Efficiency of Neyman allocation procedure over other allocation procedures in stratified random sampling

Olayiwola Olaniyi Mathew[1], Apantaku Fadeke Sola[1], Bisira Hammed Oladiran[2], Adewara Adedayo Amos[3]

[1]Department of Statistics, Federal University of Agriculture, Abeokuta, Ogun State, Nigeria
[2]Department of Mathematics, Lagos State Polytechnic, Ikorodu, Lagos State, Nigeria
[3]Department of Statistics, University of Ilorin

Email address:

laniyimathew@yahoo.com (O. M. Olayiwola)

Abstract: In sampling, we have interest in precision and in order to create the precision, we make use of prior knowledge of the population. We try to put the population into series of homogeneous groups and by this, the precision will be increased. When the population of interest can be divided into k homogeneous groups and the sample of observation is taken from each group, we have a stratified random sample and each group is called a stratum. The study was therefore designed to investigate the efficiency of Neyman allocation procedure over equal and proportional allocations. The data used for this research were primary data collected from ten Markets in Abeokuta, Ogun State, Nigeria on the prices of Peak Milk (Nigeria made). A stratified random sampling scheme was used in selecting 10 markets in Abeokuta, Ogun State, Nigeria. Each market stands as a stratum. From each stratum, independent sample was selected randomly based on equal, proportional and Neyman/Optimum allocation procedures. Statistic was obtained from each stratum and combined estimate of the separate statistic was also obtained for each of the allocation procedure. Considering the analysis and estimates obtained, the mean and variance under Neyman allocation procedure were 1356.672 and 21.45 respectively. For proportional allocation, the mean was 1349.3069 and variance was 38.98 while equal allocation gave mean of 1352 and variance of 170.3238. Neyman/Optimum allocation procedure gave the least variance. This was followed by Proportional allocation and Equal allocation. Neyman allocation procedure is the best selection procedure. Hence, for estimating the average and the variance of the prices of Peak Milk (Nigeria Made) in the markets in Abeokuta, of all the three sample allocation procedures considered in this paper, Neyman allocation procedure is the best and hence the most efficient.

Keywords: Efficiency, Stratified Random Sampling, Neyman Allocation, Procedure

1. Introduction

Sampling methods are designed to provide valid, scientific and economical tools for research problems. According to Kish (1965) and Hunt and Tyrell (2004), sampling plays a vital role in research design involving human population and commands increasing attention from social scientists, chemists, engineers, accountants, biologists and medical practitioners. Sampling problems are in general to practitioners engaged in marketing, commerce, industry, public health, biostatistics, education, public administration, economics, sociology, anthropology, psychology, political Science and even social workers.

Sampling methods are developed as means to an end originating in substantive research problems especially in the social sciences and their applications (Kish, 1965; Hunt and Tyrell, 2004). A working knowledge of practical sampling methods with an understanding of their theoretical background need to be a requirement for quantitatively oriented students in the social sciences as well as in allied fields. It is helpful although difficult to separate sampling design from the related activities involved in survey research. The sample design covers the tasks of selection and estimation for making inference from sample value to the population value. Beyond this are the problems of making inferences from one survey population to another and generally broader population, with measurements free from error. Different sampling designs

would result in different standard errors, and choosing the design with the smallest error is the principal aim of sampling design.

An effective sampling technique within a population represents an appropriate extraction of useful data which provides meaningful knowledge of the important aspects of the population (Garcia and Cortez, 2006). Probability samples are usually designed to be measurable, that is, so designed that statistical inference to population values can be based on measures of variability, usually standard errors, computed from the sample data.

In general, there is need to devise a sampling scheme which is economical and easy to operate, that yields unbiased estimates, and minimizes the effects of sampling variation.

Usually in sample surveys more than one population characteristics are estimated and these characteristics may be of conflicting nature. Stratified sampling has been designed to ensure that all important views are represented in samples. Stratification is a means of sample design by which the population of interest is divided into groups, called strata, according to some known characteristic(s). Stratified sample designs are employed for several reasons. These include: to increase the precision of estimates for the whole population for one or more key data items being collected in the survey, to obtain more precise estimates for interesting domains, to allow the use of different sampling, non-response adjustment, editing, or estimation methods for domains with differing characteristics affecting the choice of method, and to facilitate administration of the survey. Stratified sampling is always more restrictive than simple random sampling.

