© IJCIRAS | ISSN (O) - 2581-5334
September 2020 | Vol. 3 Issue. 4

VARIANCE ESTIMATION OF STRATIFIED SAMPLE MEAN ACCORDING TO APROXIMATED FREQUENCY STRATEGY

Mohammed AlRifai

Statistics Centre Abu Dhabi (SCAD), UAE

Abstract

A stratified sample design plays an important role in obtaining efficient estimators comparing with other sampling designs; it based on dividing the statistical population into independent and homogeneous subpopulations called strata.
The determination of strata boundaries has interested role in obtaining efficient estimators, several strategies of strata formation were constructed, it is proved that the approximated frequency strategy $Cum\, f^{\propto}$ ***at*** $\propto = \frac{1}{2}$ ***is the most efficient method of strata formation, Serfling (1968).***
Several researches has discussed the development of this strategy through using different values of$(\propto)$***, with different methods of sample allocations.***
This article aims to discuss the optimality of the stratified sampling according to approximated frequency strategy based on the study variable Y ***as a variable of strata formation, by constructing formulas for the approximated variance of the stratified sampling mean, for different values of*** $\propto$ ***in the interval (0, 1), using equal allocation, proportionate allocation , and ,Neyman allocation method.***
As a result, the article shows that at any value of $\propto$ ***the value of the variance of stratified sample mean in both of Neyman and equal allocation are same.***
While the variance with Neyman allocation is less than the variance with proportional allocation. Therefore, it is recommended that the frequency strategy with Nyman allocation lead to more efficient estimator for population parameters than the proportional allocation method.

Keyword: Stratified Sampling, Sample Mean, Frequency Strategy method, Sample Allocation,

1. BACKGROUND OF STRATIFIED SAMPLE DESIGHN

Of all the methods of sampling, the stratified sampling design is the most commonly used in surveys. In stratifies sampling, the population on N units is sub-divided into k subpopulations called strata, the ith sub-population having N_i units ($i = 1,2,3, \dots, k$). These sub-populations are non-overlapping so that they comprise the whole populations such that $N_1 + N_2 + N_3 + \cdots + N_k = N$
A sample is drawn from each stratum independently, the sample size within the ith stratum being n_i ($i = 1,2, \dots, k$) such that $n_1 + n_2 + n_3 + \cdots + n_k = n$
The main objective of stratification is to give a better cross-section of the population to gain a higher degree of relative precision. In order to achieve this, the following points should be examined carefully: formation of strata, number of strata, and allocation of sample size within each stratum, Anderson, D.W., Kish, L and Cornell, R.G (1979).
The strata should be non-overlapping and should together comprise the whole population. In addition, it should be done in such a way that strata are homogeneous within themselves, with respect to the characteristics under study. In many practical situations when it is difficult to stratify with respect to the characteristics under study, administrative convenience considered as the basis for stratification.

1.1.Estimation of population Mean and variance

Let i denote for the stratum and j for the sampling unit within the stratum. The following symbols refer to stratum i , Daroga Singh (1986).
N_i = total number of units in the population
n_i = total number of units in the sample
$W_i = \frac{N_i}{N}$ Stratum weight

$f_i = \frac{n_i}{N_i}$ Stratum weight

Let y_{ij} be the value of the jth unit in the stratum i, then:
$\hat{\bar{Y}}_i = \sum_j^{N_i} \frac{y_{ij}}{N_i}$, is the stratum mean
$\bar{y}_i = \sum_j^{n_i} \frac{y_{ij}}{n_i}$, is the stratum sample mean
$S_i^2 = \sum_j^{N_i} \frac{(y_{ij} - \bar{y}_i)^2}{N_i - 1}$, is the stratum variance
$V(\bar{y}_i) = \frac{S_i^2}{n_i}(1 - f_i)$ (1.1)
Suppose that the population of N units divided into K strata. The population mean per unit is as follows:
$\bar{Y} = \sum_i^k \sum_j^{N_i} \frac{y_{ij}}{N} = \sum_i^k \frac{N_i \bar{Y}_i}{N} = \sum_i^k W_i \bar{Y}_i$
An estimator $\bar{y}_{st}$ for the population mean $\bar{Y}$ is:

