ESTIMATING INCOME OF LOW INCOME HOUSEHOLDS IN A PHILIPPINE CITY USING FOOD EXPENDITURES AS PREDICTORS*

by

SOENARDI SOSROOETOYO

1. PURPOSE OF THE STUDY

In order to collect data on income and expenditure among low-income households, many countries are conducting household surveys using various methods. Experience shows that among low-income households it is usually very difficult to compare the information gathered on income and expenditure since the latter tends to be larger than the former. However, expenditure data are more easily obtained than information on total household income. It seems desirable therefore, to estimate income using the expnditure data as predicators.

The main purposes of this study are:

(i) to demonstrate the use of multiple regression in estimating income using household expenditures on food as predictors:

(ii) to determine, in particular, which of the food expenditure components can be sufficiently used as predictors of income; and lastly

(iii) to determine the weighting pattern of food expenditure for low-income household in Manila.

The data used in this study were obtained from the household expenditure survey conducted in a certain city in the Philippines throughout the year 1961.

*This paper has been derived in part from a project report submitted to the Statistical Center, U. P. by the author in partial fulfillment of the requirements for the Certificate in Statistics. The editors wish to acknowledge the kindness of Dr. E. T. Virata, Acting Director of the Center, in permitting publication of the author's results.

2. SURVEY DESIGN

Introduction

Owing to the rapid changes in the living patterns of urban household units, a household expenditure survey was conducted in 1961. In this study we shall limit the data used in the regression analysis to expenditures on food items alone, whereas, in the computation of "weights", we shall include the expenditures on alcoholic beverages, tobacco and utilities (see Appendix A).

The Sampling Design

The interval in the systematic selection. The sampling design in the household expenditure survey was based on a sub-sample of 8,092 households which reported an income of ₱2,400 and below per annum selected from a total of 12,380 households listed in the locality in October 1960. The 8,092 households represent 65.5% of the total listed households.

A total of 437 households were enumerated during this month in the survey. Assuming that the income ratio of the listed low-income bracket (₱2,400 per annum and below) to the total listed households is the same as that in the enumerated households, then 65.5% of 437 enumerated households gives 286, which is the expected number of low-income among those enumerated.

Since there are 4 weeks in a month, the 286 households were then splt into approximately equal groups, giving the expected number of 71 per group per week. Each group was interviewed four times during the year and since there are 52 weeks in a year, there were 13 different groups. The expected number of total households in 13 groups was therefore 13 times 71 or 923. Dividing the 8,092 listed low-income households by 923 we arrived at approximately nine as the interval in the systematic selection.

The method of systematic selection. The listed households of the low income bracket (8,092) were divided geographically into 20 strata. Two-stage sampling was used in the selection of the households. Firstly, by simple random sampling, 5 precincts in each stratum were selected giving 100 sample precincts. Secondly, in each sample precinct, the households were drawn systemtically with independent random starts. By this procedure, a total of 896 households were drawn as shown below.

Strata	Precinct					Total
	I	II	III	IV	V	
1	n_{11}	n_{12}	n_{13}	n_{14}	n_{15}	$n_{1.}$
2	n_{21}	n_{22}	n_{23}	n_{24}	n_{25}	$n_{2.}$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
20	$n_{20,1}$	$n_{20,2}$	$n_{20,3}$	$n_{20,4}$	$n_{20,5}$	$n_{20,1}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$	$n_{.5}$	$n_{..} = 896$

That is,

$$n_{..} = \sum_i \sum_j n_{ij} = \sum_j n_{.j} = 896 .$$

These 896 households were again separated systematically into 13 groups, the first 12 groups containing 69 households each and the last group containing 68 households.

The Survey

Each group of households was interviewed 4 times during 1961. The phasing is such that Group 1 was interviewed during weeks 1, 14, 27, 40, Group 2 was interviewed during weeks 2, 15, 28, 41, ... and finally Group 13 was interviewed during weeks 13, 26, 39, 52.

Expenditures on 29 items were collected for the week. Data on monthly income (average income) were obtained during the interviews. However, since only the data of the first 13 weeks have been processed in detail and since only totals of the different items are available for each of the 52 weeks, our analyses in the succeeding sections are limited in nature.

3. THE MODEL

The model which will be discussed in this section is constructed by using the average weekly expenditures on food as the independent variables and average monthly income as the dependent variable. About 82% of the total average expenditures are included in this construction. The average size of household having income between ₱164 and ₱200 per month is 6 persons. Thus the estimated monthly income which we will have from the model is the estimate of monthly income of a household of 6 members.

Four important food groups from which the average monthly income (Y'_j) per household in the jth week may be predicted are:

X'_{1j} = average weekly expenditures per household in the jth week on cereals and related items which consist of rice, bread and other cereal products (see item I of Appendix A).

X'_{2j} = average weekly expenditures per household in the jth week on meats, fish, poultry, milk products and eggs (items II and III of Appendix A).

X'_{3j} = average weekly expenditures per household in the jth week on vegetables and fruits (item IV of Appendix A).

X'_{4j} = average weekly expenditures per household in the jth week on miscellaneous foods (item V on Appendix A).