Moreover, in the context of stratified sampling, some multivariate approaches have been proposed whereby the sample size and its allocation within strata take diverse characteristics into consideration (Sukhatme *et al*., 1984 and Arthanari and Dodge 1981). The multivariate stratified sample design is different with two steps from the univariate stratified sample design. The first step is to decide strata using stratified variables which are multivariate. The second step is to decide a sample size and the optimal allocation, that is, to decide a sample size of each stratum using interest variables which are multivariate.

There are many methods for allocation to strata in the multivariate stratified sample design. The first method is the proportional allocation, and the second is the multivariate allocation using one interest variable which is selected of multivariate interest variables. The third is a compromise allocation which is a weighted average of sample size of strata using individual allocation (Cochran, 1977; Chatterjee, 1972). The fourth is the optimal allocation for a loss function of characteristic values which combine variances of all variables (Kish, 1976; Sukhatme *et al*, 1984; Bethel, 1989; Khan and Ahsan, 2003; and Diaz-Garcia and Cortez, 2008).

In multivariate stratified sample design, correlation is considered among interest variables. The first method is a compromise allocation weighted by correlation coefficients or covariances, and the second method is the optimal allocation for a loss function of characteristic values of variance-covariance matrix. A variation of one variable with lower correlation is more important than others. The third method is to use weighted object function by the importance of interest variables in a mathematical programming and of importance is the error of estimation, the correlation coefficient or the covariances. The multivariate stratified sample design is used for multi-objective surveys in which there is difference among the importance of interest variables.

The problem of allocation with more than one characteristic in stratified sampling is conflicting in nature, as the best allocation for one characteristic will not in general be best for others. Some compromise must be reached to obtain an allocation that is efficient for all characteristics. This problem was first considered by Neyman (1934), Dalenius (1957), Ghosh (1958), Kokan and Khan (1967), Khan and Ahsan (2003), Khan, Jahan and Ahsan (1997). Attempts were made for an acceptable allocation by either suggesting new criteria or exploring existing criteria further.

One of the problems of stratification is that loss in precision in the estimate of a characteristic increases if the characteristic in a stratum is not internally homogenous. To refrain from the increase in loss of precision is to assign a maximum weight to the jth characteristic.

2. Data Used

The data used for this research were primary data collected from ten Markets in Abeokuta, Ogun State, Nigeria on price of Peak Milk (Nigeria made).

3. Methodology

Theorem:

1. The average of stratified sampling is:

$$\bar{Y}_{st} = \sum_{h=1}^{k} W_h \, \bar{y}_h \qquad (1)$$

2. The variance of stratified sampling:

Recall $V(\bar{y}_{st}) = \frac{N-n}{N_n} S^2$ (SRSWOR)

$$var(\bar{y}_{st}) = var(\sum_{h=1}^{k} W_h \, y_h)$$

$$= \sum_{h=1}^{k} W_h^2 \, var(y_h) \qquad (2)$$

$$= \sum_{h=1}^{k} W_h^2 \, \frac{N_h - n_h}{N_h n_h} S_h^2$$

$$= \sum_{h=1}^{k} W_h^2 \, (1 - \frac{n_h}{N_h}) \frac{S_h^2}{n_h}$$

$$var(\bar{y}_{st}) = \sum_{h=1}^{k} W_h^2 (1 - f_h) \frac{S_h^2}{n_h} \qquad (3)$$

Where f_n is the sampling factor per stratum and $W_h = \frac{N_h}{N}$

If the stratified sampling is carried out under proportional allocation, the variance for mean is derived below:

