is as follows:

$$\begin{cases} k(n_1,n_2) = \min \\ u_i(n_1,n_2) \le d_{*i}, \quad i=1,...,m. \\ 2 \le n_2 < n_1 \le N \end{cases} \tag{6.56}$$

The set of the admissible solutions D can be obtained as a product $D=\prod_{i=1}^{m} D_i$, where D_i is the intersection of the simplex determined by the inequalities $2\le n_2<n_1\le N$ and the convex area determined by the expression $u_{1i}(n_1,n_2)\le d_{*i}$. The set D is convex because it is the intersection of the convex sets D_i i=1,...,m.

Let $P_{ij}(n_{1ij},n_{2ij})$ be the intersection point of the hiperboloids $u_{1i}(n_1,n_2)=d_{*i}$, $u_{1j}(n_1,n_2)=d_{*j}$ $(i\ne j=1,...m)$. Its coordinates are as follows:

$$n_{1ij} = \frac{r_i^2(r_j^2 d_{*i} - r_i^2 d_{*j})}{q_{*i}(r_j^2 d - r_i^2 d) - (1-r_i^2)\left[r_j^2(1-r_i^2) - r_i^2(1-r_j^2)\right]}, \tag{5.57}$$

$$n_{2ij} = \frac{r_j^2 d_{*i} - r_i^2 d_{*j}}{r_j^2(1-r_i^2) - r_i^2(1-r_j^2)}. \tag{5.58}$$

We assume that the set *D* is non-empty. The solution to the problem is on the edge of the set *D* because it is convex and the purpose function $k(n_1,n_2)$ is linear. The optimal solution is equal to the coordinates of the apex of the set *D* or to the coordinates of the tangence point of the plain given by the equation $k_1n_1+k_2n_2=k_m$ and the plain given by $u_{1i}(n_1,n_2)=d_{*i}$ (i=1,..,m) in the set determined by the inequality $2\le n_2<n_1\le N$. The coordinates of one of the defined points are optimal sample sizes, if the value of the purpose function denoted by k_m takes a minimal value.

6.4.3. The admissible fixed generalized variance

Let the volume of the confidence ellipsoid for the vector population means be fixed. If the ellipsoid is determined on the basis of the vector of the regression estimators $\bar{\mathbf{t}}_{BS}$, its volume is proportionate to the generalized variance of $\bar{\mathbf{t}}_{BS}$. The admissible level of the volume is proportional to d, where d is the admissible level of the generalized variance of the strategy $(\bar{\mathbf{t}}_{BS}, P_{3d})$. Hence: $\det\mathbf{V}(\bar{\mathbf{t}}_{BS}, P_{3d}) = u_5(n_1,n_2) \le d$, Our purpose is to determine such sizes $(\underline{n}_1,\underline{n}_2)$ of the samples that the cost function $k(n_1,n_2)$, given by the expression (6.19), takes a minimal value. Hence:

$$\begin{cases} k(n_1,n_2) = \min \\ u_5(n_1,n_2) \le d, \quad 2 \le n_2 < n_1 \le N \end{cases} \tag{6.59}$$

where u_5 is defined by the equation (6.32).

Wywiał (1992) derived the following optimal solution:

$$\begin{cases} \underline{n}_1 = \min\{N, n_1'', n_1^*\} \\ \underline{n}_2 = \max\{2, n_2'', n_2^*\} \end{cases} \tag{6.60}$$

where n_1'' and n_2'' are determined by the expression (6.51) and (n_1^*, n_2^*) is the solution to the following system of the equations:

$$\begin{cases} \boldsymbol{\alpha}(n_1,n_2) = -k_1 k_2^{-2} \\ u_5(n_1,n_2) = d \end{cases} \tag{6.61}$$

where:

$$\boldsymbol{\alpha}(n_1, n_2) = -\frac{F_1(n_1, n_2)}{F_2(n_1, n_2)}. \tag{6.62}$$

The first partial derivatives in respect to n_1 and n_2 of the function $F(n_1,n_2)=u_5(n_1,n_2)-d=0$ are denoted by F_1 and F_2, respectively.

The solution to the problem exists if $n_2'' < N$ and $\underline{n}_1 > \underline{n}_2$.

When $N\to\infty$, the solution to the system (6.61) can be simplified in the following way: let $n_2=wn_1$. The first equation of the system (6.61) can be transformed into the form shown by the expression (6.37) and the second equation - into the following one:

$$\frac{1}{n_1^m}\det C_{*yy}\prod_{i=1}^{m}\left(r_{Qi}^2+\frac{1-r_{Qi}^2}{w}\right)=d. \tag{6.63}$$

The solution w_* to the equation (6.37) can be obtained by means of the iterative process determined by the expression (6.41). This and the equations (6.63) and $n_2=wn_1$ lead to the following optimal solution:

$$\begin{cases} n_1^* = \sqrt{\dfrac{1}{d}\det \mathbf{C}_{*yy}\prod_{i=1}^{m}\left(r_{Qi}^2+\dfrac{1-r_{Qi}^2}{w_*}\right)} \\ n_2^* = w_* n_1^* \end{cases} \tag{6.64}$$

Let us note that the conditions of existence of the inequalities $0<w<1$ are determined by the lemma 6.2.

6.5. Minimization of total risk

The total risk function is defined by the expression:

$$u_8(n_1,n_2)=u_1(n_1,n_2) + k(n_1,n_2)$$

where the functions $k(n_1,n_2)$ and $u_1(n_1,n_2)$ are determined by the expressions (6.19) and (6.20)-(6.24), respectively. Let us consider the following problem:

$$\begin{cases} u_8(n_1, n_2) = \min \\ 2 \le n_2 < n_1 \le N \end{cases} \tag{6.65}$$

Let us introduce the following notation:

$$n_1^* = \frac{q\bar{r}}{\sqrt{k_1}}, \qquad n_2^* = q\sqrt{\frac{1-\bar{r}^2}{k_2}}.$$

where $\bar{r}$ is defined by the expressions (6.23) and (6.24).

Wywiał (1992) showed that the set of admissible solutions L is the triangle determined by the points $A_1(N,2)$, $A_2(N,N)$ and $A_3(2,2)$ without the side $\overline{A_2A_3}$. If the point $A_*(n_1^*, n_2^*) \in L$, its coordinates are the optimal solution $\underline{A}(\underline{n}_1,\underline{n}_2)$. If $A_* \notin L$ then

a) $\underline{A} = (n_1^*,2)$ if $2 \le n_1^* \le N$ and $n_2^* < 2$

b) $\underline{A} = (N, n_2^*)$ if $n_1^* > N$ and $2 \le n_2^* < N$

c) $\underline{A} = (N,2)$ if $n_1^* \ge N$ and $n_2^* \le 2$

d) $\underline{A}$ do not exist if $n_1^* < 2$ and $n_2^* < 2$, or $n_1^* > N$ and $n_2^* > N$.

6.6. Unbiased regression strategies

6.6.1. Sampling strategy I

Let U be a population consisting of N distinct and identifiable units. The vector $\mathbf{y}^T = [y_1 \dots y_N]$ consists of all the values of a variable under study. Let $\mathbf{x} = [x_{ij}]$ be the matrix of the dimensions Nxk. The matrix $\mathbf{x}$ consists of all the values of a k-dimensional auxiliary variable. The element x_{ij} is an i-th value (i=1,...,N) of a j-th auxiliary variable (j=1,...,k≥1). Let $\mathbf{J}_N$ be the column vector of the dimensions Nx1. Each element of the vector $\mathbf{J}_N$ is equal to one. Let us define:

$$\bar{y} = \frac{1}{N}\mathbf{y}^T\mathbf{J}_N, \qquad \bar{\mathbf{x}} = \frac{1}{N}\mathbf{J}_N^T\mathbf{x},$$

$$\mathbf{Y} = \mathbf{y} - \mathbf{J}_N\bar{y}, \qquad \mathbf{X} = \mathbf{x} - \mathbf{J}_N\bar{\mathbf{x}},$$

$$v_{yy} = \frac{1}{N}\mathbf{Y}^T\mathbf{Y}, \qquad \mathbf{V} = [v_{jt}] = \frac{1}{N}\mathbf{X}^T\mathbf{X}, \qquad \mathbf{v} = [v_{yj}] = \frac{1}{N}\mathbf{X}^T\mathbf{Y}.$$

The population mean of the variable under study is denoted by $\bar{y}$. The row vector $\bar{\mathbf{x}}$ consists of the population means of the auxiliary variables. The population covariance matrix of the auxiliary variables is denoted by $\mathbf{V}$. The vector $\mathbf{v}$ consists of the population covariances between the auxiliary variables and the variable under study.

Let s be the sample of the size n drawn without replacement from a population U. Let $\mathbf{y}_s^T = [y_{i_1} \dots y_{i_n}]$ be the vector of values of the variable under study observed in the sample. Similarly, the matrix

$$\mathbf{x}_s = \begin{bmatrix} x_{i_1 1} \dots x_{i_1 k} \\ x_{i_2 1} \dots x_{i_2 k} \\ \dots\dots\dots\dots \\ x_{i_n 1} \dots x_{i_n k} \end{bmatrix}$$

consists of the values of the auxiliary variables observed in the sample s. Moreover, let us define:

$$\bar{y}_s = \frac{1}{n}\mathbf{y}_s^T \mathbf{J}_n, \qquad \bar{\mathbf{x}}_s = \mathbf{J}_n^T \mathbf{x}_s,$$

$$\mathbf{Y}_S = \mathbf{y}_s - \mathbf{J}_n \bar{y}, \qquad \mathbf{X}_S = \mathbf{x}_s - \mathbf{J}_n \bar{\mathbf{x}}, \qquad \overline{\mathbf{Y}}_S = \bar{\mathbf{y}}_s - \bar{\mathbf{y}}, \qquad \overline{\mathbf{X}}_S = \bar{\mathbf{x}}_s - \bar{\mathbf{x}},$$

$$\mathbf{V}_S = \frac{1}{n}\mathbf{x}_s^T \mathbf{x}_s - \bar{\mathbf{x}}_s^T \bar{\mathbf{x}}_s, \qquad \mathbf{v}_S = \frac{1}{n}\mathbf{x}_s^T \mathbf{y}_s - \bar{\mathbf{x}}_s^T \bar{\mathbf{y}}_s,$$

or

$$\mathbf{V}_S = \frac{1}{n}\mathbf{X}_S^T \mathbf{X}_S - \overline{\mathbf{X}}_S^T \overline{\mathbf{X}}_S, \qquad \mathbf{v}_S = \frac{1}{n}\mathbf{X}_S^T \mathbf{Y}_S - \overline{\mathbf{X}}_S^T \overline{\mathbf{Y}}_S.$$

Hence, $\mathbf{V}_S = \left[v_{sij}\right]$ is the sample covariance matrix between the auxiliary variables and $\boldsymbol{v}_s$ is the column vector of the sample covariances between the auxiliary variables and the variable under study.