$$\bar{y}_{st} = \sum_i^k \frac{N_i \bar{y}_i}{N}$$

To obtain the sampling variance, we note that sampling is done independently in each stratum and, therefore
$V(\bar{y}_{st}) = V\left(\sum_i^k \frac{N_i \bar{y}_i}{N}\right) = \sum_i^k W_i^2 V(\bar{y}_i)$ (1.2)
Based on both equation (1.1) and equation (1.2), the variance of stratified sample mean formulated as follow
$V(\bar{y}_{st}) = \frac{\sum_i^k N_i(N_i - n_i)S_i^2}{N^2 n_i} = \frac{\sum_i^k (1 - f_i) W_i^2 S_i^2}{n_i}$ (1.3)

1.2. Methods of sample Allocation

In stratified sampling, the allocation of the overall sample to different strata done by the consideration of three main factors, the total number of population units in the stratum (stratum size), the variability within the stratum, and the cost in taking observation per sampling in the stratum. A good allocation is one where maximum precision obtained with minimum resources. There are different methods of allocation of sample size to different strata in stratified sampling procedures, these are : Kish, L., (1995):

Equal allocation:

This is a situation of considerable practical interest for reasons of administrative or fieldwork convenience. In this methods the total sample n divided equally among all the strata, i.e. for the i^{th} stratum

$$n_i = \frac{n}{k}$$

By substituting this value of n_i in equation (1.3), the variance of the sample mean will be:

$V(\bar{y}_{st})_{eq} = \frac{k}{n} \sum_i^k (1 - f_i) W_i^2 S_i^2$ (1.4)

Proportional Allocation

This procedure of allocation is very common in practice because of its simplicity. When other information except N_i , the total number of units in the stratum is available, the allocation of given sample of size n to different strata is done in proportional to their sizes, i.e. in the ith stratum,
$n_i = \frac{nN_i}{N}$, or $(f_i = f)$
This mean that the sampling fraction is the same for all strata. It gives self-weighting sample by which numerous estimates can made with greater speed and high degree of precision.
By substituting this value of n_i above in equation (1.3), the variance of the sample mean will be

$V(\bar{y}_{st})_{prop} = \frac{(1-f)}{n} \sum_i^k W_i S_i^2$ (1.5)

Nyman Allocation

This allocation of the total sample size to strata called minimum- variance allocation. The allocation of the samples among different strata based on a joint consideration of the stratum size and the stratum variation. In this allocation, it assumed that the sampling cost per unit among different strata is the same and the size of the sample is fixed. Sample sizes are allocated by

$$n_i = n \frac{W_i S_i}{\sum_i^k W_i S_i} = n \frac{N_i S_i}{\sum_i^k N_i S_i}$$

The variance of the sample mean in this kind of allocation is as follow:
$V(\bar{y}_{st})_{Ney} = \frac{1}{n}\left(\sum_i^k W_h S_h\right)^2$ (1.6)

2.LITERATURE REVIEW

The strata formation subject has been discussed by different researchers through finding approximated strategies serving the design of stratified sample, The most concentration of researches were on the optimum strata boundaries which lead to the minimum values of the variance of stratified sample mean, assuming that the sample size number of strata, and sample allocation method are identified.

The optimality of the sample size and the number of strata discussed by limited number of researchers Serfling (1968) and Singh. R (1971)

Dalenius & Hodge (1959) , and Dalenius and Gurney (1951) discussed the segmentation of the population into strata through determining strata boundaries leads to minimum values of the variance of stratified sample mean , by dividing the value of $Cum\sqrt{f_x}$, into equal parts assuming known number of strata and known sample allocation method , and the study variable has uniform probability density function.

Sethi (1963) , Introduced approximated solutions to the strata boundaries when the study variables have normal distribution , and Chi square distribution , he considered these two probability distributions are standard and the distribution of any study variable can be approximated by one of these two distributions.

Ghosh (1963), proposed a strategy for strata construction based on two stratification variables, he considered the measure of the stratification efficiency is the overall variance to the sample means, and the optimum stratification is the one, which produce minimum values of the overall variance.