Recall: $n_h = W_h \times n$ \hfill (4)

$$var\,(\bar{y}_{st}) = \sum_{h=1}^{k} W_h^2 \frac{N_h - n_h}{N_h n_h} S_h^2 \tag{5}$$

$$= \sum_{h=1}^{k} W_h^2 \frac{N_h - n_h}{N_h n_h} \frac{S_h^2}{N_h}$$

$$= \sum_{h=1}^{k} \frac{N_h^x}{N^2} \frac{N_h - n_h}{n_h} \frac{S_h^2}{N_h}$$

$$= \frac{1}{N} \sum_{h=1}^{k} W_h \left(\frac{N_h - 1}{n_h}\right) S_h^2$$

$$= \frac{1}{N} \sum_{h=1}^{k} W_h \left(\frac{N_h}{W_h \times n} - 1\right) S_h^2$$

$$= \frac{1}{N} \sum_{h=1}^{k} W_h \left(\frac{N_h}{\frac{N_h}{N} \times n} - 1\right) S_h^2$$

$$= \frac{1}{N} \sum_{h=1}^{k} W_h \left(\frac{N}{n} - 1\right) S_h^2$$

$$= \sum_{h=1}^{k} W_h \left(\frac{1}{n} - \frac{1}{N}\right) S_h^2 \Rightarrow \frac{\sum_{h=1}^{k} W_h S_h^2}{n} - \frac{\sum_{h=1}^{k} W_h S_h^2}{N}$$

$$= \sum_{h=1}^{k} W_h \left(\frac{N - n}{N_h}\right) S_n^2$$

$$= \sum_{h=1}^{k} W_h (1 - f) \frac{S_h^2}{n}$$

$$var(\bar{y}_{st}) = \frac{(1-f)}{n} \sum_{h=1}^{k} W_h S_h^2 \tag{6}$$

If allocation of sample over strata is done using Neyman allocation procedure, the variance for the mean of the stratified random sample is derived below. (If the cost of selecting sample per stratum is equal).
Recall:

$$n_{h = \frac{W_h S_h}{\sum_{h=1}^{k} W_h S_h}} \times n \quad \text{(Since the costs are equal)}$$

$$var(\bar{y}_{st}) = \sum_{h=1}^{k} W_h^2 \frac{N_h - n_h}{N_h n_h} S_h^2 \tag{7}$$

$$= \sum_{h=1}^{k} W_h^2 \frac{N_h - n_h}{n_h} \frac{S_h^2}{N_h}$$

$$= \frac{1}{N} \sum_{h=1}^{k} W_h \left(\frac{N_h}{n_h} - 1\right) S_h^2$$

$$= \frac{1}{N} \sum_{h=1}^{k} W_h \left(\frac{N_h}{\frac{W_h S_h}{\sum W_h S_h} \times n} - 1\right) S_h^2$$

$$= \frac{1}{N} \sum_{h=1}^{k} W_h \left(\frac{\sum W_h S_h \times N_h - W_h S_h \times n}{\frac{N_h}{N} S_h \times n}\right) S_h^2$$

$$= \frac{1}{N} \sum_{h=1}^{k} W_h S_h \left(\frac{N \sum W_h S_h}{n} - \frac{W_h S_h \times n}{W_h \times n}\right)$$

$$= \frac{1}{N} \sum_{h=1}^{k} W_h S_h \times \frac{N \sum_{h=1}^{k} W_h S_h}{n} - \frac{1}{N} \sum_{h=1}^{k} W_h S_h S_h$$

$$var(\bar{y}_{st}) = \left(\frac{\sum_{h=1}^{k} W_h S_h}{n}\right)^2 - \frac{\sum_{h=1}^{k} W_h S_h^2}{N} \tag{8}$$

$$var(\bar{y}_{st})_{op} \le var(\bar{y}_{st})_{prop} \le var(\bar{y}_{st})_{SRSWOR}$$

Proof

Recall: $var(\bar{y}_{st})_{SRSWOR} = \frac{N-n}{N_n} S^2 \tag{9}$

$$var(\bar{y}_{st})_{prop} = \frac{\sum_{h=1}^{k} W_h S_h^2}{n} - \frac{\sum_{h=1}^{k} W_h S_h^2}{N}$$

$$var(\bar{y}_{st})_{op} = \left(\frac{\sum_{h=1}^{k} W_h S_h}{n}\right)^2 - \frac{\sum_{h=1}^{k} W_h S_h^2}{N}$$