Let $\boldsymbol{S}$ be the sample space of the unordered sample s of the size n selected without replacement from the population U.

In the paragraph 2.3.10 sampling design $P_{16}(s) \propto \det\mathbf{V}_s$ is considered. The sample s is selected with a probability proportionate to the sample generalized variance of the auxiliary variables:

$$P_{16}(s) = c_1 \frac{\det \boldsymbol{V}_s}{\det \boldsymbol{V}} \tag{6.66}$$

where:

$$c_1 = \binom{N-k-1}{n-k-1}^{-1} \left(\frac{n}{N}\right)^{k+1}.$$

When k=1, the sampling plan $P_{16}(s)$ is reduced to the one proportional to the sample variance of Singh and Srivastava (1980). Let s_{k+1} be the subset of the sample s. The size of the subset s_{k+1} is equal to k+1< n.

Let us define the following quantity:

$$q_1(s_{k+1}) = \det^2\left[\mathbf{J}_{k+1} \quad \mathbf{x}_{s_{k+1}}\right] \tag{6.67}$$

where:

$$\mathbf{x}_{s_{k+1}} = \begin{bmatrix} x_{i_1 1} \ldots x_{i_1 k} \\ \ldots\ldots \\ x_{i_k 1} \ldots x_{i_k k} \\ x_{i_{k+1} 1} \ldots x_{i_{k+1} k} \end{bmatrix}.$$

Let $\mathbf{x}_{r*} = \left[x_{i_r 1} ... x_{i_r k}\right]$ be the r-th row of the matrix $\mathbf{x}_{s_{k+1}}$. After eliminating the row $\mathbf{x}_{k+1*}$ in the matrix $\mathbf{x}_{s_{k+1}}$ we obtain the matrix $\mathbf{x}_{s_k}$. After subtracting the last row of the matrix $\left[\mathbf{J}_{k+1} \quad \mathbf{x}_{s_{k+1}}\right]$ from the previous rows of this matrix we have:

$$\det\left[\mathbf{J}_{k+1} \quad \mathbf{x}_{s_{k+1}}\right] = \det\begin{bmatrix} \mathbf{0} & \mathbf{x}_{s_k} - \mathbf{x}_{k+1*}\mathbf{J}_k \\ 1 & \mathbf{x}_{k+1*} \end{bmatrix}.$$

This lets us rewrite the expression (6.67) in the following way:

$$q_1(s_{k+1}) = \det^2\left[\mathbf{x}_{s_k} - \mathbf{x}_{k+1*}\mathbf{J}_k\right]. \tag{6.68}$$

Let us note that $q_1(s_{k+1})$ is the k-dimensional measure (volume) of the parallelotop spanned by the vectors with their origins at the same point $\mathbf{x}_{k+1*}$ and the end points $\mathbf{x}_{1*},...,\mathbf{x}_{k*}$, see e.g. Borsuk (1969).

The following sampling scheme (implementing the sampling design $P_{16}(s)$) consists of the two following steps.

Step 1: Select k+1 units $s_{k+1} = \{i_1, i_2, ..., i_{k+1}\}$ with their probability of joint selection being proportional to $q_1(s_{k+1})$.

Step 2: Select (*n*-*k*-1) units from the remaining units of the population by the simple random sampling without replacement.

The sampling design proportional to $q_1(s_{k+1})$ is as follows:

$$P(s_{k+1}) = \frac{q_1(s_{k+1})}{N^{k+1} \det \boldsymbol{V}} \tag{6.69}$$

or

$$P(s_{k+1}) = \frac{(k+1)\det\left(\mathbf{U}_{s_{k+1}}^{T}\mathbf{U}_{s_{k+1}}\right)}{N^{k+1}\det \boldsymbol{V}}. \tag{6.70}$$

Particularly, if k=1, this sampling scheme is reduced to the sampling scheme proposed by Singh and Srivastava (1980). In this case $q_1(s_2) = \left(x_{i_2} - x_{i_1}\right)^2$.

The well known regression estimator of the population mean $\bar{y}$ is as follows:

$$\bar{y}_{RS} = \bar{y}_S - (\bar{\mathbf{x}}_S - \bar{\mathbf{x}})\mathbf{B}_S \tag{6.71}$$

where:

$$\mathbf{B}_S = \mathbf{V}_S^{-1}\mathbf{v}_S. \tag{6.72}$$

Let us introduce the following matrix

$$\mathrm{A_S} = \begin{bmatrix} \bar{\mathrm{y}}_\mathrm{S} - \bar{\mathrm{y}} & \bar{\mathbf{x}}_\mathrm{S} - \bar{\mathbf{x}} \\ \mathbf{v}_\mathrm{S} & \mathbf{V}_\mathrm{S} \end{bmatrix}. \tag{6.73}$$

Wywiał (1999), on the basis of well-known property of the determinant of a block matrix rewrote, the estimator $\bar{\mathrm{y}}_{\mathrm{RS}}$ in the following way:

$$\bar{\mathrm{y}}_{\mathrm{RS}} = \bar{\mathrm{y}} + \frac{\det \boldsymbol{A}_\mathrm{S}}{\det \boldsymbol{V}_\mathrm{S}}. \tag{6.74}$$

The determinant of the matrix $\boldsymbol{A}_s$ can be transformed into the following forms:

$$\det \mathbf{A}_\mathrm{S} = \mathrm{n}^{-\mathrm{k}-1} \det \begin{bmatrix} \mathrm{n}\left(\bar{\mathrm{y}}_\mathrm{S} - \bar{\mathrm{y}}\right) & \mathrm{n}\left(\bar{\boldsymbol{x}}_\mathrm{S} - \bar{\boldsymbol{x}}\right) \\ \mathbf{X}_S^T \mathbf{Y}_\mathrm{S} & \mathbf{X}_\mathrm{S}^\mathrm{T} \mathbf{X}_\mathrm{S} \end{bmatrix} \tag{6.75}$$

or

$$\det \mathbf{A}_\mathrm{S} = \mathrm{n}^{-\mathrm{k}-1} \det \begin{bmatrix} \mathbf{J}_\mathrm{n}^\mathrm{T} \\ \mathbf{X}_\mathrm{S}^\mathrm{T} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_\mathrm{S} \mathbf{X}_\mathrm{S} \end{bmatrix}.$$

The strategy $\left(\bar{\mathrm{y}}_{\mathrm{RS}}, \mathrm{P}_{16}(\mathrm{s})\right)$ is the unbiased strategy of the population mean $\bar{\mathrm{y}}$. When the sample size $\mathrm{n} \to \infty$ and the population size $\mathrm{N} \to \infty$ in such a way that $\mathrm{N-n} \to \infty$,

$$\mathrm{D}^2\left(\bar{\mathrm{y}}_{\mathrm{RS}}, \mathrm{P}_{16}(\mathrm{s})\right) \approx \frac{1}{\mathrm{n}}\left(\mathrm{v}_{\mathrm{yy}} - \mathbf{v}^\mathrm{T} \mathbf{V}^{-1} \mathbf{v}\right). \tag{6.76}$$

Let $\mathbf{R}$ be the correlation matrix of auxiliary variables and $\mathbf{r}^\mathrm{T} = \left[\mathrm{r}_{\mathrm{y1}} \ldots \mathrm{r}_{\mathrm{yk}}\right]$, where r_{yj} is the correlation coefficient between the j-th auxiliary variable and the variable under study. This lets us rewrite the expression (6.76) in the following way:

$$\mathrm{D}^2\left(\bar{\mathrm{y}}_{\mathrm{RS}}, \mathrm{P}_{16}(\mathrm{s})\right) \approx \frac{1}{\mathrm{n}} \mathrm{v}_{\mathrm{yy}} \left(1 - \mathrm{r}_\mathrm{w}^2\right) \tag{6.77}$$

where:

$$\mathrm{r_w} = \sqrt{\mathbf{r}^\mathrm{T} \mathbf{R}^{-1} \mathbf{r}}.$$

is the multiple correlation coefficient between the auxiliary variables and the variable under study. Hence, in the asymptotic case, the precision of the strategy $(\bar{y}_{RS}, P_{16}(s))$ increases when the value of the multiple correlation coefficient r_w increases, too.