Raj, D. (1964) discussed the strata formation based on equal allocation, through theorem based on four different probability distributions; he concludes that when the number of strata is high the efficiency of the estimators based on this theorem become low.

Hess, Irene (1966), introduce practical study to strata boundaries with implementation on data related to the health services in the skewed populations, he studied the selection of the stratification variable, allocation method, strata formation, and number of strata, through comparative study between four strategies, $Cum\sqrt{f_x}$ method, Sethi method, Ekman method, and Hurwitz & Madow method.

3.STRATA BOUNDRIES BY OPTIMUM STRATIFICATION STRATEGY METHOD

The optimum stratification strategy mainly aimed to determine the optimum strata boundaries (strata formation) which lead to minimum variances within strata. approximated stratification strategy $Cum(f^{\propto})$ is the most common methods that have been used by different researchers, , where f is the density function of the study variable y ,or the auxiliary stratification variable x which has high correlation with y , assuming sample size, number of strata, and sample allocation method, are pre-determined.

Dalenus &Hug 1959 proved that the approximated strategy $Cumf^{\propto}$, at $= \frac{1}{2}$, is the most efficient method

Suppose that $f(y)$ is the probability density function of the study variable Y in the domain $a \leq y \leq b$, and the values $y_1, y_2, y_3, \dots, y_{k-1}$ are the strata boundaries in the domain [a, b], then for stratum i :

$$W_i = \int_{y_{i-1}}^{y_i} f(y)dy$$

$$\mu_i = \frac{1}{W_i} \int_{y_{i-1}}^{y_i} yf(y)dy$$

$$S_i{}^2 = \frac{1}{W_i} \int_{y_{i-1}}^{y_i} y^2 f(y)dy - \mu_h{}^2$$

At $\alpha = \frac{1}{2}$, suppose that $A = \int_a^b \sqrt{f(y} \, dy$ is constant value in the domain of the density function$f$, and $A_i = \int_{y_{i-1}}^{y_i} \sqrt{f(y} \, dy$, then $A = \sum_{i=1}^{k} A_i$.

Assume that the values of $f(y)$ in the stratum i , $f(y_i)$ is defined approximately by the mean value in the stratum,

$$W_i = f(y_i)(y_i - y_{i-1})$$

$$\mu_i = \frac{(y_i + y_{i-1})}{2},$$

$$S_i^2 = \frac{(y_i - y_{i-1})^2}{12}$$

$$A_i = \sqrt{f(y_i)}(y_i - y_{i-1})$$

From formula (1.6), the variance of stratified sample mean with Nyman allocation method is given by:

$$V(\bar{y}_{st})_{Ney} = \frac{1}{n}\left(\sum_i^k W_i S_i\right)^2 = \frac{\left(\sum_{i=1}^{k} A_i^2\right)^2}{12n} \qquad (2.1)$$

Minimizing the value of the variance in equation (2.1) is equivalent to minimizing the value of $\sum_{i=1}^{k} A_i^2$, since $\sum_{i=1}^{k} A_i = A$, and A is fixed value, $\sum_{i=1}^{k} A_i^2$ has minimum value, when values of A_i are equal in all strata;

$$A_i = \frac{A}{k}; i = 1,2,3, \dots, k$$

In other word, the optimum strata boundaries with Nyman allocation method attained through dividing the domain of $Cum\sqrt{f(y)}$ into k strata, then the stratum boundary is given by;

$$y_h = \frac{h\sum_{i=1}^{k}\sqrt{f(y}}{k} \quad ; i = 1,2,3, \dots, k$$

Approximated Minimum Variance for Stratified Sample Mean Based On Frequency Strategy Method

Several researchers, Serfling (1968), Sethi (1963), Sing R & Sukhatme (1969), Al- Kassab &Al- Tauy (1994), and others developed this strategy through using different values of α, with different allocation methods. In this article, we shall construct general formula for the variance of stratified sample mean at any value ofα, for each method of sample allocation.