From analysis of variance:

$$S^2 = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}{N} \quad \text{If } N \to \infty \tag{10}$$

$$NS^2 = \sum_{h=1}^{k} \sum_{i=1}^{N_h} (y_{hi} - \bar{y}_h + \bar{y}_h - \bar{y})^2$$

$$NS^2 = \sum_{h=1}^{k} \sum_{i=1}^{N_h} (y_{hi} - \bar{y}_h)^2 + \sum_{h=1}^{k} \sum_{i=1}^{N_h} (y_h - \bar{y})^2 + 2 \sum_{h=1}^{k} \sum_{i=1}^{N_h} (y_{hi} - \bar{y})^1 (\bar{y}_h - \bar{y})$$

$$NS^2 = \sum_{h=1}^{k} \sum_{i=1}^{N_h} (y_{hi} - \bar{y}_h)^2 + \sum_{h=1}^{k} \sum_{i=1}^{N_h} (y_h - \bar{y})^2$$

Recall that: $S_h^2 = \frac{1}{N_h-1}\sum_{i=1}^{N_h}(y_{hi}-\bar{y}_h)^2$

$(N_h-1)S_h^2 = \sum_{i=1}^{N_h}(y_{hi}-\bar{y}_h)^2$

$$\therefore NS^2 = \sum_{h=1}^{k}(N_h-1)S_h^2 + \sum_{h=1}^{k}\sum_{i=1}^{N_h}(y_h-\bar{y})^2$$

$$S^2 = \sum_{h=1}^{k}\frac{(N_h-1)S_h^2}{N} + \frac{\sum_{h=1}^{k}\sum_{i=1}^{N_h}(y_h-\bar{y})^2}{N}$$

If $N_h \to large \Rightarrow N_h - 1 = N_h$

$$S^2 = \sum_{h=1}^{k}\frac{(N_h)S_h^2}{N} + \frac{\sum_{h=1}^{k}\sum_{i=1}^{N_h}(y_h-\bar{y})^2}{N}$$

$$S^2 = \sum_{h=1}^{k}W_h S_h^2 + \frac{\sum_{h=1}^{k}\sum_{i=1}^{N_h}(y_h-\bar{y})^2}{N}$$

Recall $var(\bar{y})SRSWOR = \frac{N-n}{Nn}S_y^2$

$$\frac{N-n}{Nn}S^2 = \frac{N-n}{Nn}\sum_{h=1}^{k}W_h S_h^2 + \frac{N-n}{Nn}\frac{\sum_{h=1}^{k}\sum_{i=1}^{N_h}(\bar{y}_h-\bar{y})}{N} \tag{11}$$

$$= \left(\sum_{h=1}^{k}\frac{W_h S_h^2}{n} - \sum_{h=1}^{k}\frac{W_h S_h^2}{N}\right) + \left(\frac{1-f}{n}\right)\sum_{h=1}^{k}\sum_{i=1}^{N_h}\frac{(\bar{y}_h-\bar{y})^2}{N} \tag{12}$$

$var(\bar{y})SRSWOR \geq v(\bar{y}_{st})prop$

with the variance of proportional allocation under stratified sampling.

Now compare the variance of optimal/Neyman allocation

$$\sum_{h=1}^{k}\frac{W_h S_h^2}{n} - \sum_{h=1}^{k}\frac{W_h S_h^2}{N} = \frac{(\sum_{h=1}^{k}W_h S_h)^2}{n} - \sum_{h=1}^{k}\frac{W_h S_h^2}{N} + \sum_{h=1}^{k}\frac{W_h S_h^2}{n} - \sum_{h=1}^{k}\frac{W_h S_h^2}{N} - \frac{(\sum_{h=1}^{k}W_h S_h)^2}{n} + \sum_{h=1}^{k}\frac{W_h S_h^2}{N} = 0$$

