Pak. J. Commer. Soc. Sci.
2010 Vol. 4 (2), 194-201

Estimating Income Variances by Probability Sampling: A Case Study

Akbar Ali Shah and M. Aleem
Department of statistics,The Islamia university of Bahawalpur. Pakistan
E-mails: akbar_alishah@yahoo.com, draleemiub@hotmail.com

Abstract

The main focus of the study is to estimate variability in income distribution of households by conducting a survey. The variances in income distribution have been calculated by probability sampling techniques. The variances are compared and relative gains are also obtained. It is concluded that the income distribution has been better as compared to first Household Income and Expenditure Survey (HIES) conducted in Pakistan 1993-94.

Keywords: Variability, Income distribution, Probability Sampling, Relative gain, Gini Coefficient.

1. Introduction

Estimating income variability in different strata of society has ever been a subject of research. The huge disparity in our socio economic setup has impeded out national growth. Incomes are distributed according to ability. It is assumed that ability is normally distributed so incomes are expected to be normally distributed. The inequality in income distribution in Pakistan during 1960s has decreased but in 1970s it has increased (Guisinger and Hicks 1978). The more disparity in income distribution, the lower the individual's happiness (Bigsten, 1983). Data on income distribution shows the household income shares among different income groups of society. In Pakistan the Household Income and Expenditure Survey (HIES) were conducted in 1963-64, 1966-67, 1968-69 to 1971-72, 1979, 1986-87 and 1987-88. Another Household Integrated Economic Survey was carried out in 1990-91. It is evident from the first phase of the survey spreading over 1963-71 that inequality in income distribution has narrowed while second phase hovering over 1971-79 indicating that there is stabilization at existing level of inequality. The ratio of highest to lowest 20% income group which was 7.1 in 1963-64 reduced to 4.9 in 1970-71 and it was 6.1 in 1979. In 1984-87 ratio decreased from 6.2 to 5.5. In 1987-91 the ratio of highest 20% to lowest 20% has decreased from 5.5 to 8.6. All above trends are also visible in the the Gini Co-efficient which is used as a general index for overall income distribution. Gini ratio is a measure of income disparity. It ranges from 0 to 1 meaning that each percentile of household getting the equal income and one income class has all the income and everyone else has nothing (Economic Survey 1993-94)

2. Data and Methodology

This empirical study is based on the primary source of data conducted by a survey of 350 households of Bahawalpur district , a city of Southern Punjab Province of Pakistan. This study extends over a period of one year (2007-2008) . The households of Satellite Town of Bahawalpur district is the study population. Initially the data was collected by the enumerators by simple random sampling technique with the objective that each household has equal chance of being included in the survey of 350 households. The collected data is then stratified taking income distribution as stratifying factor. The number of households falling in different income groups, the averages of income and standard deviations of the particular income groups are summarized in Table 1.

Table 1: Stratified Income Groups

Income Group	No. of Household	Average Income	Standard Deviation
Low Income Group	133	12133.85	3465.88
Middle class Income Group	104	24349.98	4075.66
High Income Group	73	42459.59	6085.49
Very High Income Group	40	842921.26	1984.35

Low Income Group 6000-18000, Middle Class Income Group 18001-30000, High Income Group 30001-54000, Very High Income Group 540001&Over.

From the collected data taking every Kth (K=7) income of the household, seven systematic samples of size 50 each are obtained to calculate variance for comparison. The results obtained by simple random sampling stratified random sampling and systematic sampling techniques are summarized in Table 2. The data collected on income distribution enables to measure the household income shares between different income groups.

The following methodologies are used for analysis :

Let variance of mean in stratified random sampling (Hansen, Hurwitz, Madow, 1953 Vol.I)

$$\frac{(1-f_h)}{n_h N^2}\frac{\sum_{h=1}^{L} N_h^2 S_h^2}{N_h} = \frac{1-f_n}{n_h N_h}\bar{S}_h^{\,2}$$

Where $S_h^2 = \dfrac{\sum_{i=1}^{N_h}\left(X_{hi} - \bar{X}_h\right)^2}{N_h - 1}$

$$\bar{S}_h = \frac{1}{N}\sum_{h=1}^{L} N_h S_h$$

$$\mathrm{N}\,\bar{S}_h = \sum_{h=1}^{L} N_h S_h$$

$$\bar{S}_h^{\,2} = \frac{\sum_{h=1}^{L} N_h^2 S_h^2}{N^2} + C.P.T.$$ as different stratum are independent to each other so C.P.T. will vanish.