Singh and Srivastava (1980) derived the unbiased estimator of the variance $D^2(\bar{y}_{RS}, P_{16}(s))$ in the case when k=1. Generalizing their result for k≥1, we can construct the following unbiased estimator of this parameter.

$$\hat{D}_1^2 = \bar{y}_{RS}^2 - \frac{N^{k-1}}{n^{k+1}} \frac{\prod_{h=1}^{k}(n-h)}{\prod_{h=1}^{k}(N-h)} \frac{\det \mathbf{V}}{\det \mathbf{V}_S} \left[\sum_{i \in S} y_i^2 + \frac{N-1}{n-1} \sum_{i \neq j \in S} y_i y_j \right]. \qquad (6.78)$$

6.6.2. Sampling strategy II

In the paragraph 2.3.9 the sampling design $P_{14}(s) \propto \det \mathbf{V}_{\#S}$ is considered, where

$$\mathbf{V}_{\#S} = \frac{1}{n} \mathbf{X}_S^T \mathbf{X}_S = \frac{1}{n} (\mathbf{x}_S - \mathbf{J}_n \bar{\mathbf{x}})^T (\mathbf{x}_S - \mathbf{J}_n \bar{\mathbf{x}}) \qquad (6.79)$$

is the sample variance-covariance matrix defined on the basis of the population means of the auxiliary variables. We can prove that

$$P_{14}(s) = c_2 \frac{\det \mathbf{V}_{\#S}}{\det \mathbf{V}} \qquad (6.80)$$

where:

$$c_2 = \frac{1}{\binom{N-k}{n-k}} \left(\frac{n}{N} \right)^k .$$

When k=1, the defined sampling design is reduced to the one proposed by Singh and Srivastava (1980).

Let s_k be a subset of the sample s and $k < n$. Let us define the following:

$$q_2(s_k) = \det^2 \begin{bmatrix} 1 & \bar{\mathbf{x}} \\ \mathbf{J}_k & \mathbf{x}_{S_k} \end{bmatrix}. \qquad (6.81)$$

From the geometrical point of view $q_2(s_k)$ is the k-dimensional measure (volume) of the parallelotop spanned by the vectors with their origins at the same point $\bar{\boldsymbol{x}}$ and the end points which determine the rows of the matrix $\mathbf{x}_{S_k}$,see, e.g., Anderson (1958) or Borsuk (1969). The expression (6.81) can be transformed into the following one:

$$q_2(s_k) = \det^2\left(\mathbf{X}_{S_k}\right) \tag{6.82}$$

It can be shown that the following sampling scheme implements the sampling design $P_{14}(s)$:

Step 1: Select k units $s_k=\{i_1, ..., i_k\}$ with their probability of joint selection being proportional to $q_2(s_k)$.

Step 2: Select (n-k) units from the remaining units of the population by simple random sampling without replacement.

When k=1, the introduced sampling scheme is reduced to one proposed by Singh and Srivastava (1980). In this case $q_2\left(s_1\right) = \left(x_j - \bar{x}\right)^2$.

Wywiał (1999) considered the following estimator:

$$\bar{y}_{\#S} = \frac{n(N-k)}{N(n-k)}\left[\bar{y}_S - \left(\bar{\mathbf{x}}_S - \bar{\mathbf{x}}\right)\mathbf{B}_{\#S}\right] \tag{6.83}$$

where:

$$\mathbf{B}_{\#S} = \mathbf{V}_{\#S}^{-1}\mathbf{v}_{\#S}, \qquad \mathbf{v}_{\#S} = \frac{1}{n}\mathbf{X}_S^T\mathbf{y}_S .$$

The statistic $\bar{y}_{\#S}$ can be transformed into the following one:

$$\bar{y}_{\#S} = \frac{n(N-k)}{N(n-k)}\frac{\det \mathbf{A}_{\#S}}{\det \mathbf{V}_{\#S}} \tag{6.84}$$

where:

$$\mathbf{A}_{\#S} = \begin{bmatrix} \bar{\mathbf{y}}_S & \bar{\mathbf{x}}_S - \bar{\mathbf{x}} \\ \mathbf{v}_{\#S} & \mathbf{V}_{\#S} \end{bmatrix} \qquad \text{or} \qquad \mathbf{A}_{\#S} = \frac{1}{n}\begin{bmatrix} \mathbf{J}_n^T \\ \mathbf{X}_S^T \end{bmatrix}\begin{bmatrix} \mathbf{y}_S & \mathbf{X}_S \end{bmatrix}.$$

In the case when k = 1, the statistic $\bar{y}_{\#S}$ is reduced to the estimator proposed by Singh and Srivastava (1980).

We can prove that $\left(\bar{y}_{\#S}, P_{14}(s)\right)$ is the unbiased strategy of the population mean $\bar{y}$. The approximate variance of the strategy $\left(\bar{y}_{\#S}, P_{14}(s)\right)$ is expressed by the right side of the equation (6.77). The unbiased estimator of the variance of the sampling strategy $\left(\bar{y}_{\#S}, P_{14}(s)\right)$ is as follows:

$$\hat{D}_2^2 = \bar{y}_{\#S}^2 - \frac{N^{k-3}}{n^{k-1}} \frac{\prod_{h=2}^{k}(n-h+1)}{\prod_{h=2}^{k}(N-h+1)} \frac{\det \mathbf{V}_{\#}}{\det \mathbf{V}_{\#S}} \left[\sum_{i \in S} y_i^2 + \frac{N-1}{n-1} \sum_{i \neq j \in S} y_i y_j \right] \tag{6.85}$$

6.6.3. Sampling strategy III

Let us assume that a population U is divided into disjoint and non-empty clusters U_g, g=1,...,G, and $U = \bigcup_{g=1}^{G} U_g$. The size of the sample U_g is denoted by $N_g > 1$ and $N = \bigcup_{g=1}^{G} N_g$. A multivariate auxiliary variable of the dimension k is observed in all first-stage units. The sample S, consisting of clusters, is selected at the first stage. The sampling design is proportional to the sample generalized variance of the auxiliary variables and it is determined by the expression (6.66). At the second stage, simple samples $Q_{g_1,\ldots}, Q_{g_n}$ are drawn without replacement from the clusters $U_{g_1,\ldots}, U_{g_n}$, selected at the first stage, respectively. The size of the sample Q_{g_j} is denoted by $1 < m_{g_j} \leq N_{g_j}$. The two-stage sample will be denoted by Q={S,Q_1,...,Q_n} and its outcome by q={s,q_1,...,q_n}. The sampling design is as follows:

$$P_{19}(q) = P_{16}(s) \prod_{g \in s} \binom{N_g}{m_g}^{-1} \tag{6.86}$$

where the sampling design $P_{16}(s)$ is determined by the equation (6.66).

Let us introduce the following notation:

$$z_g = \sum_{i \in U_g} y_i, \qquad \bar{z} = \frac{1}{G} \sum_{g=1}^{G} z_g, \qquad v_{zz} = \frac{1}{G-1} \sum_{g=1}^{G} (z_g - \bar{z})^2,$$

$$\bar{y}_{U_g} = \frac{1}{N_g} z_g, \qquad v_{U_g} = \frac{1}{N_g - 1} \sum_{i \in U_g} (y_i - \bar{y}_{U_g})^2,$$

$$z_{Q_g} = \sum_{i \in Q_g} y_i, \qquad \bar{y}_{Q_g} = \frac{1}{m_g} z_{Q_g},$$

$$\tilde{z}_{Q_g} = N_g \bar{y}_{Q_g}, \tag{6.87}$$

$$v_{Q_g} = \frac{1}{m_g - 1} \sum_{i \in Q_g} \left(y_i - \bar{y}_{Q_g} \right)^2 . \tag{6.88}$$

Let us consider the following estimator of the parameter $\bar{z}$

$$\bar{z}_{RQ} = \bar{z}_Q - (\bar{\mathbf{x}}_S - \bar{\mathbf{x}})\mathbf{B}_Q \tag{6.89}$$

where:

$$\bar{z}_Q = \frac{1}{n} \sum_{g \in S} \tilde{z}_{Q_g} , \tag{6.90}$$

$$\mathbf{B}_Q = \mathbf{V}_S^{-1} \mathbf{w}_Q , \tag{6.91}$$

$$\mathbf{w}_Q^T = \left[c_Q(z, x_1) \quad ... \quad c_Q(z, x_k) \right],$$

$$c_Q(z, x_j) = \frac{1}{n-1} \sum_{g \in S} \left(\tilde{z}_{Q_g} - \bar{z}_Q \right)\left(x_{gj} - \bar{x}_{jS} \right),$$

or

$$\mathbf{w}_Q = \frac{1}{n} \mathbf{X}_S^T \tilde{\mathbf{z}}_Q - \bar{\mathbf{X}}_S^T \bar{z}_Q \qquad \text{or} \qquad \mathbf{w}_Q = \frac{1}{n} \mathbf{x}_S^T \tilde{\mathbf{z}}_Q - \bar{\mathbf{x}}_S^T \bar{z}_Q$$

where: $$\tilde{\mathbf{z}}_Q = \begin{bmatrix} \tilde{z}_{Q_1/S} \\ ..... \\ \tilde{z}_{Q_1/S} \end{bmatrix}.$$

The regression estimator $\bar{z}_{RQ}$ from the two-stage sample is the unbiased estimator of the parameter $\bar{z}$. The variance of the strategy is as follows:

$$D^2\left(\bar{z}_{RQ}, P_{19}\right) =$$

$$= \frac{1}{n^2} E_S \left\{ \sum_{g \in S} D^2_{Q_g/S}\left(\tilde{z}_{Q_i}\right) + \bar{\mathbf{X}}_S \mathbf{V}_S^{-1} \mathbf{X}_S^T \operatorname{diag}\left(\mathbf{D}^2_{Q/S}\left(\tilde{\mathbf{z}}_Q\right)\right) \mathbf{X}_S \mathbf{V}_S^{-1} \bar{\mathbf{X}}_S^T + \right.$$

$$- 2\bar{\mathbf{X}}_S \mathbf{V}_S^{-1} \mathbf{X}_S^T \mathbf{D}^2_{Q/S}\left(\tilde{\mathbf{z}}_Q\right)\left(1 + \bar{\mathbf{X}}_S \mathbf{V}_S^{-1} \bar{\mathbf{X}}_S^T\right) +$$

$$\left. + \bar{\mathbf{X}}_S \mathbf{V}_S^{-1} \bar{\mathbf{X}}_S^T \left(2 + \bar{\mathbf{X}}_S \mathbf{V}_S^{-1} \bar{\mathbf{X}}_S^T\right) \sum_{g \in S} D^2_{Q_g/S}\left(\tilde{z}_{Q_g}\right) \right\} + D_S^2\left(\bar{z}_{RS}\right) \tag{6.92}$$

where:

$$\mathbf{D}^2_{Q/S}\left(\tilde{\mathbf{z}}_Q, P_{19}\right) = \begin{bmatrix} D^2_{Q_1/S}\left(\tilde{z}_{Q_1}, P_{19}\right) \\ \ldots\ldots\ldots \\ D^2_{Q_n/S}\left(\tilde{z}_{Q_n}, P_{19}\right) \end{bmatrix}$$

$$D^2_{Q_g/S}\left(\tilde{z}_{Q_g}, P_{19}\right) = \frac{N_g\left(N_g - m_g\right)}{m_g} v_{U_g}\,.$$

The approximate value of the variance $D^2\left(\bar{z}_{RS}, P_{19}\right)$ is as follows:

$$D^2\left(\bar{z}_{RS},\ P_{19}\right) \approx \frac{1}{n} v_{zz}\left(1 - r_w^2\right) \tag{6.93}$$

where r_w is the multiple correlation coefficient between the auxiliary variables and the variable z.

