Sindh Univ. Res. Jour. (Sci. Ser.) Vol.44 (2) 183-188(2012)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Estimation and comparison of ratio and regression estimates of true mean food expenditure data

I. A. ARSHAD, A. GHAFOOR* A. W. SHEIKH*

Department of Mathematics and Statistics, International Islamic University, Islamabad,

Corresponding author, A.W. SHAIKH., email: Awshaik786@yahoo.com Cell No. 92-3023921251

Received 8th January 2012and Revised 16th April 2012

Abstract: Data about family income and food expenditures is distributed into six income strata. Mean per unit, Ratio type and Linear Regression estimation methods are used to estimate true mean food expenditures of families in each income strata. It is statistically decided that Ratio method is appropriate for estimating mean food expenditures of families in 1st, 4th and 5th income strata and linear regression estimation is appropriate for 2nd and 3rd income strata.

In comparison of Mean per unit, ratio type and linear regression methods, it can be concluded that linear regression is more efficient as compared to ratio and mean per unit and Ratio estimation is more efficient than mean per unit estimation methods. On the basis of satisfied conditions of regression analysis on these methods, It is concluded that estimates of true mean monthly food expenditures using ratio types estimation for 1st, 4th and 5th income groups' families are 4987Rs, 18412Rs and 27470Rs with corresponding standard errors (precisions of estimates) 9.79Rs, 31.62Rs and 17.60Rs respectively and estimates based on linear regression for 2nd and 3rd strata are 10844Rs and 15192Rs with their measures of precisions 15.85Rs and 27.24Rs respectively. Estimate of true mean in last income stratum is found theoretically quite strange and estimate obtained from mean per unit is more precise.

In analysis of food consumed, income elasticity varies from 0.12 to 0.63, which means as income increases, expenditure on food also increases but at the lesser rate. Also average food purchased/adult and average food consumed/adult among strata vary from 1881Rs to 7176Rs and from 1841Rs to 7068Rs respectively. Stability exists in expenditures on cereals origin among income strata and a steady sharp uprising trend on vegetable origin and on animal origin exists. Variation in monthly food expenditures exists from 2111Rs to 11890Rs in vegetable origin and from 2980Rs to 19296Rs in animal origin. In analysis of calories intake from specific food origins, per capita/day calories intake increases in cereals, vegetables and in animal origins and vary from 1416 to1741, from 62 to 239 and from 100 to 839 respectively.

Keywords: Mean per unit; Ratio; Linear Regression; Estimation; Calories; Precision.

1 INTRODUCTION

Lot of research has been done on variation in food consumption pattern between societies and other elements of culture but in present study an attempt has been made to develop precise estimates of true mean food expenditures of families which are stratified according to their monthly income into six non overlapping strata. For this purpose Mean per unit estimation, Ratio estimation and Linear Regression estimation methods are used. Relative precision of estimates of true mean food expenditure and relative efficiencies of these methods, for different income groups, are compared. In addition to increase the precision of the estimate of mean food expenditure (y_i), income of families (X_i), is used as an auxiliary variable. Moreover various results are obtained regarding food taken cereal origin, vegetable origin and animal origin.

Watson (1937) described method of estimating the mean leaf area per leaf or per plant of a field crop by means of regression. The mean weight per leaf is determined by a large sampling and the leaf area: leaf weight ratio and its regression on leaf weight are estimated on a small subsidiary sample. Alternative methods of estimation from the mean leaf weight and either the unweighted or the weighted mean leaf area: leaf weight ratio is shown to give positively biased estimates of mean leaf area.

Eckler (1955) explained the performance of a number of estimators such as simple expansion, ratio, unbiased ratio, regression, unequal probability types and stratified sampling is examined with the simple expansion estimator in three forest tree populations.