The model is as follows.

$$E(Y'_j \mid X_{1j}, \ldots, X_{4j}) = \alpha' \, X'^{\beta_1}_{1j} \, X'^{\beta_2}_{2j} \, X'^{\beta_3}_{3j} \, X'^{\beta_4}_{4j} ,$$

$$j = 1, 2, \ldots, 52 ,$$

where α', β_1, β_2, β_3 and β_4 are the parameters to be estimated. Alternatively, by transforming the variables $Y_j = \log Y'_j$, $X_{1j} = \log X'_{1j}$, $X_{2j} = \log X'_{2j}$, $X_{3j} = \log X'_{3j}$, and $X_{4j} = \log X'_{4j}$, the model may be written as

$$E(Y_j \mid X_{1j}, \ldots, X_{4j}) = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \beta_3 X_{3j} + \beta_4 X_{4j} ,$$

where $\alpha = \log \alpha'$. Using the last model, the parameters are estimated by the method of least squares. We denote these least squares estimates by $\hat{\alpha} = a$ and $\hat{\beta}_i = b_i$, $i = 1, 2, 3, 4$.

The choice of the multiplicative model is suggested by the fact that income Y, in general, ranges from 0 to ∞. However, the logarithm of the ratio $(Y / \pi X_i^{\beta_i})$ has a range of $-\infty$ to ∞. Using the last model has been fairly well supported by our findings; namely, that the sample moment ratios appear those of a normal distribution.

4. ESTIMATION AND SIGNIFICANCE TESTS

Estimation of the Parameters

Using the model in the previous subsection, we find the mean values of the following

$$\bar{Y} = 2.24880, \quad \bar{X}_1 = 0.92653, \quad \bar{X}_2 = 0.94577,$$
$$\bar{X}_3 = 0.25306, \quad \bar{X}_4 = 0.39553.$$

The corrected sums of products of Y_j with X_{1j}, X_{2j}, X_{3j}, and X_{4j} are, respectively,

$$Q_1 = \sum_{j=1}^{52} y_j x_{1j} = 0.01043, \quad Q_2 = \sum_{j=1}^{52} y_j x_{2j} = 0.01801,$$

$$Q_3 = \sum_{j=1}^{52} y_j x_{3j} = 0.02261, \quad Q_4 = \sum_{j=1}^{52} y_j x_{4j} = 0.01764.$$

Similarly, the corrected sums of products matrix $\underset{\sim}{S} = (s_{kt})$, $s_{kt} = \sum_{j=1}^{52} x_{kj} x_{tj}$, k, t = 1, 2, 3, 4, for X_{1j}, X_{2j}, X_{3j}, X_{4j} is

$$S = \begin{bmatrix} 0.10709 & 0.05140 & 0.04767 & 0.06708 \\ 0.05140 & 0.22339 & 0.24455 & 0.14438 \\ 0.04767 & 0.24455 & 0.44399 & 0.21392 \\ 0.06708 & 0.14438 & 0.21392 & 0.18872 \end{bmatrix}.$$

The reciprocal of the matrix $\underset{\sim}{S}$ denoted by $\underset{\sim}{S}^{-1} = (s^{kt})$ is obtained by Doolittle's Method,

$$\underset{\sim}{S}^{-1} = \begin{pmatrix} 12.88354 & -1.83623 & 2.55082 & -6.06604 \\ -1.83603 & 12.68997 & -5.35258 & -2.98851 \\ 2.55076 & -5.35251 & 7.47035 & -5.27961 \\ -6.06616 & -2.98854 & -5.27957 & 15.72598 \end{pmatrix};$$

Hence, the estimates of the parameters are:

$$b_1 = s^{11}Q_1 + s^{12}Q_2 + s^{13}Q_3 + s^{14}Q_4 = .05197$$

$$b_2 = s^{21}Q_1 + s^{22}Q_2 + s^{23}Q_3 + s^{24}Q_4 = .03566$$

$$b_3 = s^{31}Q_1 + s^{32}Q_2 + s^{33}Q_3 + s^{34}Q_4 = .00598$$

$$b_4 = s^{41}Q_1 + s^{42}Q_2 + s^{43}Q_3 + s^{44}Q_4 = .04094$$

$$a = \bar{y} - b_1\bar{X}_1 - b_2\bar{X}_2 - b_3\bar{X}_3 - b_4\bar{X}_4 = 2.14922 = \hat{\alpha}.$$

Or

$$\alpha' = 141 .$$

Thus, the predicting equation for the average monthly income per household in logarithmic form is

$$\hat{Y}_j = 2.14922 + .05197\ X_{1j} + .03566\ X_{2j} + .00598\ X_{3j} + .04094\ X_{4j}$$

or, using the ordinary multiplicative form we have

$$\hat{Y}_j' = 141\ X_{1j}'^{.05197}\ X_{2j}'^{.03566}\ X_{3j}'^{.00598}\ X_{4j}'^{.04094}$$