So

$$\bar{S}_h^{\,2} = \frac{\sum_{h=1}^{L} N_h^2 S_h^2}{N^2}$$

$$f_h = \frac{n_h}{N_h} \qquad f_h = f$$

Then

1- $n_h = fN_h$ When proportional allocation is used to select elements in stratified random sampling.

2- $N \bar{S}_h \, n_h = nN_h S_h$ When Optimum allocation is used to select elements in stratified random sampling.

The variance of a simple random sample mean is

$$\frac{(1-f)}{n} S^2 \qquad (1)$$

Where $f = \frac{n}{N}$ is sampling fraction

$$S^2 = \frac{\sum_{i=1}^{N} \left(X_i - \bar{X} \right)^2}{N-1}$$

The variance of stratified random sampling when proportional allocation is used

$$\frac{1-f}{n} S_w^2 \qquad (2)$$

Where $S_w^2 = S_h \bar{S}_h$

Gain in precision using proportional allocation in stratified random sampling as compared to simple random sampling.
Considering (1)

$$\frac{1-f}{n} S^2 = \frac{1-f}{n} \frac{N}{N-1} \sigma^2$$

$$\frac{1-f}{n} \frac{N}{N-1} \left[\sigma_b^2 + \sigma_w^2 \right]$$

As $\sigma^2 = S^2$ and total population variance is written as the sum of the variances between stratum means and elementary units within strata respectively i.e. $\sigma_b^2 + \sigma_w^2$ (Hansen, Hurwitz, Madow, 1953 VolI)

$$\sigma_b^2 = \frac{\sum_{h=1}^{L} N_h \left(\bar{X}_h - \bar{X} \right)^2}{N}$$

$$\sigma_w^2 = \frac{\sum_{h=1}^{L} \sum_{i=1}^{N_h} N_h \left(X_{hi} - \bar{X}_h \right)^2}{N}$$

Assuming N is large so $\frac{N}{N-1} \cong 1$ and $\frac{N_h}{N_h - 1} \cong \frac{\bar{N}}{\bar{N}-1}$

$$S_w^2 = \frac{\bar{N}}{\bar{N}-1} \sigma_w^2$$

Subtracting (2) from (1)

$$\frac{(1-f)}{n} (S^2 + S_w^2) \qquad (3)$$

$$\frac{(1-f)}{n} \left(\sigma^2 - \frac{\bar{N}}{\bar{N}-1} \sigma_w^2 \right)$$

$$\frac{(1-f)}{n}\left(\sigma_b^2+\sigma_w^2-\frac{\bar{N}}{\bar{N}-1}\sigma_w^2\right) \quad (4)$$

Relative gain in precision due to proportional allocation in stratified random sampling as compared to simple random sampling is obtained by dividing (4) by (1)

$$\frac{\frac{(1-f)}{n}\left(\sigma_b^2-\frac{S_w^2}{\bar{N}}\right)}{\frac{(1-f)}{n}S^2}$$

$$\frac{\left(\sigma_b^2-\frac{S_w^2}{\bar{N}}\right)}{S^2}$$

The variance of optimum allocation used in stratified random sampling is

$$\frac{\left(\sum_{h=1}^{L}N_hS_h\right)^2}{nN^2}-\frac{\sum_{h=1}^{L}N_hS_h^2}{N^2} \quad \text{(Cochran, 1977)}$$

Where $\bar{S}_h' = \frac{\sum_{h=1}^{L}N_hS_h^2}{N}$

$$N\bar{S}_h' = \sum_{h=1}^{L}N_hS_h^2$$

$$\frac{\left(\sum_{h=1}^{L}N_hS_h\right)^2}{nN^2} \quad \text{without fpc} \quad (5)$$