Rao *et al.* (1967) constructed estimates for total volume, height, and crown area for test populations. Independent variables employed were diameter, height, and crown area. Relative performance was evaluated using estimates of sampling variances and biases obtained from repeated sampling of the test populations. For the larger sample sizes, linear and parabolic regressions estimators were the best. For the smaller

*Department of Statistics, Govt. College 75 S.B. Sargodha,

**Institute of Mathematics & Computer Science, University of Sindh, Jamshoro.

sample sizes, linear regression, Horvitz-Thompson pps, and ratio-of-means estimators were best.

Small-sample efficiencies of eight ratio estimators of the population ratio $\ overline(Y) /\ overline {X} $ were investigated by Monte Carlo methods, assuming a linear regression of y on x and x is normally distributed. Efficiencies of three variance estimators were also investigated. Sample criteria for which expected values exist were used. From this study, Mickey's unbiased ratio estimator and approximately unbiased ratio estimator obtained by Quenouille's method appeared promising.

Goswami (1994) investigated consumption of dairy products among different income groups across 200 households of Shillong town in Meghalaya. Families with income of Rs.4000 per month incurred highest expenditure i.e. Rs.700.08 per month on milk and milk products when compared to lower income group. Milk and milk products were expenditure elastic for all income groups. Butter and ghee were adjudged as luxury item by all respondents.

George (1980) discussed changes in consumption levels according to the socio-economic characteristics using cross-sectional data. The study found that there had been a decline in both expenditure and quantity elasticity, while quantity elasticity of cereals had increased.

Finkner (1950) determined most efficient sampling procedure for estimating commercial peach production in North Carolina. Information was obtained on the number of peach trees and the estimated total peach production in each orchard. The high correlation between two variables suggested ratio estimate and four allocation methods were compared. The variances and relative precisions were compared and method in which allocation was proportional to total number of peach trees in the stratum appeared superior to the others. Objectives of study are:

- To obtain estimates of true monthly mean food expenditures using different estimation methods, for various families and to make comparison of estimates w.r.t their standard errors.
- To develop possible ranges for true mean food expenditures for families in various income strata.
- Another objective is to access relationship between income and food expenditure and to compare calories intake/adult family unit among income groups with respect to cereals, vegetable and animal origin.
- Finally to get an idea about calories intake/adult/day in various income groups.

2 MATERIAL AND METHODS

Data about most commonly used food commodities such as cereals, pulses, fruits and fruit products, milk and milk products, meat and eggs is collected from different families by using simple random sample of size n = 569 with the help of pretested questionnaire. Collected data were then stratified into six strata according to income of families; detail is in table1. Mean per unit, ratio estimation and linear regression estimation methods are used to estimate true mean food expenditures of families in various income strata and precision of methods and as well as of estimates are compared. Moreover confidence intervals (ranges) for true mean expenditures are also discussed so as to get an idea about the true value of mean expenditures at certain level of confidence.

Table 1 Sample size in various income strata

Income groups k_i (strata)		n_k
<10,000	(1st)	132
10,001-20,000	(2nd)	145
20,001-35,000	(3rd)	110
35,001-50,000	(4th)	083
50,001–65,000	(5th)	055
65,001 >	(6th)	044

2.1 Mean per unit estimation

Estimates of true mean monthly food expenditures in each stratum by mean per unit method are defined as:

$$\hat{\mu}_k = \frac{\sum_{i=1}^{n} y_i}{n_k} \qquad 1 \le k \le 6 \qquad (1)$$

Its estimated variance which is used to measure the precision of estimate (when N is unknown) is given as:

$$v(\hat{\mu}_k) = \frac{1}{n_k}\left[s^2{}_y\right] \qquad 1 \le k \le 6 \qquad (2)$$

Also 100(1-α) % confidence interval for μ_k (for large sample size) is given in the form of probability statement as:

$$P\left[\hat{\mu}_k - Z_{\propto/2}\sqrt{v(\hat{\mu}_k)} \le \mu_k \le \hat{\mu}_k + Z_{\propto/2}\sqrt{v(\hat{\mu}_k)}\right] = 1-\propto \qquad 1 \le k \le 6 \qquad . \quad (3)$$