3.1. Minimum Variance Estimation In Equal Allocation method

Assume that $y_1, y_2, y_2, \dots, y_N$, are independent population values of the random variable, with probability density function$f(y)$, and;

$\varphi(y) = \int_{-\infty}^{\infty} f^{\alpha}(y)dy$ (3.1)

With a concentration in the interval $[a, b]$ where $f(y)$ out of this interval equal to zero with neglicable error, and the strata boundaries are $y_0 < y_1 < y_2, \dots < y_{k-1}$, we have;

$\varphi_i(y) = \int_{y_{i-1}}^{y_i} f^{\alpha}(y)dy$ (3.2)

Assume that the value of $f(y)$ in the interval i approximated by its mean value Serfling (1968), then

$W_i = f(y_i)(y_i - y_{i-1})$ (3.3)

$\mu_i = \frac{(y_i + y_{i-1})}{2}$ (3.4)

$S^2{}_i = \frac{(y_i - y_{i-1})^2}{12}$ (3.5) $S_i^2 = \frac{(y_i - y_{i-1})^2}{12}$

The value of $\varphi_i(y)$ in equation (3.2) will be;

$\varphi_i(y) = \int_{y_{i-1}}^{y_i} \mu_h^{\alpha}\, dy = \mu_h^{\alpha}(y_i - y_{i-1})$ (3.6)

The variance of the stratified sample mean, with equal allocation method (ignoring the sample fraction$f$) is given by; 1977 Cochran,

$V(\bar{y}_{st})_{eq} = \frac{k}{n} \sum_{i=1}^{k} W_i^2 S_i^2$ (3.7)

By substituting the equations (3.3) and (3.5) in (3.7), we get;

$V(\bar{y}_{st})_{eq} = \frac{k}{12n}\, \mu_i^2 (y_i - y_{i-1})^4$ (3.8)

This equation written in terms of $\varphi_i(y)$, as follows;

$V(\bar{y}_{st})_{eq} = \frac{k}{12n} \sum_{i=1}^{k} \varphi_i^{\frac{2}{\alpha}}(y)\, (y_i - y_{i-1})^{\left(4-\frac{2}{\alpha}\right)}$ (3.9)

The above equation has its minimum value, when $\varphi(y)$ value is constant for all strata, under Lagrange multiplier restriction $\varphi(y) = \sum_{i=1}^{k} \varphi_i(y)$, i.e. $\varphi_i(y) = \frac{\varphi(y)}{k}$

Then the minimum variance of stratified sample mean according to optimum stratification strategy with equal allocation method will become

$V(\bar{y}_{st})_{eq} = \frac{\varphi^{\frac{2}{\alpha}}(y)}{12nk^{\frac{2}{\alpha}-2}}$; $0 < \alpha < 1$ (3.10)

3.2. Minimum Variance Estimation In Proportional Allocation method

The variance of the stratified sample mean with proportional allocation method (ignoring the sample fraction$f$) is given by; Cochran (1977)

$V(\bar{y}_{st})_{prop} = \frac{1}{n} \sum_{i=1}^{k} W_i S_i^2$ (3.11)

By substituting the value of W_i , S_i^2, $\varphi_i(y)$, from equations (3.2), (3.3), (3.5) respectively , in equation (3.11), we get the variance of stratified sample mean in terms of $\varphi(y)$ as follows

$V(\bar{y}_{st})_{prop} = \frac{1}{12n} \sum_{i=1}^{k} \varphi_i^{\frac{1}{\alpha}}(y)\, (y_i - y_{i-1})^{\left(3-\frac{1}{\alpha}\right)}$ (3.12)

The above equation has its minimum value, when $\varphi(y)$ value is equal for all strata, under Lagrange multiplier restriction $\varphi(y) = \sum_{i=1}^{k} \varphi_i(y)$, i.e $\varphi_i(y) = \frac{\varphi(y)}{k}$

By substituting the value of $\varphi_i(y)$, in equation (3.12) above, we get the minimum variance of stratified sample mean in proportional allocation as follows

$V(\bar{y}_{st})_{prop} = \frac{\varphi^{\frac{1}{\alpha}}(y)}{12nk^{\frac{1}{\alpha}-1}}$ (3.13)

3.3. Minimum Variance Estimation In Neyman Allocation method

The variance of the stratified sample mean with proportional allocation method (ignoring the sample fraction (f); is as follows;

$V(\bar{y}_{st})_{Ney} = \frac{1}{n}\left(\sum_i^k W_i\, S_i\right)^2$ (3.14)