The variance of proportional allocation used in stratified random sampling is

$$\frac{\sum_{h=1}^{L}N_hS_h^2}{nN}-\frac{\sum_{h=1}^{L}N_hS_h^2}{N^2}$$

$$\frac{\sum_{h=1}^{L} N_h S_h^2}{nN} \quad \text{without fpc} \qquad (6)$$

Relative gain in precision due to optimum allocation in stratified random sampling as compared to proportional allocation is obtained by subtracting (5) from (6) and dividing by (5)

$$\frac{\dfrac{\sum_{h=1}^{L} N_h S_h^2}{nN} - \dfrac{\left(\sum_{h=1}^{L} N_h S_h\right)^2}{nN^2}}{\dfrac{\sum_{h=1}^{L} N_h S_h^2}{nN}}$$

$$\left(\frac{N\sum_{h=1}^{L} N_h S_h^2 - \left(\sum_{h=1}^{L} N_h S_h\right)^2}{N} \right) \Big/ \sum_{h=1}^{L} N_h S_h^2$$

$$\left(\frac{N^2 \left(\bar{S}'_h - \bar{S}_h^{\,2}\right)}{N} \right) \Big/ N\bar{S}'_h \qquad (7)$$

The variance of systematic sampling is

$$V(\bar{y}_{sy}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{Y})^2$$

Using above methodologies the results are obtained and summarized in Table 2.

Table 2: Variances and Relative Gains

Techniques	Variance	Relative Gain
Simple Random Sampling	561989560	Optimum allocation over simple random sampling 0.9943
Stratified Random Sampling Optimum Allocation (n_1 =8, n_2 =7, n_3=8, n_4 = 13)	3195050	Proportional allocation over optimum allocation 0.7063
Stratified Random Sampling Proportional Allocation (n_1 =13, n_2 =10, n_3=8, n_4 = 5)	10877372	Proportional allocation over systematic sampling 0.7591
Systematic Sampling	93453986	Systematic sampling over simple random sampling 0.8337

3. Discussion and Concluding Remarks

Analysis show that the variance in income distribution groups by simple random sampling is large as compared to stratified random sampling and systematic sampling among population of Bahawalpur district. The relative gain by optimum allocation of stratified random sampling over simple random sampling is higher as compared proportional allocation of stratified random sampling and systematic sampling. The relative gain by systematic sampling over simple random is higher as compared to proportional allocation and optimum allocation of stratified random sampling. This is due to larger variation in income distribution among systematic sample (Cochran, 1977).

Table 1 shows that the income distribution has narrowed as compared to the first Household Income and Expenditure Survey (HIES) 1963-64 in Pakistan. It is evident from the survey that 38% of population falls in low income group which strengthens the fact that 45% of the population in Asia are living below poverty line while 30% of the population falls in middle income group. They could manage to live with low standard of health and education. Their basic economic needs are hardly fulfilled. A small fraction of society is falling in high income group. They have reasonably good standard of living with health and education. They have all the amenities of life. A very small segment of population (11%) of Bahawalpurdistrict fall in very high income group. The rural population have lower standard of living compared to their urban counterpart. The reason is that their poor possession of economic assets, shortage of socio physical infrastructure and high rate of underemployment. The income distribution has improved in rural areas as compared to urban areas in all the Household Income and Expenditure Survey (HIES) except 1990-91. The contribution of lowest 20% in the rural area is consistently higher than in the urban area while the highest 20% which attributed lower share in rural areas as compared to urban areas. (Economic Survey 1993-94, Economic Survey 2006-07).

Acknowledgments. Authors are thankful to the editors and anonymous reviewers for their insightful comments to improve the earlier draft of this paper.

References

Bigsten, A. (1983). *Income Distribution and Development*. Heinemann London.

Guisinger,S.and Hicks,N.L.(1978). Long term trend in income distribution in Pakistan, World Development Report, Vol.6.

Hansen, M.H , Hurwitz, W.N. and Madow, W.G. (1953). *Sample Survey Methods and Theory.* Vol. I. Methods and Applications. John Wiley & Sons Inc. London.

Hansen, M.H , Hurwitz, W.N. and Madow, W.G. (1953). *Sample Survey Methods and Theory*. Vol. II. Methods and Applications. John Wiley & Sons Inc. London.

Cochran, W.G. (1977). *Sampling Techniques*. 3rd Edition. John Wiley & Sons Inc. New York.

Economic Survey (1993-94). Government of Pakistan Finance Division Economic Advisor’s Wing Islamabad.

Economic Survey (2006-07). Government of Pakistan Finance Division Economic Advisor’s Wing Islamabad.