2.2 Ratio type estimation

In ratio estimation of true mean food expenditure, income of families X_i is taken as auxiliary variable which is correlated with food expenditures Y_i in each stratum. The objective of using auxiliary variable in ratio type estimation is to obtain increased precision of estimates by taking the advantage of correlation between food expenditure and income in each stratum and is given as:

$$\hat{\mu}_{r(k)} = \frac{\bar{y}_k}{\bar{x}_k}\mu_{x(k)} \qquad 1 \le k \le 6 \qquad (4)$$

Where "r" stands for ratio

Ratio estimates are consistent and biased which both are negligible in large sample. Estimated variance of estimate given in (4), when N is unknown, is given as:

$$v(\hat{\mu}_{r(k)}) = \frac{1}{n}\left[s^2{}_{y(k)} + \hat{R}^2{}_{(k)}\, s^2{}_{x(k)} - 2\hat{R}_{(k)} s_{xy(k)}\right] \qquad 1 \le k \le 6 \qquad (5)$$

$$where \qquad \hat{R}_{(k)} = \frac{\bar{y}_k}{\bar{x}_k}$$

Limiting sampling distribution of ratio estimates becomes normal as n tends to infinite. So for large sample size 100(1-α) % confidence interval for $\mu_{r(K)}$ is given in the form of probability statement as:

$$P\left[\hat{\mu}_{r(k)} - Z_{\propto/2}\sqrt{v(\hat{\mu}_{r(k)})} \le \mu_{r(k)} \le \hat{\mu}_{r(k)} + Z_{\propto/2}\sqrt{v(\hat{\mu}_{r(k)})}\right] = 1-\propto \qquad 1 \le k \le 6 \qquad (6)$$

2.3 Linear regression estimation

Linear regression estimation of true mean food expenditures also use auxiliary information X_i, which is correlated with food expenditures Y_i. In practice when

- Linear relationship between Y_i and X_i exists.
- The regression line does not pass through origin.

Then estimates based on linear regression are appropriate. In this case, true mean of covariate *i.e.* income μ_x must be known and is given as:

$$\hat{\mu}_{lr(k)} = \hat{\mu}_k + b_{yx(k)}(\mu_{x(k)} - \bar{x}_k) \qquad 1 \le k \le 6 \qquad (7)$$

where "*lr*" stands for linear regression

Where, $b_{yx\ (k)}$ is regression estimate that measures average change in food expenditures when income is increased by one unit in kth stratum. An unbiased sample estimate of V ($\hat{\mu}_{lr(k)}$) is given as:

$$v(\hat{\mu}_{lr(k)}) = \frac{1}{n_k}\left[s^2{}_{y(k)} + b^2{}_{yx(k)}\, s^2{}_{x(k)} - 2b_{yx(k)} s_{xy(k)}\right] \qquad 1 \le k \le 6 \qquad (8)$$

100 (1-α) % confidence interval for $\mu_{lr(k)}$ for large sample is given in the form of probability statement as:

$$P\left[\hat{\mu}_{lr(k)} - Z_{\propto/2}\sqrt{v(\hat{\mu}_{lr(k)})} \le \mu_{lr(k)} \le \hat{\mu}_{lr(k)} + Z_{\propto/2}\sqrt{v(\hat{\mu}_{lr(k)})}\right] = 1-\propto \qquad 1 \le k \le 6 \qquad (9)$$

Confidence intervals (ranges) given in (3), (6) and (9) also can be used for testing particular value of true mean food expenditure.

2.4 Income elasticity (IE)

Income Elasticity w.r.t. consumption expenditure discussed by Richard *et al.* (1991) measures, how much consumption expenditure increases with the increase in income and is defined as:

$$\boldsymbol{I.E} = \frac{\dfrac{\Delta\, \boldsymbol{EXP.}}{\boldsymbol{Sum\ of\ Expenditures\ of\ two\ adjecent\ groups}}}{\dfrac{\Delta\, \boldsymbol{INCOME}}{\boldsymbol{Sum\ of\ Income\ of\ same\ two\ groups}}}$$

Where
Δ EXP = Change in Expenditure.
Δ INCOME = Change in Income.