Let us consider the following statistic:

$$\breve{z}_{Q_g} = \frac{m_g\left(N_g - 1\right)}{N_g\left(m_g - 1\right)}\left(\tilde{z}^2_{Q_g} - \frac{N_g^2}{m_g}\frac{N_g - m_g}{N_g - 1} M_{Q_g}\right) \tag{6.94}$$

where:

$$M_{Q_g} = \frac{1}{m_g}\sum_{i \in Q_g} y_i^2\,.$$

The following statistic is the unbiased estimator of the variance $D^2\left(\bar{z}_{RQ}, P_{19}\right)$:

$$\hat{D}^2_{3,Q} = \bar{z}^2_{RQ} - \frac{N^{k-1}}{n^{k+1}} \frac{\prod_{h=1}^{k}(n-h)}{\prod_{h=1}^{k}(N-h)} \frac{\det \boldsymbol{V}}{\det \boldsymbol{V}_S}\left[\sum_{g \in S} \breve{z}_{Q_g} + \frac{N-1}{n-1}\sum_{g \neq t \in S} \tilde{z}_{Q_g}\tilde{z}_{Q_t}\right]. \tag{6.95}$$

Finally, the unbiased estimator of the population total is as follows:

$$t_{RQ} = G\bar{z}_{RQ}\,, \tag{6.96}$$

The unbiased estimator of population average is determined by the expression:

$$\bar{t}_{RQ} = \frac{G}{N}\bar{z}_{RQ}\,, \tag{6.97}$$

The variances of these estimators can be easily derived on the basis of the expression (6.92). The equation (6.95) leads to construction of the estimators of these variances.

Let us note that the statistics $\bar{t}_{RQ}$, defined by the expression (6.96), can be treated, in some sense, as a particular case of the regression estimator of the total value from two stage sample considered by Särndal, Swensson, and Wretman (1992).

6.6.4. Example of simulation analysis of accuracy

Let us consider the example of average estimation by means of regression strategies: $(\bar{y}_{RS}, P_3(s))$, $(\bar{y}_{RS}, P_{16}(s))$, $(\bar{y}_{\#S}, P_3(s))$ and $(\bar{y}_{\#S}, P_{14}(s))$, where $P_3(s)$ is the sample design of the simple sample drawn without replacement determined by the equation (1.28). The variable under study is: the revenues from 1985 municipal taxation (in millions of kronor) - y, the auxiliary variables are: the number of Conservative seats in municipal councils –x_1, the number of Social Democratic seats in municipal councils - x_2 and the real estate values according to 1984 assessment (in millions of kronor) - x_3. These variables are observed in the population of Swedish municipalities. The population is divided into eight strata according to geographical regions of Sweden. The data are published by Sarndal Swensson, Wretman (1992). We are going to consider only the data consisting of 15 municipalities in the seventh stratum. The tables 4.1-4.6 let compare the accuracy of the estimation of the mean of the revenues from 1985 municipal taxation. The relative efficiencies are determined by the expression:

$$e = \frac{D^2(\bar{y}_{RS}, P_{16})}{D^2(\bar{y}_{RS}, P_3)} 100\% \qquad \text{or} \qquad e = \frac{D^2(\bar{y}_{\#S}, P_{14})}{D^2(\bar{y}_{\#S}, P_3)} 100\% .$$

Table 6.1

The accuracy of strategies $(\bar{y}_{RS}, P_3(s))$ and $(\bar{y}_{RS}, P_{16}(s))$.

The auxiliary variable x_2

The size of the sample	3	4	5	6	7
The bias under the plan P_3	-45	-28	-20	-16	-11
The variance under the plan P_3	14867	5508	33230	2415	1785
The variance under the plan P_{16}	6814	4288	2976	2160	1606
The relative efficiency e	45.8	77.9	89.4	89.5	90.0

Analysis of the tables 6.1-6.3 leads us to the following conclusions: absolute value of the bias of $(\bar{y}_{RS}, P_3(s))$ decreases when the number of variables increases. Similarly, the absolute value of the bias decreases

when the size of the sample increases for particular sets of auxiliary variables. The variances of both strategies $(\bar{y}_{RS}, P_{16}(s))$ and $(\bar{y}_{RS}, P_3(s))$ decrease when sample size becomes greater and greater. In all the cases the precision of the $(\bar{y}_{RS}, P_{16}(s))$ strategy is better than the precision of the $(\bar{y}_{RS}, P_3(s))$ strategy.

Table 6.2

The accuracy of $(\bar{y}_{RS}, P_3(s))$, $(\bar{y}_{RS}, P_{16}(s))$ strategies.

The auxiliary variables: x_1, x_2

The size of the sample	4	5	6	7
The bias under the plan P_3	-16	-12	-6	-5
The variance under the plan P_3	6021	2847	2537	1561
The variance under the plan P_{16}	2780	1739	1434	1020
The relative efficiency e	46.2	61.1	56.5	65.3

Table 6.3

The accuracy comparison of $(\bar{y}_{RS}, P_3(s))$ and $(\bar{y}_{RS}, P_{16}(s))$ strategies.

The auxiliary variables: x_1, x_2 and x_3

The size of the sample	5	6	7
The bias under the plan P_3	0	-4	-3
The variance under the plan P_3	33769	3214	1499
The variance under the plan P_{16}	2036	1242	855
The relative efficiency e	6.0	38.6	57.0

The relative efficiency coefficient increases when the sample size increases. Hence, we can expect that the accuracy of the $(\bar{y}_{RS}, P_{16}(s))$ strategy is not much better than the accuracy of the $(\bar{y}_{RS}, P_3(s))$ strategy in the large sample.

Table 6.4

The accuracy of $(\bar{y}_{\#S}, P_3(s))$ and $(\bar{y}_{\#S}, P_{14}(s))$ strategies.

The auxiliary variable: x_2

The size of the sample	2	3	4	5	6	7
The bias under the plan P_3	-39.2	-29	-22	-18	-14	-12
The variance under the plan P_3	29508	14355	8689	5778	4045	2912
The variance under the plan P_{14}	42285	17499	9727	6104	4084	2831
The relative effic. e	143.3	121.9	112.0	105.7	101.0	97.3

Table 6.5

The accuracy of $(\bar{y}_{\#S}, P_3(s))$, $(\bar{y}_{\#S}, P_{14}(s))$ strategies.
The auxiliary variables x_1, x_2

The size of the sample	3	4	5	6	7
The bias under the plan P_3	-39	-32	-26	-21	-16
The variance under the plan P_3	33461	15014	8846	5813	4003
The variance under the plan P_{14}	33912	14026	7692	4717	3051
The relative efficiency e	101.4	93.4	87.0	81.1	76.2

Table 6.6

The accuracy comparison of $(\bar{y}_{\#S}, P_3(s))$ and $(\bar{y}_{\#S}, P_{14}(s))$ strategies.
The auxiliary variables: x_1, x_2 and x_3

The size of the sample	4	5	6	7
The bias under the plan P_3	-60	-45	-34	-25
The variance under the plan P_3	34551	16940	10380	6764
The variance under the plan P_{14}	41342	16663	8844	5191
The relative efficiency e	119.7	98.4	85.2	76.7

The analysis of the tables 6.1-6.6 lets us infer that in small samples the regression estimator $\bar{y}_{\#s}$ from the simple sample can be more precise than the $(\bar{y}_{\#s}, P_{14}(s))$ strategy. In all the cases the variance of the $(\bar{y}_{RS}, P_{16}(s))$ strategy is smaller than the variances of the $(\bar{y}_{\#S}, P_{14}(s))$, $(\bar{y}_{\#S}, P_3(s))$ and $(\bar{y}_{RS}, P_3(s))$ strategies. The same is with their mean square errors.

Simulation analysis of the accuracy of the considered strategies was evaluated by Gamrot and Wywiał (2002).

SUMMARY IN POLISH

W praktyce badań reprezentacyjnych zwykle mamy do czynienia z problemem wnioskowania o wielu parametrach analizowanych cech populacji. Rzadko celem takiego badania jest ocena wartości jednego parametru, chociaż temu właśnie przypadkowi jest głównie poświęcana większość prac z metody reprezentacyjnej. Bynajmniej nie oznacza to, iż te prace mijają się z praktycznymi potrzebami badań statystycznych, ponieważ otrzymywane wyniki dotyczące wnioskowania o pojedynczym parametrze jednowymiarowej cechy można w wielu zagadnieniach bezpośrednio uogólnić na przypadek wielowymiarowy. W tej dziedzinie są jednak problemy jednoczesnego wnioskowania o wielu parametrach, które wymagają szczególnego podejścia. Należą do nich problem sposobu oceny dokładności estymacji wektora parametrów oraz interpretacja używanych do tego celu wskaźników. Kluczowe znaczenie ma także usystematyzowanie podstawowych wiadomości pozwalających na porównywanie dokładności estymatorów wektorowych. Następna kwestia dotyczy optymalizacji badań próbkowych, a zwłaszcza optymalizacji rozmiarów prób złożonych, gdy występują ograniczone nakłady na badania reprezentacyjne oraz żądania spełnienia wymaganej dokładności oceny parametrów.