By substituting the value of W_i , S_i^2, $\varphi_i(y)$, from equations (3.2), (3.3), (3.5) respectively , in equation (3.14), we get the variance of stratified sample mean in terms of $\varphi(y)$ as follows

$V(\bar{y}_{st})_{Ney} = \frac{1}{12n}\left[\sum_{i=1}^{k} \varphi_i^{\frac{1}{\alpha}}(y)\, (y_i - y_{i-1})^{\left(2-\frac{1}{\alpha}\right)}\right]^2$ (3.15)

This equation has its minimum value when $\varphi(y)$ value is equal for all strata, under Lagrange multiplier restriction $(y) = \sum_{i=1}^{k} \varphi_i(y)$, i.e $\varphi_i(y) = \frac{\varphi(y)}{k}$, by substituting the value of $\varphi_i(y)$ in equation (3.15) we get

$V(\bar{y}_{st})_{Ney} = \frac{\varphi^{\frac{2}{\alpha}}(y)}{12nk^{\frac{2}{\alpha}-2}}$ (3.16)

Based on equations (3.10), (3.13), (3.16) , we compare the variance of the stratified sample mean between different methods of allocation;

Since $R_1 = \frac{V(\bar{y}_{st})_{Ney}}{V(\bar{y}_{st})_{eq}} = 1$, $k \geq 1$

Then;

$V(\bar{y}_{st})_{Ney} = V(\bar{y}_{st})_{eq}$, for all $0 < \alpha < 1$

On other side let;

$R_2 = \frac{V(\bar{y}_{st})_{Ney}}{V(\bar{y}_{st})_{prop}} = \frac{\varphi^{\frac{1}{\alpha}}}{k} \leq 1$,

Hence;

$V(\bar{y}_{st})_{Ney} \leq V(\bar{y}_{st})_{prop}$, for all $0 < \alpha < 1$, $k \geq 1$ and $\varphi^{\frac{1}{\alpha}} \leq k$.

As a result, we conclude that the variance of stratified sample mean with Neyman allocation method has minimum variance than the proportional or equal allocation methods, for all for all $0 < \alpha < 1$

4.IMPLEMENTATION SECTION

In this section, we shall discuss the variance of stratified sample mean in different allocation method, assuming that study variable Y has an exponential distribution function in the first case and normal distribution function in the other.

a.Variance of Stratified Sample mean in Exponential distribution

Assume that $y_1, y_2, y_2, \dots, y_N$, are independent population values of the random variable Y, which has exponential distribution with probability density function $f(y) = \lambda e^{-\lambda x}$, $0 \leq x \leq \infty$, $\lambda > 1$, then

$$\varphi(y) = \int_{-\infty}^{\infty} f^{\alpha}(y)dy = \int_{0}^{\infty} \lambda^{\alpha} e^{-\lambda\alpha x} dy$$

Or;

$\varphi(y) = \frac{\lambda^{\alpha-1}}{\alpha}$; $0 < \alpha < 1$ (4.1)

By substituting the value of $\varphi(y)$ (4.1) , in the equations (3.10), (3.13) ,and (3.16) we get the values of the minimum variance of stratified sample mean in the different allocation methods as follows.

$$V(\bar{y}_{st})_{eq} = \frac{\lambda^{\frac{2(\alpha-1)}{\alpha}}}{12n\alpha^{\frac{2}{\alpha}}k^{\frac{2}{\alpha}-2}}$$

$$V(\bar{y}_{st})_{prop} = \frac{\lambda^{\frac{\alpha-1}{\alpha}}}{12n\alpha^{\frac{1}{\alpha}}k^{\frac{1}{\alpha}-1}}$$

$$V(\bar{y}_{st})_{Ney} = \frac{\lambda^{\frac{2(\alpha-1)}{\alpha}}}{12n\alpha^{\frac{2}{\alpha}}k^{\frac{2}{\alpha}-2}}$$