3 RESULTS AND DISSCUSSIONS

Collected data contains information about monthly family income, expenditure on food in rupees (Rs.) and quantity of food consumed (grams) and analyzed in different dimensions to achieve desired interpretations.

For this purpose income-food expenditure data is distributed into k = 6 non-overlapping income strata. Estimated monthly income, estimated monthly food expenditure with their standard errors (as measure of precision) and quantity of food consumed for each stratum is obtained and relative efficiencies of various methods are compared.

3.1 Residual analysis

Regression and correlation analysis is performed on income-food data in each stratum separately. Residual analysis for all income strata reveals that linear

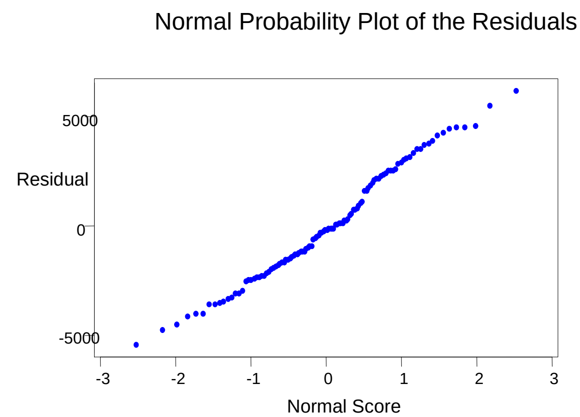

Fig.1 Normal probability plot for (20001-35000)

regression model for income (X_i) and food expenditure (Y_i) is appropriate *i.e.* there is no lack of fit and normal probability plot of residuals (for illustration in third income stratum) given in **(Fig.1),** shows the tendency of points along a straight line through origin having angle of 45^0 with X-axis and gives an indication of normality of residuals as well as of observation which is desired assumption for making test of significance. Summary of results obtained from regression and correlation analysis along with their probability values (P-values) are presented in **(Table 2).**

Table 2 Regression Analysis of X_i and Y_i

Strata K_i	$b_{yx(k)}$ (P-Value)	$a_{yx(k)}$ (P-Value)	$r_{xy(k)}$ (P-Value)
1st	0.85 (0.001)	1880.88 (0.416)	0.94 (0.000
2nd	0.51 (0.000)	2989.60 (0.009)	0.69 (0.000)
3rd	0.40 (0.008)	2539.52 (0.000)	0.80 (0.007)
4th	0.70 (0.000)	3601.04 (0.255)	0.54 (0.007)
5th	0.58 (0.000)	6485.28 (0.194)	0.94 (0.008)
6th	0.51 (0.001)	7176.00 (0.033)	0.84 (0.000)

It is clear from table 2 that y-intercept of regression of food expenditures on income for strata (< 10,000), (35,001-50,000) and (50,001-65,000) are highly non-significant and also have highly significant positive correlation between income and food expenditures. It means that for these income strata, regression line passes through origin and we can obtain ratio type estimates for true mean food expenditures. But on the other hand y-intercept of regression line of food expenditures on income for income strata (10,001-20,000), and (20,001-35,000) are highly significant with high positive significant correlation between income and food expenditures. So we can conclude that for these two income strata, regression line does not passes through origin and so linear regression type estimates for true mean food expenditures are appropriate. It can also be observed that y-Intercept of regression line in last income strata (65,001 >) has p-value that lies between 1% and 5% , it means that y-intercept is significant at 1% level of significance but it is non-significant at 5% level of significance. So for this income stratum both ratio and linear regression estimates for true mean food expenditures are appropriate.