W ogólności właśnie wymienionym problemom jest poświęcona niniejsza praca. Prezentowano w niej głównie zagadnienia dotyczące jednoczesnej estymacji wielu parametrów cech w populacji. Nacisk położono na prezentację wyników otrzymanych przez autora.

W pracy ograniczono się głównie do analizy problemu estymacji wektora wartości średnich w populacji. Otrzymane na tym polu wyniki można jednak łatwo przenieść na zagadnienie oceny innych ważnych z punktu widzenia praktyki parametrów, takich jak suma wartości cechy w populacji, ilość elementów z cechą wyróżnioną w populacji, częstość względna występowania określonego zjawiska w populacji.

W pierwszym rozdziale przedstawiono podstawowe definicje związane z rozkładami cech w populacji ustalonej, jak i w tzw. nadpopulacji. Szczegółowiej potraktowano problem interpretacji miar zróżnicowania wartości wielowymiarowej zmiennej. Przedstawiono podstawowe wiadomości o własnościach planów, schematów losowania oraz strategii losowania. Ponadto, są prezentowane definicje i twierdzenia, które zwykle są bezpośrednimi uogólnieniami na przypadek wielowymiarowy odpowiednich określeń znanych z przypadku wnioskowania o jednowymiarowym parametrze. Szczególny nacisk położono na problem porównywania dokładności estymatorów wektorowych.

Podstawowe parametry rozkładu wektora znanych estymatorów Horvitza-Thompsona są prezentowane w drugim rozdziale. Wyznaczano tu również w przybliżony sposób wariancje tego estymatora dla wybranych planów losowania zależnych od cech pomocniczych. Uwzględniono również plany losowania prób z populacji przestrzennych zależne od położenia względem siebie elementów populacji.

W trzecim rozdziale prezentowano podstawowe własności rozkładu wektora estymatorów z próby warstwowej. Konstruowano i rozwiązywano zadania optymalnej lokalizacji prób w warstwach. Zagadnienie optymalnego tworzenia warstw w populacji ograniczono do problemu wykorzystania formalnych metod grupowania na podstawie obserwowanych cech w populacji do wyodrębniania w niej warstw. Wskazano także na możliwość wykorzystania takich metod do warstwowania próby prostej wylosowanej z populacji. Z tak utworzonych warstw w następnym kroku są losowane próby proste, w których już obserwuje się cechy badane. Problem ten jest podobny do znanego zagadnienia warstwowania próby po jej wylosowaniu. Przedstawiono również estymację wartości średniej na podstawie kombinacji liniowej przeciętnych z podprób wyróżnianych w pierwotnie wylosowanej próbie prostej z populacji. Te podpróby również są używane do wyodrębniania jednorodnych warstw w populacji na podstawie cech dodatkowych. Liczebności tych warstw prowadzą do wyznaczenia współczynników kombinacji liniowej. Do tworzenia warstw są wykorzystywane m.in. odpowiednie metody klasyfikacji danych. Wprowadzono tutaj również kryterium reprezentatywności próby konstruowane na podstawie definicji zanurzenia punktu w zbiorze punktów.

Podstawowe parametry rozkładu wektora średnich z próby grupowej prezentowano w rozdziale czwartym. Parametry te przedstawiono jako funkcje tzw. współczynników korelacji wewnątrz-grupowej. Wprowadzono również współczynnik jednorodności rozkładu wartości wielowymiarowej zmiennej. Podjęto problem optymalnego wyróżniania grup w populacji bądź w wylosowanej próbie prostej na podstawie obserwacji cech pomocniczych. Podobnie jak w uprzednim punkcie, do tego celu wykorzystano formalne metody grupowania zbiorów. Analizowano problem estymacji przeciętnej cechy badanej na podstawie próby dwufazowej, przy czym w pierwszej fazie próbę prostą racjonalnie dzielono na równoliczne grupy na podstawie obserwowanej w niej cech pomocniczych. Potem w drugiej fazie spośród tych grup losowano już próbę, w której obserwowano zmienną badaną.

Rozdział piąty dotyczy estymacji wektora średnich w populacji na podstawie wektora średnich z próby dwustopniowej. Oprócz podstawowych własności rozkładu wektora tych estymatorów zaprezentowano rozwiązania zadań optymalizacji rozmiarów prób składowych próby dwustopniowej.

Szósty rozdział jest poświęcony wektorowym estymatorom różnicowym i regresyjnym. Prezentowane są ich macierze wariancji i kowariancji w przypadkach, gdy losowana jest próba prosta bądź podwójna. Gdy estymatory te są wyznaczane na podstawie obserwacji cech badanych i pomocniczych w próbie podwójnej, formułowano i rozwiązywano zadania optymalne-

go ustalania liczebności obu prób składowych próby podwójnej. Ponadto przedstawiono własności trzech strategii będących uogólnieniami regresyjnych strategii Singha i Srivastawy. Strategie te uwzględniają więcej niż jedną zmienną dodatkową oraz jedna z nich zależy od dwustopniowego planu losowania próby.

Rezultaty otrzymane w pracy powinny przyczynić się do racjonalizacji badań reprezentacyjnych populacji. Prezentowane własności estymatorów wektorowych i planów losowania prób winny ułatwić wybór najodpowiedniejszych spośród nich w praktyce badań statystycznych. Celem optymalnego tworzenia warstw bądź grup w populacji jest w konsekwencji zwiększenie dokładności oceny parametrów na podstawie prób warstwowych bądź grupowych. Zastosowanie tych procedur przyczyni się również do oszczędności nakładów przeznaczanych na badania reprezentacyjne. W tej kwestii zwłaszcza mają bezpośrednie znaczenie analizowane w pracy zagadnienia optymalnego ustalania liczebności prób złożonych z uwzględnieniem kosztów badania i żądanej dokładności ocen parametrów.

INDEX OF EXPRESSIONS

$\bar{\mathbf{y}} = [\bar{y}_1 ... \bar{y}_m]^T$, $\bar{\mathbf{y}} = N^{-1}\mathbf{Y}^T\mathbf{J}_N$: the vector of mean values from a population;

$\tilde{\mathbf{y}} = N\bar{\mathbf{y}}$: the vector of global values from a population;

$\mathbf{C}_{*yy} = \mathbf{C}_*(y)=[c_{*tj}]$ (t,j=1,...,m), where:

$c_{*tj}=c_*(y_t, y_j) = \frac{1}{N-1}\sum_{i=1}^{N}(y_{it} - \bar{y}_t)(y_{ij} - \bar{y}_j)$: the matrix of variances and covariances of variables

$v_*(y_j)=c_*(y_j, y_j)$: the variance of a variable y_j ;

$\mathbf{R} = \mathbf{R}(y) = [r_{tj}]$ (t,j=1,...,m), where $r_{tj} = \frac{c(y_t, y_j)}{\sqrt{v_*(y_t)v_*(y_j)}}$: the correlation matrix;

$q_*(y) = \sqrt{\mathrm{tr}\mathbf{C}_*(y)}$: the mean radius of an m-dimensional;

$g(y) = \det(\mathbf{C}_*(y))$: the generalized variance of an m-dimensional variable;

$\rho(y) = \sqrt{\lambda_1}$: the spectral radius is equal to the square root of the maximal eigenvalue λ_1 of the covariance matrix $\mathbf{C}_*(y)$,

$P(\underline{s})$: sampling design of an ordered sample $\underline{s}$,

$P(s)$: sampling design of an unordered sample s,

π_k: inclusion probability of order one,

π_{kl}: inclusion probability of order two,

$P_1(\underline{s})$:the sampling design of the simple and ordered sample, see the expression (1.26),

$P_2(\underline{s})$: the sampling design of the simple and ordered sample with a fixed effective size, see (1.27),

$P_3(s)$: the sampling design of the simple and unordered sample with a fixed effective size, see (1.28),

E(.), D^2(.) and Cov(.): expected value, variance and covariance determined on the basis of a sampling design,

E(.), D^2(.) and *Cov*(.): expected value, variance and covariance evaluated on the basis of superpopulation model,

$\mathbf{t}_S$: an estimator of a parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a parameter space,

$\mathbf{t}_s$: a value of the estimator $\mathbf{t}_S$,,

($\mathbf{t}_S$, P(s)): the sampling strategy,

$(\bar{\mathbf{y}}_S, P_3(s))$: the vector of simple sample means where

$$\bar{\mathbf{y}}_S = [\bar{y}_{S1} ... \bar{y}_{Sm}\ ]\ ,\ \bar{y}_{iS} = \frac{1}{n}\sum_{k\in S} y_{ki}\ ,\ i=1,...,m,$$

$\mathbf{V}(\mathbf{t}_S)=E(\mathbf{t}_S-E(\mathbf{t}_S))^T(t_S-E(\mathbf{t}_S))$: the covariance matrix of an estimator $\mathbf{t}_S$,

$\mathbf{V}_{SR}(\mathbf{t}_S)= E(\mathbf{t}_S-\boldsymbol{\theta})^T(t_S-\boldsymbol{\theta})$: the matrix of the second mixed moments of estimation errors,

$q^2_{SR}(\mathbf{t}_S) = tr\mathbf{V}_{SR}(\mathbf{t}_S)$, the mean square radius,

$q(\mathbf{t}_S)= \sqrt{tr\mathbf{V}(\mathbf{t}_S)}$: the mean radius of the estimator $\mathbf{t}_S$,

$g_{SR}(\mathbf{t}_S) = det\mathbf{V}_{SR}(\mathbf{t}_S)$: the generalized mean square error,