Test of Hypothesis that $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

$$\sum_{j=1}^{52} (Y_j - \alpha - \beta_1 X_{1j} - \beta_2 X_{2j} - \beta_3 X_{3j} - \beta_4 X_{4j})^2$$

which is

$$\left(\sum_{j=1}^{52} Y_j^2 - 52\bar{Y}^2\right) - b_1Q_1 - b_2Q_2 - b_3Q_3 - b_4Q_4$$

$$= .00793 - .00204$$

$$= .00589 .$$

If the hypothesis $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ is true then the minimum value of $\sum_{j=1}^{52} (Y_j - \alpha)^2$ is $\sum_{j=1}^{52} Y_j^2 - 52\bar{Y}^2 = .00793$, which is the total sum of squares with $(n-1) = 51$ degrees of freedom. The reduction in the above sum of squares due to regression is .00204 with 4 degrees of freedom. Hence, the analysis of variance table for the null hypothesis that $\beta_i = 0$, $i = 1, \ldots, 4$ is:

Sources of Variation	Degrees of freedom	Sum of Squares	Mean Sum of Squares	F
Due to regression	4	.00204	.000510	4.08
Residuals	47	.00589	.000125	
Total	51	.00793		

The variance ratio 4.08 with 4 and 47 degrees of freedom is significant at the 1% level, which shows that the variables considered jointly above are useful in prediction.

Test of Hypothesis that $\beta_i = \beta$, **all i**

More conveniently, this is the test to see if the independent variables have equal predictive values. The minimum value of

$$\sum_{j=1}^{52}\left[Y_j - \alpha - \beta(X_{1j} + X_{2j} + X_{3j} + X_{4j})\right]^2 \ldots$$

has to be determined. The normal equation giving the estimate of β is

$$\left[s_{11} + s_{22} + s_{33} + s_{44} + 2(s_{12} + s_{13} + s_{14} + s_{23} + s_{24} + s_{34})\right] b = Q,$$

where $Q = Q_1 + Q_2 + Q_3 + Q_4$.

Thus,

$$2.50119\ b = .06869$$
$$b = .02746$$
$$bQ = .00189\ .$$

The minimum value with $(n-2) = 50$ degrees of freedom is

$$\left[\left(\sum_{j=1}^{52} Y_j^2 - 52\bar{Y}^2\right) - bQ\right] = .00793 - .00189$$
$$= .00604\ .$$

The analysis of variance table for equal predictions value of the different components is,

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F
Deviation from equality	3	.00015	.000050	.04
Residuals	47	.00589	.000125	
Total	50	.00604		

The variance ratio $F = .04$ with 3 and 47 degrees of freedom is not significant at the 5% level, so there is no evidence from the data to conclude that β_1, β_2, β_3 and β_4 are different. The difference if any are very small and can be attributed to chance. Hence, we can use the predictive formula

$$\hat{Y}_{.}'' = 151.2\, (X'_{1j}\, X'_{2j}\, X'_{3j}\, X'_{4j})^{.02746},$$

where 151.2 is obtained from

$$a' = \hat{\alpha} = \bar{Y} - b(\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \bar{X}_4) = 2.17958$$

or $\alpha'' = 151.2$, by assuming $\beta_1 = \beta_2 = \beta_3 = \beta_4$

Test of Hypothesis that $\beta_3 = 0$

From section 3 we noticed that the coefficient $\hat{\beta}_3$ is very small compared with the other coefficients (β_1, β_2 and β_4). Perhaps its is better to eliminate β_3 to reduce the number of predictors rather than use four independent variables. This is equivalent to testing the hypothesis that $\beta_3 = 0$.

The estimate $\beta_3 = b_3 = .00598$ has the variance $s^{33}\sigma^2$. The variance ratio with 1 and 47 degrees of freedom is

$$\frac{(b_3)^2}{s^{33}\sigma^2} = \frac{(.00598)^2}{7.47035} \cdot \frac{1}{.000125}$$

$$= \frac{.000036}{.000934}$$

$$= .0385,$$

where for σ^2 the estimate based on 47 degrees of freedom is used. This is not significant at the 5% level. Thus, we can accept the hypothesis that β_3 is zero. Then the best estimates of b_1, b_2 and b_4 have to be revised, using the model

$$E(Y_j''' \mid X_{1j}, X_{2j}, X_{4j}) = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \beta_4 X_{4j}\,,$$

$$j = 1, 2, \ldots, 52\,.$$

For this purpose we can reduce the original $\underset{\sim}{S}^{-1}$ matrix,

$$S^{-1} = \begin{bmatrix} 12.88354 & -1.83623 & 2.55082 & -6.06604 \\ -1.83603 & 12.68997 & -5.35258 & -2.98851 \\ 2.55076 & -5.35251 & 7.47035 & -5.27961 \\ -6.06616 & -2.98854 & -5.27957 & 15.72598 \end{bmatrix}$$

by the pivotal condensation [8], starting from the element at the intersection of the third row and third column. That is using s^{33} as the pivot, we have

$$\begin{bmatrix} s^{11} - \dfrac{s^{31}s^{13}}{s^{33}} & s^{12} - \dfrac{s^{32}s^{13}}{s^{33}} & s^{14} - \dfrac{s^{34}s^{13}}{s^{33}} \\[2ex] s^{21} - \dfrac{s^{31}s^{23}}{s^{33}} & s^{22} - \dfrac{s^{32}s^{23}}{s^{33}} & s^{24} - \dfrac{s^{34}s^{23}}{s^{33}} \\[2ex] s^{41} - \dfrac{s^{31}s^{43}}{s^{33}} & s^{42} - \dfrac{s^{32}s^{43}}{s^{33}} & s^{44} - \dfrac{s^{34}s^{43}}{s^{33}} \end{bmatrix} =$$

$$\begin{bmatrix} 12.01256 & -0.00857 & -4.26327 \\ -0.00839 & 8.85484 & -6.77140 \\ -4.26344 & -6.77135 & 11.99469 \end{bmatrix}$$