3.2 Estimation based on Mean per unit, Ratio Type and Linear regression methods

Estimates of true mean food expenditures and their standard errors in various income strata by mean per unit and ratio type method are presented in table 3. It can be observes that estimated mean food expenditures vary from 7523Rs to 35588Rs with their standard errors from 23.95Rs to 48.91Rs in mean per unit method and average estimated food expenditures vary from 4987Rs to 27470Rs with corresponding standard errors vary from 9.79Rs to 17.60Rs in ratio type method in first five income strata. So it can be concluded that estimated mean food expenditures by ratio method are more precise as compared to estimates obtained by mean per unit because of having less standard error of estimates in first five income strata.

Table 3 Estimates of Mean per unit and Ratio method

Strata K_i	Mean/unit $\hat{\mu}_K$ (Rs.)	Standard Error $\sqrt{var(\hat{\mu}_{(K)})}$	Ratio type $\hat{\mu}_{R(K)}$ (Rs.)	Standard Error $\sqrt{var(\hat{\mu}_{R(K)})}$
1st	7523	23.95	4987	9.79
2nd	11958	21.79	10438	16.69
3rd	15237	32.27	15134	28.06
4th	21606	47.03	18412	31.62
5th	33588	48.91	27470	17.60
6th	35880	64.23	28599	116.53

The estimate obtained in last income stratum is surprisingly theoretically different and in this case estimate of true mean food expenditures by mean per unit method is 35880Rs with standard error 64.23Rs is more precise than estimate obtained by ratio method, because of having smaller standard error of estimate. Estimates along with their standard errors (measures of precisions) by mean per unit and linear regression methods are presented in table 4. It can be observed that estimated mean food expenditures vary from 4116Rs to 26409Rs with corresponding standard errors vary from 7.91Rs to 15.54Rs in the first five income strata, which are much lesser than mean per unit method. It means that linear regression estimation method provided more precise estimates of mean food expenditures in first five strata because of having less standard error of estimates.

Table 4 Mean/unit and Linear regression estimates.

Strata K_I	Mean/unit $\hat{\mu}_K$ (Rs.)	Standard Error $\sqrt{var(\hat{\mu}_K)}$	Linear Reg. $\hat{\mu}_{lr(K)}$ (Rs.)	Standard Error $\sqrt{var(\hat{\mu}_{lr(K}$
1st	7523	23.95	4116	7.91
2nd	11958	21.79	10844	15.82
3rd	15237	32.27	15162	27.24
4th	21606	47.03	16444	28.10
5th	33588	48.91	26409	15.54
6th	35880	64.23	26467	109.98

Since Mean per unit provides theoretically unbiased estimates and both ratio and linear regression gives biased estimators but in present study linear regression estimates were observed more precise as compared to mean per unit and ratio method because of smaller measures of precision in first five income strata.

Here again from table 3 and table 4, estimate of true mean obtained in last income stratum is surprisingly theoretically different and in this case estimate of true mean food expenditures by mean per unit is 35880Rs with standard error 64.23Rs, and is more precise than estimates from ratio type and linear regression methods.

The reason of theoretically strange results in last income stratum by mean per unit, ratio type and linear regression method may be due to exaggeration of food expenditure by families or it may be due to small sample size or it may be due to open-ended class in last stratum.

3.3 Ranges of true mean expenditures

So far we have discussed point estimates of true mean foods expenditures, but now interval estimates of true mean expenditures for various strata by using relation (3) ,(6) and (9) in mean per unit, ratio and linear regression estimation methods are explained. Interval estimates provides a possible range or confidence intervals (C.I.) of values in which true mean food expenditures are expected to lie under certain level of confidence. Ranges of true mean expenditures by mean per unit, ratio and linear regression estimation methods are presented in **(Table 5)**.

