A STUDY OF EXPENDITURE PATTERNS OF CALCUTTA MIDDLE CLASS FAMILY BUDGETS

By S. JANARDHAN POTI

Statistical Laboratory, Calcutta

The traditional method in family expenditure studies is to start with some classification based on traits fixed arbitrarily from our knowledge of socio-economic conditions such as community, province or some arbitrary economic levels. From such classified data according to immediately noticeable traits attempts are made to calculate average expenditure patterns of each group and to explain the variation of the expenditure patterns due to the basic classification itself or due to causes indirectly involved in the classification itself. In the middle-class budget survey conducted in 1945 by the Indian Statistical Institute data was collected from 1785 family schedules in three independent subsamples. Analysis of the data pertaining to one of the three subsamples on the basis of a classification of families according to total family expenditure reveals that the family expenditure patterns do not appreciably differ from group to group. Table I giving expenditure patterns compiled from data collected from Calcutta diet survey (1945) reveals the truth of the above statement.

TABLE I. EXPENDITURE PATTERNS.

percentage of expenditure on	expenditure groups in rupees.								
	0—50	51—100	101—150	151—200	201—250	251—300	301—400	401—600	600& above
essential food items	60·02	47·56	41·72	39·36	36·30	31·72	31·57	27·17	23·74
non-essential food items	9·00	17·40	20·66	21·08	23·28	25·31	26·14	27·96	28·69
rent & fuel	18·29	16·45	16·45	16·21	14·51	14·51	13·52	13·16	9·51
miscellaneous & clothing	12·69	18·59	21·17	23·35	25·91	28·46	28·77	31·71	38·06
total	100·00	100·00	100·00	100·00	100·00	100·00	100·00	100·00	100·00
average family size	3·27	4·11	4·83	7·14	8·22	8·98	10·36	11·77	14·12

Though the figures show that the patterns consistently vary with total expenditure an examination of the individual family variation reveals that only a small fraction of it is ascribable to total expenditure. There are other causes of variation such as family size, family composition, community, social status etc. which are also equally effective in causing the variation of expenditure patterns. The average family size increases consistently with the expenditure level as shown by the same table. In fact one may state that the classification adopted here does not separate the families into distinct types with respect to their expenditure patterns. Hence any conclusions based on the above type of analysis may not reveal the actual state of affairs. In fact cost of living indices for 1945 for all the groups calculated with 1939 as the base year were found to lie between 280 and 290 though one should expect the cost of living index should be high for high expenditure levels where proportionately more is spent on items (like fish, meat etc.) the prices of which have arisen by 5 or 6 times that of the pre-war price.

Therefore, an attempt has been made in this paper to classify families according to expenditure pattern, and then to investigate further in what respects families possessing the same pattern agree between themselves and in what manner they contrast with families possessing a different pattern. *This method of tackling the problem may help us to classify families into distinct types by means of some common traits.* If no such common traits are found as may happen sometimes when manifold classifications are required to explain the variations of expenditure pattern then we may classify the families according to expenditure pattern itself and study the effect of several factors on each group separately at different times. This is a better method of study of the economic conditions of families at different intervals of time than the one usually adopted which will give a distorted picture of the existing conditions.

The expenditure pattern studied here consists of sets of four variates which are (a) proportion of expenditure on essential food items (by essential food items is meant cereals, pulses, vegetables, spices and gur) to total expenditure (b) proportion of expenditure on non-essential food items (all food items other than included in (a)) to total expenditure (c) proportion of expenditure on rent and fuel to total expenditure (d) proportion of expenditure on miscellaneous and clothing to total expenditure. This effects a classification of all items into four homogeneous classes with price relatives of items in each class close together. As all the items of expenditures are covered up by (a), (b), (c), (d) it is not necessary to consider the four variates as they will add up to one. It is sufficient, therefore, to deal with any three of them only in our statistical analysis.

Since the statistical analysis with a multiplicity of variates is complicated we shall try to search for a function of these variates which can explain a major portion of the variance of the sets of variates. This function we shall call as the most discriminating function and we shall replace the original set of three variates which represented the family pattern by this function.

Each family can be represented by a point in three dimensions with coordinates p_1, p_2, p_3 representing the proportional expenditures on (a), (b) and (c). The variance of such points can be resolved into three components by taking any three mutually orthogonal axes. We shall choose that system of axes for which the sum of squares of the projections of the point vectors on each of the axes is stationary. If the variance of the projections on an axis constitutes a major component of the total variance then the original points may be replaced by their projections on this line. Having obtained the equation of the line we shall get the distance of each projected point from the origin and thus replace the set of variates (p_1, p_2, p_3) by this one variate, namely the distance of the projected point on this line from the origin.