The new coefficients are

$$b_1' = 12.01256\ Q_1 - .00857\ Q_2 - 4.26327\ Q_4 = .04993$$

$$b_2' = .00839\ Q_1 + 8.85484\ Q_2 - 6.77140\ Q_4 = .03994$$

$$b_4' = -4.26344\ Q_1 - 6.77135\ Q_2 + 11.99469\ Q_4 = .04393$$

$$a'' = \bar{y} - b_1'\bar{X}_1 - b_2'\bar{X}_2 - b_4'\bar{X}_4 = 2.14739$$

$$\alpha''' = 140.4\ .$$

Thus, after eliminating β_3, the predicting equation becomes

$$Y''' = 2.14739 + .04993\ X_{1j} + .03994\ X_{2j} + .04393\ X_{4j}$$

or

$$Y_j'''' = 140.4\ X_{1j}'^{.04993}\ X_{2j}'^{.03994}\ X_{4j}'^{.04393}$$

where $Y_j''' = \log Y_j''''$ and $X_{ij} = \log X_{ij}'$ as before.
The residual sum of squares is

$$\sum_{j=1}^{52} (Y_j - \bar{Y})^2 - b_1'Q_1 - b_2'Q_2 - b_4'Q_4 = .00590\ .$$

If to the residual sum of squares .00589 (before we reduce the matrix $\underset{\sim}{S}^{-1}$) we add the sum of squares due to $b_3 = (b_3)^2 / s^{33} = .000005$, the result is $.005895 \approx .00590$. Thus, the calculation of b_1, b_2 and b_4 is correct.

5. COMPUTATION OF WEIGHTS

Before we compute the weights, we construct first the analysis of variance table to see whether there are significant differences among the expenditures of the household groups.

The analysis of variance table is constructed (using Appendix B) as follows

a) The correction factor

$$C = (26{,}539.89)^2/117$$
$$= 704{,}365{,}761.2121/117$$
$$= 6{,}020{,}220.1813$$

b) Total sum of squares $= (640.50)^2 + (773.15)^2 + \ldots + (57.55)^2 - C$

$$= 10{,}906{,}341.2211 - 6{,}020{,}220.1813$$
$$= 4{,}886{,}121.0398$$

c) Household's group sum of squares

$$= \frac{(2{,}357.35)^2 + (2{,}578.72)^2 + \ldots + (1{,}765.87)^2}{9} - C$$
$$= (54{,}914{,}495.5771/9) - 6{,}020{,}220.1813$$
$$= 6{,}101{,}610.6197 - 6{,}020{,}220.1813$$
$$= 81{,}390.4384$$

d) Commodities sum of squares

$$= \frac{(7{,}639.41)^2 + (7{,}633.40)^2 + \ldots + (1{,}470.40)^2}{13} - C$$
$$= (138{,}671{,}123.4957/13) - 6{,}020{,}220.1813$$
$$= 10{,}667.009.5997 - 6{,}020{,}220.1813 = 4{,}646{,}789.4184$$

e) Errors sum of squares = (b) − (c) − (d) = 157,941.1830.

Analysis of Variance

Sources of Variation	Freedom Degrees of	Sum of Squares	Mean Sum of Squares	F
Household groups	12	81,390.4384	6,782.5365	4.12
Commodities	8	4,646,789.4184	580,848.6772	
Errors	96	157,941.1830	1,645.2206	
Totals	116	4,886,121.0398		

The F-ratio of 4.12 with 12 and 96 degrees of freedom is significant even at 1% level. Thus, there appears to be significant differences among the expenditures of the household group (weeks) in the first 13 weeks of 1961 (assuming that the expenditures of each household during these 13 weeks are homogeneous). By inspection of the analysis variance table above, we see that the expenditures by commodity groups are even more highly significant.

There is some reason to have separate "weights" for the expendtures on the commodities by weeks from the results obtained in the analysis of variance. If data for 52 weeks are available it is also likely that expenditures by months or quarters will also be found significant. Hence, in practice, expenditures are adjusted for seasonal variations. Our present data, however, does not permit us to remove seasonal patterns.

On the other hand, expenditures on the 9 groups of commodities are decidedly significant. The average percentage expenditures in some cases, however, are influenced by the high variability of expenditures by items within the same commodity group during the first 13 weeks. A more stable pattern of expenditure may be obtained from the median in these cases.

Consider Appendix D. It seems that during the first 13 weeks, most of the average percentage expenditures are more or less equal to their medians. Hence the average percentage expenditures may be used as weights. But within the same commodity classifications it appears better to use the medians rather than the averages in some cases. Take as an example commodity No. 7, i.e., **other fresh meats.** Ten out of 13 average percentage expenditures (or about 77%) are less than the average of their percentage expenditures. In other words, the distribution of average weekly percentage expenditure for this item is highly skewed. If we assume that this is the type of distribution prevailing during the entire year, projecting expenditures for a week to the whole year would be subject to considerable sampling fluctuation.