$g(\mathbf{t}_S) = det\mathbf{V}(\mathbf{t}_S)$: the generalized variance of the estimator $\mathbf{t}_S$,

$\rho_{SR}(\mathbf{t}_S)$: spectral radius of the matrix: $\mathbf{V}_{SR}(\mathbf{t}_S)$,

$\rho(\mathbf{t}_S)$= spectral radius of the covariance matrix: $\mathbf{V}(\mathbf{t}_S)$,

$$deff(t_S, P(s)) = \frac{v_{SR}(t_S, P(s))}{D^2(\bar{y}_S, P_1(s))}$$: the deff coefficient,

$deff(\mathbf{t}_S, P(s)) = \rho\left(\mathbf{V}_{SR}(\mathbf{t}_S, P(s))\mathbf{V}^{-1}(\bar{\mathbf{y}}_S, P_1(s))\right)$: generalized of the deff coefficient into the multivariate case,

$\boldsymbol{V}(\mathbf{T}_s)=E(\mathbf{T}_s-E(\mathbf{T}_s)]^T[(\mathbf{T}_s-E(\mathbf{T}_s)]$: a ξ-covariance matrix of prediction errors,

$q_{SR}(\mathbf{T}_S) = \sqrt{E\mathbf{V}_{SR}(\mathbf{T}_S)}$: the mean square radius of a strategy $(\mathbf{T}_S,p)$,

$E[\mathbf{V}_{SR}(\mathbf{T}_S)] = EE(\mathbf{T}_S - \boldsymbol{\Theta})^T(\mathbf{T}_S - \boldsymbol{\Theta})$: the matrix of mixed second moments of prediction errors,

$$t_{HTS} = \frac{1}{N}\sum_{k=1}^{N}\frac{a_k y_k}{\pi_k}$$: the Horvitz and Thompson estimator,

$P_4(s)$: Lahiri 's sampling design, see expression (2.24),

$P_5(s)$: the sampling design proportional to total of values of auxiliary variable which are not observed in sample, see (2.38),

$P_6(s)$: Singh and Srivastava's sampling design, see (2.43),

$P_7(s)$: the sampling design proportional to function of sample variance of an auxiliary variable, see (2.49),

$P_8(s)$: the sampling design proportional to squared estimation error of auxiliary variable mean, see (2.59),

$P_9(s)$: the sampling design proportionate to decreasing function of squared estimation error of auxiliary variable mean, see (2.64),

$P_{10}(s),...,P_{13}(s)$: space sampling designs, see the expressions (2.70), (2.71), (2.72), (2.75),

$P_{14}(s),...,P_{18}(s)$: Sampling designs dependent on the determinant of sample covariances matrix, see the expressions (2.78), (2.81), (2.82), (2.85), (2.86),

$P_{19}(s)$: the two stage-Singh-Srivastava's sampling design, see the expression (6.86),

$P_w(s)$: the sampling design of the stratified sample, see the expressions (3.1), (3.2),

P'_w : the sampling design of the stratified sample in the case of proportional allocation of the samples in strata,

$(\bar{\mathbf{y}}_{gS}, P_g)$: the cluster sample strategy, see the chapter 4.1,

$\bar{\mathbf{y}}_{wS} = \sum_{h=1}^{H} w_h \bar{\mathbf{y}}_{hS}$: the vector of stratified sample means,

the regression model, see the expressions: (1.19)-(1.21), (3.64), (3.65),

$\mathbf{R}^{(w)} = \left[r_{ij}^{(w)}\right]$: the matrix of the coefficients of within-cluster correlation, see the expressions (4.5), (4.6), (4.12),

$\mathbf{C}_m = \left[c_m\left(y_i, y_j\right)\right]$: the between-cluster matrix of the covariances, see expression (4.23),

$\mathbf{C}_w = \left[c_w\left(y_i, y_j\right)\right]$: the within-cluster matrix of the covariances,

$\boldsymbol{\Delta}$: the homogeneity coefficient of multidimensional variable, see expression (4.27),

$\left(\tilde{\mathbf{y}}_{gS}, P_d\right)$: the two stage sampling strategy, see the chapter 5.1,

$(\mathbf{t}_{BS}, P_3)$: the simple regression strategy, see the chapter 6.1,

$\hat{r}^2$: the mean determination coefficient, see the expression (6.10),

$\left(\bar{\mathbf{t}}_{BS}, P_{3d}\right)$: the vector of regression estimators from double sample, see the chapter 6.2,

$\left(\bar{y}_{RS}, P_{16}(s)\right)$: the first generalised Singh-Srrivastava's regression strategy, see the expressions (6.66), (6.71)-(6.74),

$\left(\bar{y}_{RS}, P_{16}(s)\right)$: the second generalised Singh-Srrivastava's regression strategy, see the expressions (6.80), (6.83), (6.84),

$(\bar{t}_{RQ}, P_{19}(s))$: the two stage-Singh-Srivastava's sampling strategy, see the expression (6.86)-(6.97).

REFERENCES

Anderson T.W. (1958): *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York.

Anderson D.W., Kish L., Cornell R.G. (1980): On stratification, grouping and matching. *Scandinavian Journal of Statistics*, vol. 7, pp. 61 – 66.

Arwanitis L.G., Afonia B. (1971): Use of generalized variance and the gradient projection method in multivariate stratified sampling. *Biometrics*, vol. 27, pp. 119-127.

Beardwood J., Halton J.H., Hammersley J.M. (1959): The shortest path through many points. *Proceedings of Cambridge Phil. Soc*., vol. 55.

Bethel J. (1989): Sample allocation in multivariate surveys. *Survey Methodology*, vol. 15, nr 1, pp. 47-57.

Bethleham, J.G. (1988): Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, 4, pp.251 – 260.

Blythe J.R.H. (1945): The economics of sample size applied to the scaling of sowlongs. *Biometrics Bulletin*, vol. 1, s. 67-70.

Borovkov A.A. (1984): *Mathematical statistics. Estimation. Testing Hypothesis* (in Russian). Nauka, Moscow.

Borsuk K. (1976): *Multidimensional Analytical Geometry*. PWN, Warszawa.

Bracha Cz. (1978): Estimation of linear regression parameters on the basis of sample drawn without replacement from finite population (in Polish). *Przegląd Statystyczny*, vol. 25.

Bracha Cz. (1982): Estimation of linear regression parameters on the basis of two-stage sample (in Polish). *Przegląd Statystyczny*, vol. 29.

Bracha Cz. (1983): Linear Regression in Survey Sampling (in Polish). *Prace i Materiały Instytutu Cybernetyki i Zarządzania. SGPiS*, Warszawa.

Bracha Cz. (1987): Using Auxiliary Information in Survey Sampling. (in Polish). ZBS-GUS i PAN, Warszawa.

Bracha Cz. (1991): Selected problems of stratifying sampling. (in Polish) *Prace i Materiały Instytutu Cybernetyki i Zarządzania*. Tom 20. SGPiS, Warszawa.

Bracha Cz. (1996): *Teoretyczne podstawy metody reprezentacyjnej*. PWN, Warszawa.

Cassel C.M., Sarndal C.E., Wretman J.H. (1977): *Foundation of Inference in Survey Sampling*. John Wiley & Sons, New York-London-Sydney--Toronto.

Chatterjee S. (1968): Multivariate stratified surveys. *Journal of the American Statistical Association*, vol. 63.

Cochran W.G. (1961): Comparison of methods for determining stratum boundaries. *Bulletin of International Statistical Institute*, vol. 38, No. 2, pp. 345-358.

Cochran W.G. (1963): *Sampling Techniques*. John Wiley & Sons, New York.

Cramer H. (1958): *Mathematical Methods in Statistic*. (in Polish) PWN, Warszawa.

Czerniak W. (1971): On independend sampling with replacement (in Polish). *Biblioteka Wiadomości Statystycznych*. Tom 15, pp. 40-86.

Dalenius T. (1950): The problem of optimum stratification. *Journal of the American Statistical Association*, vol. 54, pp. 88 – 101.

Dalenius T. (1953): The multivariate sampling problem. *Scandinavisk Aktuarietidskrift*, vol. 36, pp. 92-102.

Dalenius T. (1957): *Sampling in Sweden. Contribution to Methods and Theories of Sample Survey Practice.* Almqvist & Wiksells, Stockholm.

Dalenius T., Gurney M. (1951): The problem of optimum stratification II. *Scandinavisk Aktuarietidskrift*, vol. 34, pp. 133-148.

Dalenius T., Hodges J.L.Jr. (1959): Minimum variance stratification. *Journal of the American Statistical Association*, vol. 54, pp. 88-101.

Dayal S. (1985): Allocation of sample using values of auxiliary characteristic. *Journal of Statistical Planning and Inference*, vol. 11, pp. 321-328.

Demidowicz B.P., Maron I.A. (1965): *Numerical Methods.* Vol. I. (in Polish) PWN, Warszawa.

Deville C.M., Särndal C.E., (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, vol. 87, nr 418, pp. 376-382.

Donoho D.L., Gasko M. (1992): Breakdown Properties of Location Estimate Based on Halspace Depth and Projected Outlyingness. *Annals of Statistics* 20, pp.1803 – 1827.

Efron B., Tibshirani R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York, London.

Everitt B.S. (1998): *The Cambridge Dictionary of Statistics.* Cambridge University Press, Cambridge.

Fisz M. (1963): Probability Theory and Mathematical Staitrstics. Wiley, New York.

Fisz M. (1967): *Introduction to Probability and Mathematical Statistics* (in Polish). PWN, Warszawa.

Flachsmeyer J. (1977*): Combinatorics* (in Polish). PWN, Warszawa.

Gamrot W., Wywiał J. (2002). Comparison of the accuracy of some two-stage sampling schemes by means of computer simulation. *Zeszyty Naukowe*, no. 21, Akademia Ekonomiczna w Katowicach, p. 11-28.

Geary R.C. (1949): Sampling methods applied to Irish agricultural statistics. *Technical Series*.