Table 5 Range of true mean food expenditures (μ_y)

k_i	Mean/unit C.I.	Ratio type C.I.	Linear Reg. C.I.
1st	7477-7570	4968-5006	4101-4132
2nd	11916-12001	10406-10472	10814-10876
3rd	15174-15300	15079-15189	15109-15215
4th	21514-21699	18350-18474	16389-16499
5th	32331-32522	27436-27505	26379-26440
6th	35482-36278	28371-28827	26251-26682

Ranges for true mean expenditures given in table 5 are wider in mean per unit and ratio estimations as compared to linear regression estimation method because of having smallest standard errors of estimates in linear regression method. So we can conclude that true mean food expenditures for first five income strata lies in the ranges (4101-4132), (10814-10876), (15109-15215), (16389-16499) and (26379-26440) respectively but possible range of true mean for last stratum is (26251-26682) under 95% level of confidence.

3.4 Analysis of food purchased/consumed

Average food purchased/adult and average food consumed/adult family unit for various income strata in table 6 vary from 1881Rs to 7176Rs and from 1841Rs to 7068Rs respectively and shows a steady increase in expenditures with the increase in income.

Table 6 Average food purchased and consumed

Strata (ki)	Average food purchased /adult	Average food consumed /adult	I.E
1st	1881	1841	0.58
2nd	2990	2930	0.51
3rd	2540	2490	0.63
4th	3601	3521	0.12
5th	6485	4381	0.34
6th	7176	7068	0.41

In table 6 income elasticity (I.E) also varies from 0.12 to 0.63 and lies within limits from zero to one which is logically and theoretically true and can be interpreted as income increases, expenditure on food also increases but at a lesser rate. Also note that I.Es of last three income strata are less than 0.50 which indicates of saving in high income families but on the other hand low income families have much less saving because their IE are above 0.50 and only fulfill their food requirements from income.

3.5 Expenditures on specific food origins

Average expenditure on various food origins such as cereals origin (includes flour, pluses and rice), vegetable origin (includes vegetables and vegetables products fruits and fruit products) and animal origin (includes meat eggs milk and milk products) for each income strata are shown in table 7. Expenditures on cereals origin vary from 2432Rs to 5612Rs in various income strata which show first increasing and then decreasing trend. It can also be viewed from Fig 2 that there exists approximately stability in expenditures on cereals origin among various income strata.

Table 7 Expenditures from various origins

Strata (k_i)	Cereals	Vegetables	Animals
1st	2432	2111	2980
2nd	4141	3560	4257
3rd	3870	4048	7318
4th	5612	7088	8905
5th	5538	9754	17133
6th	4693	11890	19296

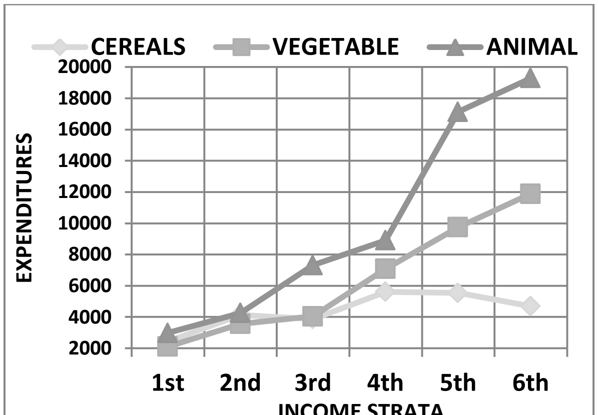

Fig. 2 Expenditures on major Food origins

On the other hand there is a steady increasing and sharp uprising trend in expenditures on vegetable origin and animal origins in Fig 2 respectively. Expenditures vary from 2111Rs to 11890Rs in vegetable origin and from 2980Rs to 19296Rs in animal origin.

Stability in expenditures in cereals origin and a steady increasing trend in expenditures on vegetable origin is observed because flour rice and pulses and vegetables and vegetables products are consumes as a major food in all income strata than food on animal origin. So we can conclude that as income increases, people spend more and more on food items especially on meat, eggs, milk and milk products.

3.6 Calories intake from specific food origins

Per capita calories intake/day for specific food origins are determined through quantity consumed (grams)/adult/day and given in table 8. It can be observed that calories intake/adult/day increases in cereals, vegetable and in animal origins and vary from 1416 to1741 (in cereals origin), from 62 to 239 (in vegetable origin) and from 100 to 839 (in animal origin) respectively. It means that low income families fulfill most of their calories from cereals and vegetable origins but on the other hand high income families (last three) fulfill their most of calories from animal origin.