To keep the total percentage expenditure of the whole commodities equal to 100%, we use the procedure shown in Tables A, B, and C.

TABLE A. WEIGHTS FOR MEATS, FISH, ETC.

Commodity (1)	Average (2)	Average and Median (3)	Weights (4)
Meats, Fish, etc.	28.71	28.46	28.71
4. Pork, fresh	7.16	7.16	7.22
5. Pork, smoked or cured	.47	(.34)	.34
6. Beef	2.94	2.94	2.97
7. Other fresh meats	.28	(.16)	.16
8. Canned meats	1.39	1.39	1.40
9. Fish and sea foods	13.14	13.14	13.26
10. Canned fish	1.27	1.27	1.28
11. Poultry	2.06	2.06	2.08

TABLE B. WEIGHTS FOR MISCELLANEOUS FOODS

Commodity (1)	Average (2)	Average and Median (3)	Weights (4)
Miscellaneous foods	9.55	9.46	9.55
18. Sugar	2.11	2.11	2.13
19. Other sweets	.32	(.23)	.22
20. Coffee	2.22	2.22	2.24
21. Tea and chocolate	.25	.25	.25
22. Carbonated drinks	1.05	1.05	1.06
23. Fats and oils	1.92	1.92	1.94
24. Other miscellaneous foods	1.68	1.68	1.70

TABLE C. WEIGHTS FOR ALCOHOLIC BEVERAGES AND TOBACCO

Commodity (1)	Average (2)	Average and Median (3)	Weights (4)
Alcoholic beverages and tobacco	8.36	8.11	8.36
25. Tobacco products	7.23	7.23	7.45
26. Alcoholic beverages	1.13	(.88)	.91

The weights of commodities obtained from the median are enclosed in parentheses ( ). They are pork (smoked or cured) and other fresh meats (see Table A), other sweets (see Table B), and alcoholic beverages (see Table C). Column 2 shows the average percentage expenditures, while column 3 shows both the averages and the medians. Since we are to keep the total percentage expenditure of the commodity category equal to 100% in column 4, we make adjustments among the weights within the category. Each value in Column 3 is multiplied by the ratio of total percentage in column 2 by the total percentage in column 3. The results are recorded in column 4. Thus, in Table A, each value in column 3 is multiplied by 28.71/28.46.

Of course, the results will be much better if we used the whole 52 weeks data, not only those of the first 13 weeks. But since one main purpose of this project study is to introduce a method of finding weights, we have used the data of the first 13 weeks only. By the same procedure the weights of the other commodities may also be computed.

The results of the computation of weghts are shown in Table D.

TABLE D. THE WEIGHTS OF FOOD ITEMS DERIVED FROM THE FIRST 13 WEEKS

Item	Median	Average	Weight
I. Cereals	28.79	28.82	28.82
1. Rice	20.98	20.94	20.94
2. Bread	6.52	6.46	6.46
3. Other cereal products	1.31	1.42	1.42
II. Meat, Fish, etc.	28.71	28.71	28.71
4. Pork, fresh	7.13	7.16	7.22
5. Pork, cured or smoked	.34	.47	.34
6. Beef	2.82	2.94	2.97
7. Other fresh meats	.16	.28	.16
8. Canned meats	1.37	1.39	1.40
9. Fish and sea foods	13.76	13.14	13.26
10. Canned fish	1.17	1.27	1.28
11. Poultry	2.01	2.06	2.08
III. Milk products and eggs	7.00	6.97	6.97
12. Milk, cream and ice cream	4.35	3.93	3.93
13. Butter and cheese	1.04	1.08	1.08
14. Eggs	2.02	1.96	1.96
IV. Vegetables and Fruits	1.08	8.00	8.00
15. Vegetables	5.43	5.20	5.20
16. Fruits and nuts	2.62	2.54	2.54
17. Canned and dried fruits	.23	.26	.26
V. Miscellaneous foods	9.27	9.55	9.55
18. Sugar	2.13	2.11	2.13
19. Other sweets	.23	.32	.23
20. Coffee	2.35	2.22	2.24
21. Tea and chocolate	.23	.25	.25
22. Carbonated drinks	1.00	1.05	1.06
23. Fats and oils	1.94	1.92	1.94
24. Other miscellaneous foods	1.70	1.68	1.70
VI. Alcoholic beverages and tobacco	8.36	8.36	8.36
25. Tobacco products	6.75	7.23	7.45
26. Alcoholic beverages	.88	1.13	.91
VII. 27. Utilities	2.23	2.28	2.28
VIII. 28 Other and unclassifiable foods	1.59	1.75	1.75
IX. 29. Food consumed away from home	5.02	5.56	5.56

Note: The classification of the foodstuffs was adopted from the International Labour Office recommendation (see Reference 6).