Ghosh J.K. (1963): A game theory approach to the problem of optimum allocation in stratified sampling with multiple characters. *Calcutta Statistical Association Bulletin*, vol. 12, pp. 4-12.

Ghosh S.P. (1958): A note on stratified random sampling with multiple characters. *Calcutta Statistical Association Bulletin*, vol. 8, pp. 81-90.

Godambe V.P., Joshi V.M. (1965): Admissibility and Bayes estimation in sampling finite populations I. *Annals of Mathematical Statistics*, vol. 36, pp. 1707-1722.

Greń J. (1963): Localization of sample in the case of multiparameter stratified sampling (in Polish). *Przegląd Statystyczny*, vol. 10, pp. 291-302.

Greń J. (1964): On some methods of localization of sample in the case of multiparameter stratified sampling(in Polish). *Przegląd Statystyczny*, vol. 11, pp. 361-369.

Greń J. (1966): On application of non-linear programming to survey sampling (in Polish). *Przegląd Statystyczny*, vol. 13, pp. 203-217.

Greń J. (1969): Multidimensional regression estimator of mean values (in Polish). *Przegląd Statystyczny*, vol. 16.

Greń J. (1970): Multidimensional regression estimator of mean values in finite population (in Polish). *Przegląd Statystyczny*, vol. 17, pp. 73-78.

Greń J., Koźniewska I. (1964): Solution to some recurrence equation connected with two-parameter stratified sampling (in Polish). *Przegląd Statystyczny*, vol. 11, pp. 169-176.

Hartigan J.A. (1975). *Clustering Algorithms*. New York. J. Wiley.

Hartley H.O. (1965): Multiple purpose optimum allocation in stratifed sampling. *Proceedings of the American Statistical Association, Social Statistics Section*, pp. 258-261.

Hartley H.O., Rao J.N.K. (1969): A new estimation theory for sample surveys II. In: *New Developments in Survey Sampling*. Edited by N.L. Johnson and H. Smith. Wiley-Interscience, New York, pp. 147-169.

Hastings C. (1955). *Approximation for digital computations. Princenton University Press, Princenton.*

Herzel A. (1986): Sampling without replacement with unequal probabilities: sample designs with preassigned joint inclusion probabilities of any order. *Metron*, vol. 44, no. 1-4, pp. 49-68.

Hess I., Sethi V.K., Balakrishnan T.R. (1966): Stratification - a practical investigation. *Journal of the American Statistical Association*, vol. 61.

Horvitz D.G., Thompson D.J. (1952): A generalization of sampling without replacement from finite universe. *Journal of the American Statistical Association*, vol. 47, pp. 663-685.

Huddleston H.F., Claypool P.L., Hocking R.R. (1970): Optimal sample allocation to strata using convex programming. *Applied Statistics*, vol. 19, pp. 273-278.

Hughes E., Rao J.N.K. (1979): Some problems of optimal allocation in sample surveys involving inequality constrains. *Communication in Statistics*, vol. A8, pp. 1551-1571.

Huisman M. (2000): Post-stratification to correct for nonresponse: classification of ZIP code areas. Compstat 2000. *Proceedings in Computational Statistics, 14th Symposium held in Utrecht, The Netherlands. Edited by J.G. Bethleham and P.G.M. van der Heijden*, pp. 325 – 330.

Jaganathan R. (1965): The programming approach in multiple characters studies. *Econometrica*, vol. 33, pp. 236-237.

Jaganathan R. (1965a): A method for solving a nonlinear programming problem in sample surveys. "Econometrica", vol. 33, s. 841-846.

John S. (1969): On multivariate ratio and product estimators. *Biometrika*, vol. 56, pp. 533-537.

Jonin B.G., Jonina N.P., Zhuravlev N.M. (1978): Ispolsovanye protsedur optimisatsyi pri klassifikatsy obyektov i faktorov. *In: Ekonomika i statisticheskoye modyeli w prognozirovyi planirowanyi promyshlennogo proisvodstva*. Nauka, Moskwa.

Joshi V.M. (1965): Admissibility and Bayes estimation in sampling finite populations II i III. *Annals of Mathematical Statistics*, vol. 36, pp. 1658-1670.

Joshi V.M. (1966): Admissibility and Bayes estimation in sampling finite populations IV. *Annals of Mathematical Statistics*, vol. 37, pp. 1658-1670.

Kish L. (1961): Efficient allocation of multipurpose sample. *Econometrica*, vol. 29, pp. 363-385.

Kish L. (1965): *Survey sampling*. John Wiley & Sons, Inc. New York-London-Sydney.

Kokan A.R. (1963): Optimum allocation in multivariate surveys. *Journal of the Royal Statistical Society*, vol. A 126, pp. 557-565.

Kokan A.R., Khan S. (1967): Optimum allocation in multivariate surveys: an analytical solution. *Journal of the Royal Statistical Society*, vol. B 29, pp. 115-125.

Konijn H.S. (1962): Regression analysis in sample surveys. *Journal of the American Statistical Association*, vol. 57.

Konijn H.S. (1973): *Statistical theory of sample survey and analysis*. North-Holland Publishing Company, Inc., Amsterdam-London, American Elsevier Publishing Company, Inc., New York.

Krzyśko M. (2000). *Multidimensional Statistical Analysis* (in Polish). Uniwersytet im Adama Mickiewicza, Poznań.

Kubik L.T., Krupowicz A. (1982): *Introduction to Probability and Its Applications* (in Polish). PWN, Warszawa.

Lahiri G.W. (1951): A method for sample selection providing unbiased ratio estimator. *Bulletin of the International Statistical Institute*, vol. 33, pp. 133-140.

Lehman E.L. (1991): *Theory of Estimation* (in Polish). PWN, Warszawa.

Lipski W., Marek W. (1986): *Combinatoric Analysis* (in Polish). PWN, Warszawa.

Liu R. (1990): On a Notation of Data Depth Based Random Simplices. *Annals of Statistics* 18, pp. 405 – 414.

Lynch G.W. (1978): The choice of auxiliary variables in multivariate ratio and regression estimators. *Proceedings of the Section on Survey Research Methods*. American Statistical Association.

Mahalanobis P.C. (1944): On large scale sample surveys. *Phil. Trans. Roy. Soc*. London, vol. B 231, pp. 329-451.

Melaku A. S. (1987): L d-norm and other methods for sample allocation in multivariate stratified surveys. *Computational Statistics & Data Analysis*, vol. 5, pp. 415-423.

Mikhail N.N., Mir. M.A. (1981): Unbiased estimates of the generalized variance for finite population. *Journal of the Indian Statistical Association*, vol. 19, pp. 85-92.

Miller R., G. (1981). *Simultaneous Statistical Inference.* Springer Verlag, New York- Berlin- Heildelberg-London-Paris-Tokyo-Hong Kong--Barcelona.

Mukerjee R., Rao T.J. (1985): On a problem of allocation of sample size in stratified random sampling. *Biometrical Journal*, vol. 27, nr 3, pp. 327-331.

Murthy M.N. (1977): *Sampling theory and practice*. Statistical Publishing Society, Calcutta.

Neyman J. (1934): On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, vol. 97, pp. 558-606.

Olkin I. (1958): Multivariate ratio estimation for finite populations. *Biometrika*, vol 45, pp. 154-165.

Ralston A. (1975): *Introduction to Numerical Methods* (in Polish). PWN, Warszawa.

Ramakrishnan M.K. (1975): Choice of an optimum sampling strategy I. *Annals of Statistics*, vol. 3, pp. 669-679.

Rao C.R. (1973). *Linear Statistical Inference and Its Applications.* John Wiley & Sons, Inc., New York – London – Sydney – Toronto.

Rao C.R. (1982): *Linear Models of Mathematical Statistics* (in Polish). PWN, Warszawa.

Rao J.N.K. (1973): On double sampling for stratification and analytical surveys. *Biometrika*, vol. 60, nr 1, pp. 125-133.

Rao J.N.K. (1985): Conditional inference in survey sampling. *Survey Methodology*, Vol. 11, No. 1, pp. 15 – 31.

Rao J.N.K., A.J. Scott (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association*,
vol. 76, no. 374, pp. 221-230.

Rao T.V.H. (1962): An existence theorem in sampling theory. *Sankhya*, vol. A 24, pp. 327-330.

Royall R.M. (1970): On finite population sampling theory under certain linear regression models. *Biometrika*, vol. 57, pp. 377-387.

Sarndal C.E. (1976): On uniformly minimum variance estimation in finite populations. *Annals of Statistics*, vol. 4, pp. 993-997.

Sarndal C.E., B. Swensson, J. Wretman (1992): *Model Assisted Survey Sampling. Springer Verlag, New York-Berlin-Heidelberg- London-Paris--Tokyo-Hong Kong- Barcelona-Budapest.*

Schneeberg H., Pollot J.P. (1985): Optimum stratification with two variables. *Statistical Papers*, vol. 25, pp. 97 – 117.

Sen A.R. (1953): On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, vol 5, pp. 119-127.

Serfling R.J. (1968): Approximately optimum stratification. *Journal of the American Statistical Association*, vol. 63.

Singh P., Srivastava A.K. (1980): Sampling schemes providing unbiased regresion estimators. *Biometrika*, vol. 67, pp. 205-209.

Singh R. (1971): Approximately optimal stratification on the auxiliary variable. *Journal of the American Statistical Association*, vol. 66, pp. 829 – 833.

Skibicki M. (2002). Maximization of measure of allowable sample sizes region in stratified sampling. In: *Classification, Clustering, and Data Analysis*. Eds. K. Jajuga, A. Sokołowski, H.H. Bock. Springer Berlin, Heidelberg, New York, Barcelona, Hong Kong, London, Milan, Paris, Tokyo, 2002, pp. 263-269.

Skibicki M. (2003). Optimatization of sample sizes drawn from strata on the basis of spectral radius of covariance matrix of means estimator vector. *Statistics in Transition.* In printing.