The mathematical derivation of the linear components described above is given below. Let $l_1 p_1+l_2 p_2+l_3 p_3=t$ be one of the components required, then variance of t is

$$\sum_{i=1}^{3}\sum_{j=1}^{3} l_i l_j v_{ij} = Q \tag{1}$$

where v_{ii} is the variance of p_i and v_{ij} is the covariance between p_i and p_j. Also since the sum of the variances of the three components is equal to the total variances of the set p_1, p_2, p_3 we must have

$$l_1^2+l_2^2+l_3^2-1=v=0 \quad .. \quad (2)$$

So then our object is to maximise Q with respect to l_1, l_2 and l_3 subject to the restriction (2). Using a Lagrangian multiplier we obtain l_1, l_2 and l_3 as solutions of (2) and (3)

$$\frac{\partial Q}{\partial l_i}-k\frac{\partial v}{\partial l_i}=0 \;\Big\}\; i=1, 2, 3 \quad .. \quad (3)$$

On simplifying (3) we get

$$\left.\begin{array}{l} (V_{11}-k)l_1+V_{12}l_2+V_{13}l_3 = 0 \\ V_{21}l_1+(V_{22}-k)l_2+V_{23}l_3 = 0 \\ V_{31}l_1+V_{32}l_2+(V_{33}-k)l_3 = 0 \end{array}\right\} \quad .. \quad (4)$$

If solutions other than $l_1=l_2=l_3=0$ exist the determinant of the set of equations (4) must vanish. This gives

$$\begin{vmatrix} (V_{11}-k) & V_{12} & V_{13} \\ V_{21} & (V_{22}-k) & V_{23} \\ V_{31} & V_{32} & (V_{33}-k) \end{vmatrix} = 0 \tag{5}$$

which is a cubic equation in k called the characteristic equation and it is a well known result that such equations have all the roots real and positive. Let k_1, k_2 and k_3 represent the three roots of (5). Substituting any one of them in (4) and multiplying 1st member of (4) by l_1, 2nd by l_2 and 3rd by l_3 and adding up we get $Q-k(l_1^2+l_2^2+l_3^2)=0$ i.e. $Q=k$. Hence each of these roots stand for the variance of each one of the components possessing the properties described above. Substituting each one of the roots k_1, k_2 and k_3 successively in the first two members of (4) and using (2) we determine the corresponding l's by solving in the usual manner. Let us denote the first, second, and third components thus obtained as $l_{i1}p_1+l_{i2}p_2+l_{i3}p_3$, $i=1, 2$ and 3 respectively. It is also easy to see that these components are orthogonal to each other i.e. $\Sigma\, l_{ik}l_{jk}=\delta_{ij}$ where $\delta_{ij}=0$ if $i\neq j$ and one if $i=j$.

One of the three subsamples collected in Calcutta diet survey consisting of 625 family schedules is taken up for analysis. Since 128 family schedules belonged to house owning families whose housing expenses are naturally very low were omitted in our analysis and the results of the analysis on the remaining 497 family schedules are given below. Division of the sums of squares or sums of products by total sample size is avoided to facilitate computational work. The characteristic equation derived from the covariance matrix is

$$\begin{vmatrix} (8{\cdot}4554-k) & -4{\cdot}2233 & {\cdot}0963 \\ -4{\cdot}2233 & (2{\cdot}7997-k) & -1{\cdot}0383 \\ {\cdot}0963 & -1{\cdot}0383 & (1{\cdot}4905-k) \end{vmatrix} = 0$$

which is same as $\quad k^3-12{\cdot}7267k^2+21{\cdot}5217k-{\cdot}4018=0$

The roots of this equation are $k_1=10{\cdot}7461$, $k_2=1{\cdot}9807$, $k_3=0{\cdot}0189$. Total variance is equal to $10{\cdot}7461+1{\cdot}9807+0{\cdot}0189=12{\cdot}7267$. The contribution by the first component to the total variance of the set is 84·4% and that by other two components together is only 15·6% of the total variance. Hence replacing the family pattern (p_1, p_2, p_3) by a single component as the first one we are leaving only about 16% of the total variation of the expenditure pattern unaccounted and this cannot vitiate the results of our statistical analysis. The three components as well as their variance contributions are given below.

components	direction cosines of the axes l_1	l_2	l_3	contribution to variance
1	·8780	−·4748	·0621	84·4%
2	·2933	·4302	−·8537	15·5%
3	·3785	·7678	·5170	0·1%

Also it can easily be verified that these components are orthogonal to each other.