$$\sum_{h=1}^{k}\frac{W_h S_h^2}{n} - \frac{(\sum_{h=1}^{k}W_h S_h)^2}{n} = 0$$

Multiply through by n

$$\sum_{h=1}^{k}W_h S_h^2 - \left(\sum_{h=1}^{k}W_h S_h\right)^2 = 0$$

Since $\sum_{h=1}^{k}W_h S_h^2 - \left(\sum_{h=1}^{k}W_h S_h\right)^2 \geq 0$

$$\sum_{h=1}^{k}W_h S_h^2 \geq \left(\sum_{h=1}^{k}W_h S_h\right)^2 \tag{13}$$

Hence $var(\bar{y}_{st})prop \geq var(\bar{y}_{st})opt$

$\therefore var(\bar{y}_{st})opt \leq var(\bar{y}_{st})prop \leq var(\bar{y})SRSWOR.$
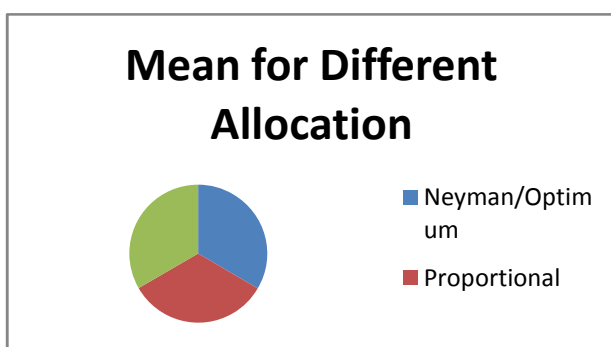
4. Summary of Results

In the selection of sample in each stratum, for equal allocation, 16 samples were taking from each stratum making total of 160 samples across 10 strata (markets). For proportional allocation, sample size varies depending on the total number of milk sellers in each market. Cost of selecting sample from each stratum (market) was put into consideration in optimum allocation. Few samples were selected from the market where the cost of sample selection is high, while many samples were selected from where there is low cost of sample selection. The results obtained are as shown in the table 1.

Table 1: *Stratification of abeokuta markets into ten strata with corresponding sample size for different allocation procedures*

Strata	Equal allocation sample size	Proportional allocation sample size	Neyman/optimum allocation sample size
1	16	14	17
2	16	20	8
3	16	13	29
4	16	10	24
5	16	25	13
6	16	14	22
7	16	18	16
8	16	20	10
9	16	14	17
10	16	12	4

Table 2: *Mean per stratum based on different allocation procedures*

Strata	Equal allocation	Proportional allocation	Neyman/optimum allocation
1	1287	1297.78	1297.778
2	1331	1336.00	1336.000
3	1282	1290.00	1290.000
4	1350	1346.67	1346.667
5	1270	1265.63	1263.333
6	1369	1376.67	1362.222
7	1375	1375.33	1373.333
8	1444	1452.50	1452.500
9	1446	1466.67	1466.667
10	1370	1337.33	1337.333

Table 3: Variance per stratum based on different allocation procedures

Strata	Equal allocation	Proportional allocation	Neyman/optimum allocation
1	1801.0687	1532.95	1535.948
2	1128.4896	2268.57	2268.571
3	2076.6249	1,057.14	1057.143
4	6666.6245	8380.95	8380.952
5	2333.2764	1906.25	2309.524
6	5248.8576	3317.65	2630.065
7	1805.4851	1698.09	1738.095
8	2471.0841	2500.00	2500
9	6048.7951	6494.12	6494.118
10	30064.092	21878.09	31878.1

Table 4: Combine mean for stratified sampling using different allocation procedures

Neyman/optimum	1356.6720
Proportional	1349.3069
Equal	1352.0000

Mean for Different Allocation

- Neyman/Optimum
- Proportional

Table 5: Combine variance for stratified sampling using different allocation procedures

Neyman/optimum	21.45
Proportional	38.98
Equal	170.3238

Variance for Different Allocation

- Neyman/Optimum
- Proportional
- Equal
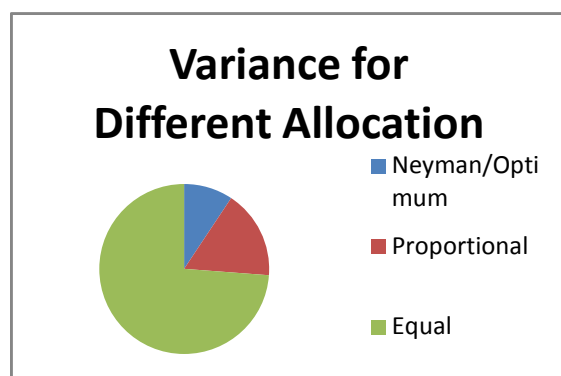
5. Discussion of Results

In the selection of sample in each stratum, for equal allocation, 16 samples were taking from each stratum making total of 160 samples across 10 strata (markets). For proportional allocation, sample size varies depending on the total number of milk sellers in each market. Cost of selecting sample from each stratum (market) was put into consideration in optimum allocation. Few samples were selected from the market where the cost of sample selection is high, while many samples were selected from where there is low cost of sample selection.