Skibicki M., J.Wywiał. (2001): On using clustering methods to stratification of population. (in Polish). *Wiadomości Statystyczne*, 2001, 8, pp. 4-11.

Skibicki M., Wywiał J. (2002). On optimal sample allocation in strata. W: *Statystyka regionalna w służbie samorządu lokalnego i biznesu* pod. red. Jana Paradysza. *Akademia Ekonomiczna w Poznaniu, Internetowa Oficyna Wydawnicza, Centrum Statystyki Regionalnej*, Poznań.

Srikantan (1963): Problems in optimum allocation. *Operational Research*, vol. 11, pp. 265-273.

Theil H. (1979): Principles of Econometrics (in Polish). PWN, Warszawa.

Thomsen I. (1976): A comparison of approximately optimal stratification given proportional allocation with other method of stratification and allocation. *Metrika*, vol. 23, pp. 15 - 25.

Thompson M.E. (1997). *Theory of Sample Surveys.* Chapman & Hall, London – Weinhein – New York –Tokyo – Melbourne – Madras.

Tillé Y. (1998): Estimation in surveys using conditional inclusion probabilities: simple random sampling. *International Statistical Review*, Vol. 66, No. 3, pp. 303 – 322.

Tillé Y. (1999). Estimation in surveys using conditional inclussion probabilities: Complex Design. *Survey Methodology*, vol. 25, No 1, pp. 57-66.

Tripathi T.P. (1973): Double sampling for indusion probabilities and regression method of estimation. *Jourlnal of the Indian Statistical Association*, vol. 10, pp.33-46.

Tripathi T.P. (1976): On double sampling for multivariate ratio and difference methods at estimation. *Journal of the Indian Society of Agricultural Statistics*. vol. 28, nr 1, pp. 33-54.

Tschuprow A.A. (1923): On the mathematical expectation of the moments of frequency distribution in the case of correlated observations. *Metron*, vol. 2, pp. 461-493 i s. 646-683.

Wagner W., Kobylińska M. (2000): On using depth measure in desribing statistic. (in Polish). In: *Wyzwania i dylematy statystyki XXI wieku pod red. W. Ostasiewicza.* Akademia Ekonomiczna we Wrocławiu.

Ward J.H. (1963): Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, vol. 58.

Williams W.H. (1962): The variance of an estimator with part-stratified weighing. *Journal of the American Statistical Association*, Vol. 57, pp. 622 – 627.

Wilks S.S. (1932): Certain generalization in the analysis of variance. *Biometrika*, vol. 24, pp. 471-494.

Wilks S.S. (1962): *Mathematical statistics*. John Wiley & Sons, Inc., New York-London.

Wywiał J. (1988): Estimation of vector of means on the basis of difference and regression estimators (in Polish). *Przegląd Statystyczny*, vol. 35, pp. 19-35.

Wywiał J. (1988a): Location of sample in strata minimizing spectral radius of variance-covariance matrix (in Polish). *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, nr 404, pp. 195-200.

Wywiał J. (1989): On minimization of generalized variance of vector of means from stratified sample (in Polish). *Wiadomości Statystyczne*, nr 11, pp. 23-24.

Wywiał J. (1990): On optimization of sizes of sample drawn from strata in the case of estimation of vector of mean values in finite population (in Polish). *Zeszyty Naukowe Akademii Ekonomicznej w Katowicach*, nr 117, pp. 19-32.

Wywiał J. (1991): Properties of cluster sample drawn from population clustered on the basis of auxiliary variable. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, nr 559, pp. 107-112.

Wywiał J. (1991a). On sampling design proportionate to sample mean of auxiliary variable (in Polish). *Wiadomości Statystyczne, pp. 21-22.*

Wywiał J. (1991b): On optimal stratification of fixed population on the basis of multidimensional auxiliary variable (in Polish). *Badania Operacyjne i Decyzje* 1991, no 2, pp.63-69.

Wywiał J. (1992). *Survey Sampling in Economic Research (in Polish).* Prace Naukowe Akademii Ekonomicznej w Katowicach.

Wywiał J. (1992a): On some coefficients measuring spread of multidimensional variable and generalization of Friedman-Rubin's clustering method of finite population (in Polish). *Zeszyty Naukowe Akademii Ekonomicznej w Katowicach*, nr 120, 129-149.

Wywiał J. (1993): Prediction of superpopulation average and clustering population on the basis of auxiliary variables. (in Polish). *Przegląd Statystyczny*, Vol. XL, zeszyt 3-4, pp. 303-308.

Wywiał J. (1995). *Multidimensional Aspects in Iurvey Sampling (in Polish).* Ossolineum, Warszawa-Wrocław-Kraków.

Wywiał J. (1995a): *On optimal stratification of population on the basis of auxiliary variables*. Statistics in Transition. Vol. 2, No. 2, pp. 831-837.

Wywiał J. (1996): Unbiased estimators of generalized variance from simple random sample drawn without replacement from finite population. *Journal of the Indian Statistical Association*, vol. 34, pp. 125-129.

Wywiał J. (1996a): Generalization of the Ward method and its application to decomposition of regression relationship. *Proceedings of 15th International Conference on Multivariate Statistical Analysis-MSA'96.* University of Łódź. ,December 4-5, 1996, Łódź, pp. 59-75.

Wywiał J. (1996b): On space sampling. *Statistics in Transition* vol. 2 No. 7, pp. 1185-1191.

Wywiał J. (1996c): On two-phase sampling for stratification. *Statistics in Transition.* Vol. 2 No. 6, pp. 971-977.

Wywiał J. (1997): Decomposition of time series on the basis of modified grouping method of Ward. *Acta Universitatis Lodziensis, Folia Oeconomica* 141, p. 137-148.

Wywial J. (1997a): Sampling Design proportional to the Sample Generalized Variance of Auxiliary Variables. *Proceedings of 16th International Conference on Multivariate Statistical Analysis- MSA'97.* Edited by Cz. Domanski and D. Parys. November 27-29 1997. Department of Statistical Methods, Institute of Econometrics and Statistics, University of Łódź, Polish Statistical Association. November 27-29. 1997r., pp. 129-143

Wywiał J. (1998): On stratification of a Population by means of Cluster Methods. *Statistics in Transition* vol. 3 no. 3, pp. 569-574.

Wywiał J. (1998a): Estimation of Population Average on the basis of Strata Formed by Means of Discrimination Functions. *Statistics in Transition (Journal of the Polish Statistical Association)* vol. 3, No. 5, pp. 903-912.

Wywiał J. (1999): Generalization of Singh and Srivastava's schemes providing unbiased regression estimatiors". *Statistics in Transition* vol. 2, No. 2, pp. 259-281.

Wywiał J. (1999a): "Sampling designs dependent on the sample generalized variance of auxiliary variables". *Journal of the Indian Statistical Association.* Vol. 37, pp. 73-87.

Wywiał J. (2000): "On precision of Horvitz-Thompson strategies. *Statistics in Transition (Journal of the Polish Statistical Association)* vol. 4, nr 5, pp. 779-798.

Wywiał J. (2000a). On optimal stratification of population in the case of estimation of mean vector (in Polish). *Prace Naukowe Akademii Ekonomicznej we Wrocławiu* nr 857, pp. 217-223.

Wywiał J. (2000b). On conditional estimation of population average on the basis of stratified population through clustering the selected sample. *Proceedings in Computational Statistics 14th Symposium held in Utrecht, the Netherlands, 2000. Short Communications and Posters. Edited by W. Jansen and J.G. Bethlehem.* Statistics Netherlands.

Wywiał J. (2001). Stratification of population after sample selection. *Statistics in Transition* vol. 5, nr 2, 2001, pp. 327-348.

Wywiał J. (2001a). Using of auxiliary variables to estimation of population parameters. *Report from grant KBN 1H02B 008 16,* Akademia Ekonomiczna w Katowicach .

Wywiał J. (2001b). On estimation of population mean in the case when non-respondents are present P*race Naukowe Akademii Ekonomicznej we Wrocławiu nr 906, Taksonomia 8,* pp. 13-21.

Wywiał J. (2001c). Estimation of population mean on the basis of non-simple sample when non-response error is present. *Statistics in Transition* vol. 5, no. 3, pp. 443-450.

Wywiał J. (2002): On stratification of population on the basis of auxiliary variable and the selected sample. *Acta Universitatis, Folia Oeconomica* 156, pp. 83-90.

Wywiał J. (2002a): On estimation of population average on the basis of cluster sample. In: *Classification, Clustering, and Data Analysis.* Eds. K. Jajuga, A. Sokołowski, H.H. Bock. Springer Berlin, Heidelberg, New York, Barcelona, Hong Kong, London, Milan, Paris, Tokyo, 2002, pp. 271-277.

Wywiał J. (2002b). On the accuracy of mean estimation on the basis of two-phase sampling for stratification. W: *Statystyka regionalna w służbie samorządu lokalnego i biznesu* pod. red. Jana Paradysza. *Akademia Ekonomiczna w Poznaniu, Internetowa Oficyna Wydawnicza, Centrum Statystyki Regionalnej*, Poznań.

Wywiał J., Kończak G. (1994): On location of sample in strata minimizing spectral radius of variance-covariance matrix. (in Polish). In: Proceedings of the *XI* Conference devoted to. *Professor Zbigniew Pawłowski.* Trzemieśnia 24-26 III. Akademia Ekonomiczna w Krakowie, s. 85-92.

Yates F., Grundy P.M. (1953): *Selection without replacement from within strata with probability proporcional to size. Journal of the Royal Statistical Societ", vol. B15, pp. 235-261.*

Yates F. (1960): *Sampling methods for censuses and surveys.* Griffin & Company Lth., London.

Zasępa R. (1972): *Survey Sampling (in Polish).* PWE, Warszawa.

Zieliński R. (1979). *Generators of random pseudovalues (in Polish*). WNT, Warszawa.

ISBN 83-7246-273-9