Table 8 Calories from various origins

Strata (k_i)	Cereal origin	Vegetable Origin	Animal origin
1st	1553	127	119
2nd	1741	62	100
3rd	1416	155	365
4th	1607	198	283
5th	1483	183	587
6th	1552	239	839

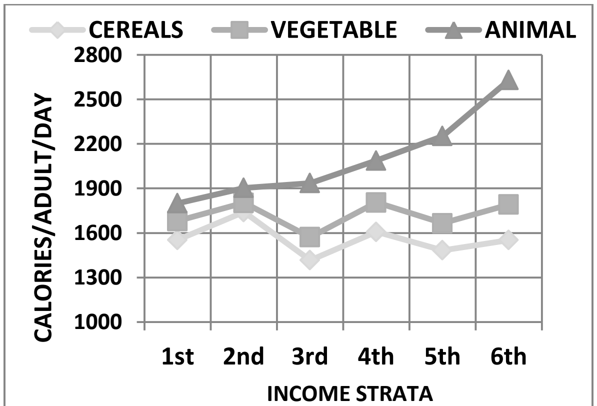

Fig 3 Calories intakeo n various food origins

Calories intake/adult family unit form specific food origins are also shown in **(Fig. 3)** and It is clear that there is an increasing trend in calories intake from vegetable origin but a sharp upward increasing trend exists in calories intake/adult/day in families from animal origin.

3.7 Final Recommendations

It was recommended that ratio type estimation is suitable for estimating true mean food expenditures in 1st, 4th and 5th income strata which are 4987Rs with range of true mean (4101-4132) Rs, 18412Rs with range of true mean (16389-16499) Rs and 27470 Rs with range of true mean (26379-26440) Rs respectively and are found more efficient than mean per unit estimates but approximately as precise as linear regression estimates. It was also observed that linear regression estimation method is appropriate for estimating true mean monthly food expenditures in 2nd and 3rd income strata and it is concluded that estimates of true mean expenditures for 2nd and 3rd income strata are 10844Rs with range of true mean (10814-10876) Rs and 15162Rs with range of true mean (15109-15215) Rs respectively. These estimates are found more efficient as compared to estimates obtained by mean per unit and ratio type methods. But standard error of estimate of true mean in last income stratum by mean per unit method is least as compared to the ratio type and linear regression method and so it can be concluded that estimated true mean food expenditures for last income stratum is 38880Rs with range of true mean (35482-36278)

In analysis of food for specific food origin, there exists approximately stability in expenditures on cereals origin among various strata but there is a steady increasing and sharp uprising trend in expenditures on vegetable origin and animal origins. Per capita calories intake/day for specific food origins are determined and it was observed that low income families fulfill most of their calories from cereals and vegetable origins but high income families (last three strata) fulfill their most of calories from animal origin.

REFERENCES:

Eckler, A. R. (1955) Rotation sampling. *Ann. Math. Stat.*, **(26):** 644-685.

Finkner, A. L. (1950) Method of sampling for estimating commercial peach production in North Carolina. *North Carolina Agri. Exp. Stat Tech. Bull*., 91Pp.

Goswami, S.N. (1994) Differences in consumption pattern for milk and milk products among different groups. *Jour. of dairy Sci.,* 47 **(1):** 62-64.

George, P.S. (1980) The changing pattern of consumer demand for food grains in India. *Indian Jour. Agri. Econ.,* 35 **(1):** 53-68.

Rao, J. N. K. and L. D. Beagle, (1967) A Monte Carlo study of some ratio estimators. *Sankhya,* **(B29)** 47-56.

Richard, H. and W. Left, (1991), Price Theory and Resource Allocation, *McGraw Hill Comp*. London.

Watson, D. J. (1937) The estimation of leaf areas. *Jour. agri. Sci*. **(27):** 474.