6. INTERPRETATION OF RESULTS AND SOME SUGGESTIONS

Interpretation and Suggestion Concerning the Sampling Design

Intuitively the design is reasonably sound. The stratification likely reduces the variance of error, if the assumption of homogeneous units within the stratum is reasonable. The systematic selection may bring a lot of simplification of the sample taking and may lead to a more precise estimate, if the randomnization is done before the samples were taken in such a way that the intraclass correlation, p , may reasonably be assumed to be zero. Otherwise the analysis will be too complicated for a first step sample survey. The two-stage sampling may reduce considerably cost and time and other practical aspect of the survey.

From a practical as well as a theoretical point of view, however the design is likely to be unsound. The sophistication led to several difficulties for the survey staff, especially in providing the necessary information of the estimate. No derived formula appropriate to the estimation of the standard error was given. It is a common mistake of non-statistical research workers in thinking that a design which is intuitively sound can always bring precise estimates.

A scientific sample survey will completely lose its value if necessary information regarding the precision of the estimate cannot be assessed. We may be able to derive an appropriate formula for that design, with considerable complication. Consequently, computational error will probably be introduced, and the inference therefrom may be subject to very doubtful reliability. Indeed, for a statistical worker, reliability of the data has more important practical consideration than the manageable precision (in terms of standard error) of the estimate.

Improvement of the design may be suggested from both practical and theoretical considerations.

Two-stage stratified sampling with probability proportional to size in both stages is suggested with the precinct as the first stage sampling unit and the household as the second stage sampling unit. A smaller number of strata may considerably reduce the labor of analysis. Aside from geographical location, stratification is recommended by taking into consideration the homogeneity feature of the first stage sampling unit (precincts) with respect to the characteristic under study. The size of strata may differ from stratum to another. Hence the suggestion is that precincts should be selected with probability proportional to the number of households.

The second stage units should be selected from the sample precinct, again with probability proportional to the number of households. An appropriate formula may be adapted from reference (2) with the complete account of the suggested method.

Comparison with previous surveys, however, cannot be done, since inspite of our inability to provide a measure of precision in the former design, the new suggested design does not give a complete account of the stratification and selection features. But at least the new design will furnish a more sounder basis and considerable simplfication. Obtaining a measure of precision is a predominant advantage.

So far, we are not talking about the cost of the survey. We assume that the cost available is such that we may manage it in terms of, say, only the sample size and, thereby, the cost analysis may follow.

It might be pointed out that in the systematic selection, certain households with more than ₱2,400 per annum income have been drawn into the sample. As a matter of fact there are approximately 25% of these households. The prediction or estimates of household income could have been better if these households with income of more than ₱2,400 per annum have been replaced by those with income of ₱2,400 per annum or less, so that our sample would conform strictly to our definition of low income household.

Interpretations and Suggestions Concerning the Models

In section 4, we derived three models, which are

$$(1)\quad \hat{Y}' = 141\, X_{1j}'^{.05197}\, X_{2j}'^{.03566}\, X_{3j}'^{.00598}\, X_{4j}'^{.04094},$$

$$(2)\quad \hat{Y}'' = 151.2\, (X_{1j}'\, X_{2j}'\, X_{4j}')^{.02746},$$

$$(3)\quad \hat{Y}_j'''' = 140.4\, X_{1j}'^{.04993}\, X_{2j}'^{.03994}\, X_{4j}'^{.04393}.$$

we can use these three models to estimate the income of a certain household consisting of 6 members. The last model seems much better to use than the other two, since here we only use three independent variables as predictors. The computations involved are easier than in the first model. We will not suggest using the second model, since as can be seen from the first model, the value of β_3 is very different from the other three.

Hence, we can now say that the average weekly expenditures on cereals, meats. fish, poultry, milk products, eggs, and miscellaneous foods (i.e., X_{1j}', X_{2j}', X_{4j}') can be sufficiently used as predictors of monthly income of the low-income households in Manila.

Interpretations and Suggestions Concerning Weights

It has been shown in section 5 that there is a significant difference among the expenditures of household groups. One may expect such a thing to happen since the survey was conducted in different weeks (time). Experience also shows that expenditures are subject to seasonal variation (time). For instance one may expect that the expenditures during November and December will be very different from the other months.

Appendix D shows that the average percentage expenditure among the groups are not too far from their medians. Since almost all of the average percentage expenditure are more or less the same as the median, the weights were constructed using both average percentage expenditure and median.

REFERENCES

(1) Anderson, R. L. and T. A. Bancroft (1952). **Statistical Theory in Research,** McGraw-Hill Book Co., Inc., New York, Toronto, London.

(2) Cochran, William G. (1951). **Sampling Techniques,** John Wiley and Sons, Inc., New York, London.

(3) Croxton, Frederick E. and Dudley P. Cowden (1961). **Practical Business Statistics,** Prentice-Hall, Inc., Maruzen Co., Ltd., Englewood Cliffs, New York, Tokyo, Japan.

(4) Ezekiel, Mordecai and Karl A. Fox (1959). **Methods of Correlation and Regression Analysis,** John Wiley and Sons, Inc., New York, London.

(5) Goulden, Cyril H. (1959). **Methods of Statistical Analysis,** John Wiley and Sons, Inc., Charles E. Tuttle Co., New York, Tokyo.