For example, at $\alpha = \frac{1}{2}$, we have;

$$V(\bar{y}_{st})_{Ney} = V(\bar{y}_{st})_{eq} = \frac{4}{3n\,\lambda^2 k^2}$$

While in proportional allocation the variance is;

$$V(\bar{y}_{st})_{prop} = \frac{3}{4n\lambda k}$$

In the same manner, at $\alpha = \frac{1}{3}$, we have;

$$V(\bar{y}_{st})_{Ney} = V(\bar{y}_{st})_{eq} = \frac{243}{4n\,\lambda^4 k^4}$$

And;

$$V(\bar{y}_{st})_{prop} = \frac{27}{12n\,\lambda^2\,k^2}$$

b. Variance of Stratified Sample Mean in Normal Distribution

Assume that $y_1, y_2, y_2, \dots, y_N$, are independent population values of the random variable Y, which has normal distribution with probability density function

$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$, $-\infty \leq x \leq \infty$

the value of $\varphi(y)$ is as follows

$$\varphi(y) = \int_{-\infty}^{\infty} f^{\alpha}(y)dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy$$

Hence;

$\varphi(y) = \frac{(2\pi)^{\frac{1-\alpha}{2}}}{\sqrt{\alpha}} \sigma^{1-\alpha}$ (4.2)

By substituting the value of $\varphi(y)$ in equation (3.10), or (3.16) we get the variance of stratified mean in both equal and Neyman allocation method as follows

$V(\bar{y}_{st})_{Ney} = V(\bar{y}_{st})_{eq} = \frac{(2\pi)^{\frac{1-\alpha}{\alpha}} \sigma^{2\left(\frac{1-\alpha}{\alpha}\right)} \alpha^{-\frac{1}{\alpha}}}{24nk^{\frac{2}{\alpha}-2}}$ (4.3)

In the same manner, the variance of stratified sample mean with proportional allocation method is given by substituting the value of $\varphi(y)$ in equation (4.2) into equation (3.14) to get;

$$V(\bar{y}_{st})_{prop} = \frac{(2\pi)^{\frac{1-\alpha}{2\alpha}} \sigma^{\left(\frac{1-\alpha}{\alpha}\right)} \alpha^{\frac{-1}{2\alpha}}}{12nk^{\frac{1}{\alpha}-1}} \qquad (4.4)$$

For example at $\alpha = \frac{1}{2}$, based on equations (4.3, (4.4) we have;

$$V(\bar{y}_{st})_{Ney} = V(\bar{y}_{st})_{eq} = \frac{\pi\sigma^2}{3nk^2}$$

And;

$$V(\bar{y}_{st})_{prop} = \frac{\sqrt{2\pi}\sigma}{6nk}$$

While at $= \frac{1}{3}$, we have

$$V(\bar{y}_{st})_{Ney} = V(\bar{y}_{st})_{eq} = \frac{9\pi^2\sigma^4}{2nk^4}$$

And;

$$V(\bar{y}_{st})_{prop} = \frac{\sqrt{3}\pi\sigma^2}{2nk^2}$$

REFERENCES

[1] Anderson, D.W., Kish, L and Cornell, R.G (1979)" Quantity gain from stratification for optimum and approximately optimum strata "Journal of the American Statistical Association, 71, 887-892.

[2] Daroga Singh (1986)" Theory and Analysis sample Survey Designs "1ed Wiley Eastern Limited.

[3] Cochran, W.C, (1977)" Sampling techniques "3rd edition, John Wiley, New York.

[4] Dalenius, T. and Hodges, I.L. (1959)" Minimum Variance Stratification "Journal of the American Statistical Association, 54, 88-101.

[5] Kish, L., (1995)" Survey Sampling"1st edition, John Wiley & Sons, New York.

[6] Serfling, R. J. (1968)" Approximately Optimal Stratification "Journal of the American Statistical Association, 63, 1298-1309.

[7] Al-Kassab, M. M. and Al-Tauy, H. (1994" Approximately Optimal Stratification Using Neyman Allocation". Journal of the Tanmiat Al-Tafidian,19, 31-37., 63, 1298-1309.

[8] Sethi (1963)," A note on Optimum Stratification of Population for Estimating the Population Mean "Journal of the American Statistical Association, 5, 20-30.

[9] Singh, R. and Sukhatme, BV (1969," Optimum Stratification "Journal Ann. Inst. Stat. Math, 21, 515-528.