Taking the first component which explains 84·4% of the total variance to represent the family pattern we find that with increase of family size this component increases but with increase of total expenditure it decreases as is shown by Table 2 giving mean values of the principal component of expenditure pattern by total expenditure groups and family sizes.

TABLE 2. VARIATION OF X (PRINCIPAL COMPONENT) WITH FAMILY SIZE AND TOTAL EXPENDITURE.

family size	expenditure group in rupees 0—50	51—100	101—150	151—200	201—250	251—300	301—400	401—400
2	·41	·30	·10	·02	—	·01	—	·01
3	·52	·30	·16	·08	·06	—	—	—
4	·00	·34	·23	·15	·20	·12	..	·06
5	—	·39	·29	·21	·26	·18	·03	

Except in the case of the groups 151-200 and 201-250 in family sizes 4 and 5 we find that variations are consistent with the statement made above. Such discrepancies as observed are due to samples being too small in those groups.

Let us next study the variation in expenditure pattern due to total expenditure by family sizes. Table 3 below gives the regression equation of x, the principal component on t, the total expenditure for each family size and the contribution of variance due to total expenditure.

TABLE 3. REGRESSION OF THE PRINCIPAL COMPONENT OF EXPENDITURE PATTERN ON TOTAL EXPENDITURE.

family size	no. of families	regression equation	variance of x	variance of x due to t	percentage variation due to t	correlation between x and t
(1)	(2)	(3)	(4)	(5)	(6)	(7)
2	34	$(x-\cdot 2603)=-\cdot 0013(t-93\cdot 59)$	·0480	·3715	39·15	−·0257
3	43	$(x-\cdot 2723)=-\cdot 0029(t-101\cdot 67)$	1·4133	·9732	68·86	−·8311
4	70	$(x-\cdot 2602)=-\cdot 0011(t-130\cdot 94)$	1·3842	·3615	26·23	−·6019
5	72	$(x-\cdot 2092)=-\cdot 0015(t-131\cdot 93)$	1·0540	·5575	52·86	−·7271
6	60	$(x-\cdot 2778)=-\cdot 0018(t-157\cdot 06)$	1·2260	·8084	65·80	−·8117
7	46	$(x-\cdot 2559)=-\cdot 0010(t-189\cdot 99)$	·6974	·2867	41·11	−·6412
8	46	$(x-\cdot 2417)=-\cdot 0012(t-220\cdot 47)$	·4311	·2490	57·76	−·7600
9	35	$(x-\cdot 2090)=-\cdot 0004(t-271\cdot 33)$	1·2219	·2169	17·75	−·4213
10	32	$(x-\cdot 2418)=-\cdot 0002(t-323\cdot 72)$	·6330	·1057	16·70	−·4086

It can be seen that even for a given family size the variation brought about by total expenditure is not very appreciable except in the case of families of sizes 3, 5, 6 and 8. If we had analysed the variance irrespective of family size the variance contribution by total expenditure would have been very much smaller as is shown by the following analysis.

Combining families of all sizes we find the correlation between total expenditure and the principal component reduces to −·3290 and the variance contributed by total expenditure to the total variance of the pattern is only 10·82%. This explains why, when classified according to total family expenditure, average family patterns showed little variation between expenditure groups.

Similar analysis has not been done on families whose size exceeds 10 since samples are very small in those cases. The variation from one family size to another cannot be given a functional form (as in the above case we had given a linear fit) since families of a given size belong to a special type distinct from others with respect to composition. As family composition determines to a very larger extent the variation in pattern the variation due to size gets distorted and hence the variation due to family size and total expenditure have not been estimated by applying partial regression formulae to the data.

SUMMARY.

(1) Since the principal component $l_{11}p_1+l_{12}p_2+l_{13}p_3$ explains a major portion of the variance, there exists a single economic factor which influences to a great extent the expenditure pattern. This component which increases with family size and decreases with total expenditure can be used to classify families according to poverty levels for budgetary studies.

(2) Though family size and total expenditure together explain in many cases a large part of the variation in expenditure patterns we find that there exist other factors like family composition etc., which too in no small measure determine the poverty level mentioned above.

(3) Since classification by expenditure levels alone cannot divide the families into distinct types with respect to expenditure pattern, any study of changes of pattern of families in a given expenditure level after the lapse of a period leads to erroneous judgement. Therefore, it would be better to group the families according to poverty levels determined by the principal component derived in this article and study how they change their groups and assess improvement or deterioration.

Paper received : 10 May, 1946.