(6) Internationl Labour Office (1949). **Methods of Living Studies, I.L.O.** Publications, Geneva.

(7) Mood, Alexander McFarlane (1950). **Introduction to the Theory of Statistics,** McGraw-Hill Book Co., Inc., New York, Toronto, London.

(8) Roa, C. Radhakrishna (1952). **Advanced Statistical Methods in Biometric Research,** John Wiley and Sons, Inc., New York, Chapman and Hall, Ltd., London.

(9) Williams, E. Y. (1959). **Regression Analysis,** John Wiley and Sons, Inc., New York.

APPENDIX A

CLASSIFICATION OF FOODSTUFFS

I. Cereals:
  1. Rice
  2. Bread
  3. Other cereal products

II. Meat, fish, etc.:
  4. Pork, fresh
  5. Pork, smoked or cured
  6. Beef
  7. Other fresh meats
  8. Canned meats
  9. Fish and sea foods
  10. Canned fish
  11. Poultry

III. Milk products and eggs:
  12. Milk cream and ice cream
  13. Butter and cheese
  14. Eggs

IV. Vegetables and fruits:
  15. Vegetables
  16. Fruits and nuts
  17. Canned and dried fruits

V. Miscellaneous foods:
  18. Sugar
  19. Other sweets
  20. Coffee
  21. Tea and chocolate
  22. Cabonated drinks
  23. Fats and oils
  24. Others

VI. Alcoholic beverages and Tobacco:
  25. Tobacco products
  26. Alcohloic beverages

VII. 27. Utilities

VIII. 28. Other and unclassifiable foods

IX. 29. Foods consumed away from home

APPENDIX B. WEEKLY EXPENDITURES OF THE THIRTEEN GROUPS OF HOUSEHOLDS IN THE FIRST THIRTEEN WEEKS BY MAJOR ITEM OF COMMODITIES

Week	Cereals I	Meats, Fish, etc. II	Products Milk & eggs III	Vegeta- Fruits bles & IV	Miscel- laneuos foods V	Alcoholic beverages & Tobacco VI	Utilities VII	Other & unclas- sifiable foods VIII	Foods consumed away from home IX	Total
1	640.50	629.30	161.20	205.45	281.85	197.20	62.40	22.35	156.10	2357.35
2	773.15	819.55	145.70	167.65	261.97	218.75	66.75	27.15	98.05	2578.72
3	622.36	729.20	117.75	194.00	195.33	196.30	65.05	59.95	96.05	2275.99
4	582.76	564.38	145.66	168.84	188.15	198.65	54.45	31.35	147.45	2081.69
5	622.90	544.70	142.40	172.85	205.85	160.20	67.15	33.46	150.00	2099.51
6	534.10	522.90	139.30	158.90	200.10	172.65	44.60	33.85	135.85	1942.25
7	515.05	424.55	138.83	150.40	179.55	139.45	40.15	73.30	155.70	1816.98
8	552.25	586.65	158.39	186.70	202.30	150.55	49.55	13.55	85.10	1985.04
9	584.60	618.45	146.40	150.40	187.00	164.95	25.80	20.80	122.60	2021.00
10	576.98	512.85	120.05	164.05	172.10	163.05	32.35	35.80	91.25	1868.48
11	591.26	589.60	159.40	146.10	178.00	226.30	23.40	49.80	89.55	2053.41
12	522.10	529.25	109.50	124.40	153.00	107.95	40.10	20.15	85.15	1691.60
13	521.40	562.02	149.80	129.00	137.05	134.85	39.30	34.90	57.55	1765.87
Total	7639.41	7633.40	1834.38	2119.74	2542.25	2230.85	611.05	456.41	1470.40	26539.89

APPENDIX D. THE PERCENTAGE EXPENDITURE LOWEST TO THE

	1	2	3	4	5
I. Cereals	27.17	27.35	27.50	27.82	27.99
1. Rice	17.54	19.04	19.64	19.78	19.86
2. Bread	5.34	5.69	6.03	6.27	6.34
3. Other cereal products	.58	1.00	1.00	1.08	1.11
II. Meats,Fish, etc.	23.36	25.94	26.69	26.93	27.11
4. Pork, fresh	5.07	5.73	5.84	5.99	6.47
5. Pork, smoked or cured	.16	.16	.20	.22	.23
6. Beef	1.78	2.37	2.44	2.46	2.54
7. Other fresh meats		.03	.05	.11	.11
8. Canned meats	.59	1.14	1.15	1.20	1.27
9. Fish and sea foods	9.88	10.01	10.75	11.14	12.48
10. Canned fish	.86	.98	1.03	1.06	1.15
11. Poultry	1.22	1.29	1.52	1.60	1.92
III. Milk products and Eggs	5.18	5.65	6.42	6.46	6.78
12. Milk, cream and ice cream	1.30	2.67	2.69	3.50	3.89
13. Butter and cheese	.67	.68	.87	.92	.93
14. Eggs	1.18	1.51	1.76	1.84	1.86
IV. Vegetables and Fruits	6.50	7.11	7.31	7.34	7.44
15. Vegetables	3.42	4.55	4.71	4.90	4.90
16. Fruits and nuts	1.63	2.03	2.18	2.29	2.37
17. Canned and dried fruits	.06	.07	.14	.14	.16
V. Miscellaneous foods	7.76	8.58	8.67	9.04	9.21
18. Sugar	1.78	1.90	1.97	1.97	2.00
19. Other sweets	.08	.19	.19	.21	.21
20. Coffee	.75	.86	1.80	2.05	2.20
21. Tea and chocolate	.02	.07	.14	.14	.16
22. Carbonated drinks	.43	.48	.62	.69	.74
23. Fats and oils	1.33	1.72	1.74	1.78	1.83
24. Other miscellaneous foods	1.30	1.50	1.51	1.55	1.55
VI. Alcoholic beverages and Tobacco	6.37	7.58	7.63	7.63	7.67
25. Tobacco products	6.06	6.21	6.29	6.45	6.48
26. Alcoholic beverages	.16	.31	.62	.70	.79
VII. 27. Utilities	1.14	1.28	1.73	2.21	2.30
VIII. 28. Other and unclassifiable foods	.68	.95	1.03	1.05	1.19
IX. 29. Foods consumed away from home	3.26	3.80	4.22	4.29	4.36