The mean and variance under Neyman allocation procedure were 1356.672 and 21.45 respectively. For proportional allocation, the mean was 1349.3069 and variance was 38.98 while equal allocation gave mean of 1352 and variance of 170.3238. Neyman/Optimum allocation procedure gave the least variance. This was followed by Proportional allocation and Equal allocation. Neyman/Optimum allocation procedure is the best selection procedure in this paper.
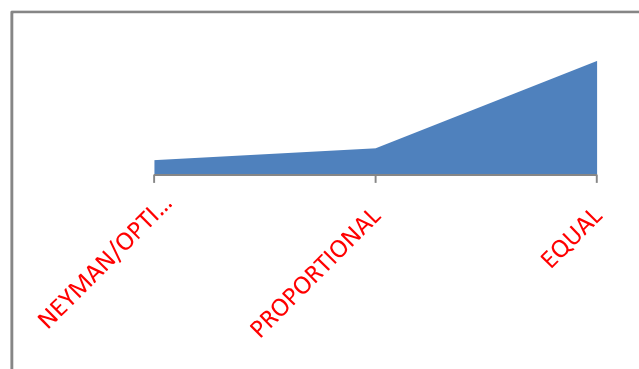
Fig1. Variance for different allocation procedures

6. Conclusion

Finally, for estimating the average and the variance of the prices of Peak Milk (Nigeria Made) in the markets, out of all the three sample allocation procedures considered in this paper, Neyman/Optimum allocation procedure is the best and hence the most efficient.

References

[1]   Kish, L. 1965. Survey sampling. *New York, Wiley.*

[2]   Hunt, N. and Tyrell, S. (2004), Stratified sampling. *Coventry UniversityPress.*
http://www.coventry.ac.uk/ec/¬nhunt/meths/strati.html (accessed February 28, 2011)

[3]   Arthanari, T.S and Dodge, Y. 1981.   Mathematical programming on statistics. *A Wiley-Interscience, Publication, John Wiley & Sons Inc.*

[4]   Bethel, F. 1989. Bayes and Minimax prediction in finite population. *Journal of Statistical Planning*, 60, 127 – 135.

[5]   Chatterjee, S. 1972. A study of optimal allocation in multivariate stratified surveys. *Skand Akt*. 73, 55 – 57.

[6]   Cochran, W. G. 1977. Sampling Techniques (3rd Edition), New York, Wiley.

[7] Dalenius, T. 1957. Sampling in Sweden: Contributions to the methods and theories of sample survey practice, *Almavist and wicksell, Stockholm*.

[8] Diaz-Garcia, J.A and Cortez, L.U. 2008. Multi-objective optimisation for optimum allocation in multivariate stratified sampling. *Survey Methodology,* Vol. 34, No 2, 215-222.

[9] Ghosh, S.P., 1958. A note on Stratified Random Sampling with Multiple Characters. Col.Stat. Bull, 8, 81-89.

[10] Hunt, N. And Tyrell, S. 2004. Stratified sampling. *Coventry UniversityPress.* *http://www.coventry.ac.uk/ec/~nhunt/meths/strati.html (accessed February 28, 2011).*

[11] Khan, M.G.M and Ahsan, M.J. 2003. *A note on Optimum Allocation in Multivariate   Stratified Sampling. South*

Pacific Journal Natural Science, 21, 91-95.

[12] Khan, M.G.M, Jahan, N. and Ahsan, M.J. 1997. Determining the optimum cluster size. *Journal of the Indian Society of Agricultural Statistics.* Vol. 50 (2), 121-129.

[13] Kish, L. 1965. Survey sampling. *New York, Wiley*.

[14] Kokan, A.R and Khan, S.U., 1967. Optimum allocation in mutivariate surveys. An analytical solution. *Journal of Royal Statistical Society*. Series B, 29, 115-125.

[15] Neymaan, J. 1934. On the two different aspects of the representative methods. The method stratified sampling and the method of purposive selection. *Journal of Royal Statistical Society*, 97, 558-606.

[16] Sukhatme, P.V, Sukhatme, B.V, Sukhatme, S., and Asok, C. 1984. *Sampling Theory of Survey with Applications.* *3rd Edition. Ames, Iowa: Iowa State University Press*.