*Median

OF THE FIRST 13 WEEKS ARRANGED FROM THE HIGHEST VALUE

6	7*	8	9	10	11	12	13	Average
28.34	28.79	28.92	29.52	29.67	29.98	30.79	30.88	28.82
20.13	30.98	21.44	21.55	21.85	22.57	23.45	24.43	20.94
6.37	6.52	6.58	6.79	6.87	6.93	7.09	7.15	6.46
1.15	1.31	1.33	1.63	1.72	1.76	2.10	2.72	1.42
27.45	28.71	29.55	30.60	31.20	31.79	31.82	32.04	28.71
7.03	7.13	7.44	7.53	7.71	8.51	9.09	9.57	7.16
.25	.34	.37	.44	.60	.64	.84	1.72	.47
2.63	2.82	2.85	3.19	3.37	3.62	3.96	4.20	2.94
.14	.16	.17	.27	.27	.51	.65	.94	.28
1.28	1.37	1.46	1.46	1.69	1.76	1.77	1.99	1.39
12.67	13.76	13.99	14.06	14.68	15.43	15.58	16.35	13.14
1.16	1.17	1.30	1.40	1.40	1.44	1.48	2.09	1.27
1.99	2.01	2.08	2.26	2.61	2.64	2.77	2.93	2.06
6.84	7.00	7.17	7.24	7.65	7.77	7.98	8.48	6.97
4.24	4.35	4.45	4.46	4.58	4.68	5.09	5.10	3.93
.95	1.04	1.06	1.13	1.18	1.35	1.37	1.85	1.08
1.91	2.02	2.03	2.13	2.22	2.30	2.34	2.47	1.96
8.11	8.18	8.23	8.28	8.52	8.76	8.78	9.41	8.00
5.29	5.43	5.59	5.60	5.69	5.78	5.87	5.92	5.20
2.45	2.62	2.75	2.79	2.80	2.84	3.06	3.19	2.54
.20	.23	.25	.29	.31	.37	.43	.67	.26
9.25	9.27	9.81	9.89	10.16	10.19	10.30	11.96	9.55
2.07	2.13	2.16	2.20	2.20	2.24	2.36	2.39	2.11
.22	.23	.25	.32	.40	.43	.72	.77	.32
2.25	2.35	2.36	2.49	2.72	2.78	2.93	3.37	2.22
.17	.23	.24	.25	.26	.36	.61	.62	.25
.76	1.00	1.22	1.24	1.43	1.58	1.59	1.84	1.05
1.92	1.94	1.99	2.02	2.07	2.11	2.17	2.19	1.92
1.60	1.70	1.74	1.77	1.77	1.81	1.99	2.10	1.68
8.16	8.36	8.48	8.62	8.73	8.89	9.54	11.02	8.36
6.65	6.75	7.47	7.54	7.80	7.94	8.04	10.32	7.23
.85	.88	.93	1.19	1.74	2.15	2.17	2.19	1.13
2.36	2.23	2.50	2.59	2.62	2.65	2.86	3.20	2.28
1.51	1.59	1.74	1.92	1.98	2.43	2.63	4.03	1.75
4.88	5.02	6.08	6.62	6.99	7.08	7.15	8.57	5.56

COMING SOON

The U.P. Statistical Center announces the forthcoming publication of Dr. Tito A. Mijares' **Percentage Points of the Sum of s Roots (s = 1 — 50): a unified table for testing univariate hypotheses of various types.**

The book consisting of about 250 pages contains a new set of tables that are presently unavailable in the current literature. This timely publication is expected to fill an important gap in the unification of treatment of various types of hypotheses testing common in the statistical analyses of univariate and multivariate data. A comprehensive account of the distribution theories under the various null hypotheses is given in the introduction, with numerical examples in each case that users in both the academic and non-academic fields will find highly useful.

The price is tentatively set at ₱24.00 U. S. (or $8.00 including postage) per cloth-bound copy. Members of the Philippine Statistical Association are entitled to a 20 per cent discount. Orders and further inquiries may be addressed to:

The Director
Statistical Center, University of the Philippines
P. O. Box 479, Manila