MATHEMATICAL FOUNDATION FOR A METHOD OF STATISTICAL ANALYSIS OF HOUSEHOLD BUDGETS

By JOHN W. BOLDYREFF
Harvard University

The object of this paper is to offer a satisfactory method of statistical analysis of household budgets in accordance with the general principles of mathematical logic. I have, therefore, taken these words of Fourier: "Mathematics has no symbols for confused ideas"[1] as my guiding light, and set out to effect a simple and comprehensive analysis of the general type of statistical data which is included under the heading "household budgets," i. e. monetary incomes and expenditures of these incomes.

I have tried to lay the greatest stress, accordingly, on the clarity and terseness of the exposition rather than inclusiveness, attempting to diminish to the utmost the number of undefined ideas and the undemonstrated propositions. I make no special claim to originality and base my method upon the works of numerous previous investigators, summarizing analytically old principles and ideas on the bases of mutual consistency and reducibility to more fundamental principles. This paper is specially framed to relieve the feeling of intellectual discomfort which of late has been troublesome to conscientious investigators in our field, so overcrowded with revelations of numerous parts, rather than with indications of the mode of combination of the major components within the whole. I address here the properly instructed mind and so dispense at times with the elaboration of some statements.

In this summary, therefore, we shall be concerned with laying down a rigid method for analysing the budgetary data, defining their scope formally to include only the monetary incomes and

[1] Quoted from J. A. Schumpeter, *Die Wirtschaftstheorie der Gegenwart*, Wien, I, 11, 1927.

the relative amounts of these incomes spent in a defined manner. This, naturally, excludes all reference to economic theory (e. g. utility, demand curves, etc.) from our discussion; I do so not because of a desire to depreciate the importance of that kind of belief, but because I do not wish to consider it here.

Obviously, "it is never a mathematical proposition which we need, but we use mathematical propositions *only* in order to infer from propositions which do not belong to mathematics to others which equally do not belong to mathematics."[2] Moreover, it is also true that nothing can be purely logical or mathematical (unless we follow Hilbert and define mathematics as a game with meaningless marks on paper); all propositions involve some psychological terms such as defining, meaning, asserting or naming. The method and scope of a mathematical analysis is in a like manner dependent on the purpose for which it is to be undertaken.

The purposes in the study of budgetary data assume varying emphasis depending on the point of view of approach, that of economics, home economics, social welfare, and sociology.[3] All of these approaches are concerned with the relation between the sizes of incomes and the relative amounts spent for certain goods and services.

Generally, the classification of expenditures of an income is made as to the amounts (or proportions) spent for food, clothing, rent, light, education, health, recreation, savings, and amusement. Some investigators limit their classifications to five items: food, clothing, rent, fuel and light, and sundries (everything not included under the first four). Others prefer to subdivide the classification further and break up each of the above nine types of expenditure into what they deem to be its component parts, and proceed to study these new relationships and to generalize from them. On my part, I judge the latter performances ex-

---

[2] Wittgenstein, *Tractatus Logico-Philosophicus*, 6, 211.
[3] C. C. Zimmerman, *Am. J. Soc.*, vol. XXXIII, 6, 1928.

tremely dangerous. It seems to me, that the analysis of the major components of income's expenditure in their relationships to the size of income and to each other should be developed and perfected beforehand, and then only gradually extended to apply to the minor items. Moreover, the splitting up of a few variables (the types of expenditures) into many introduces other difficulties —aside from the fact that a study of simple relationships is apt to be more clarifying—the introduction of a component part of the whole variable as a new variable, immediately raises the question why this component is isolated and not the other. None of the arguments that can be generally cited (and usually no arguments are cited) are really decisive, and the position is extremely unsatisfactory to anyone with real curiosity about the fundamental relationships. Unless we wish the analysis of the budgetary data to remain self-contradictory and meaningless, we must adopt a limiting method, and study not more than two variables at a time. Then, and only then, can we hope to establish or discover any "laws," or functional relationships.

In my experiments to develop a satisfactory method of analysis I would begin generally with five classes of expenditures: food, clothing, rent, fuel and light, and sundries. Later, I have come to the conclusion that some of these tend to have a sort of complementary relationship between them. Thus, "fuel and light" are often higher or lower with a higher or lower "rent," and in some cases a part of "rent" covers "fuel and light," in other cases the discomfort and monetary cost of "fuel and light" lowers the "rent" expenditure. Likewise, some complementary relationship is observed between "fuel and light" and "clothing" (especially in submarginal households) and between "clothing" and "rent" (e. g. social demand of the stylish residential district). These are merely a few examples which led me to question the validity of initial isolation of these three items (clothing, fuel and light, and rent) from each other. Accordingly, I suggest to limit our

investigation to the study of possible relationships between: (1) the size of the income and (a) the amount spent for food, (b) the amount spent for sundries; (2) the amount spent for food and the amount spent for sundries—assuming temporarily for convenience and analysis all other itemized expenditures under rent, clothing, and fuel and light, to be not subject to individual isolation.

As to the unit in household budget, the variety of units employed bewilders at first a mathematical student. Of these, the old scale of two children for one adult, the various other "adult equivalents" (e.g. Engel's quet scale of 3.0 for woman of 20 and 3.5 for man of 25 years; Atwater's scale of 10, 8, 7, 5, 2.5; then the scales of Voit, U.S.D. of L., H. C. Sherman and L. H. Gillett, G. Lusk, L. Emmett Holt, and others—each scale giving "adult equivalents" for children, male and female), all clearly show inability of investigators to agree on a scale to determine the size of a family in standard units. It seems to me that the inventors of such scales forget somehow that "taking an arbitrary individual in the living nature—a man, an animal, a plant—it will generally be found impossible to find out another individual in all respects identical to the first one chosen."[4] The standard scale in budgetary studies is less valid than usual statistical abstractions, for such factors as geographic space (climate, nutritive ratio, energy value, cost), social space (stratification and differentiation), economic space (size of incomes), occupational space (caloric requirement, etc.), time factor (daily, weekly, monthly, seasonal, and longer fluctuations), as well as age (for there is a great latitude in "adult" ages and a corresponding variability in "requirements") and sex differences, are admittedly affecting each budgetary individual in a variety of unknown ways. In view of the complexity of the problem and the enormousness of human population, any "adult equivalent" scale will appear

[4] C. V. L. Charlier, *Acta Universitatis Ludensis*, 1905-6, XVI, 5, p. 3.

to be based on samples obtained in gross violation of the sampling theory, for it is very doubtful that a sufficiently large and representative sample can be secured and it is very hard to see how it can escape being greatly biased. Besides, most of these scales are based on energy requirement only, and refer to "food" but not at all to other types of expenditure; therefore, they would be of little general significance even if they were valid in their specific aspect. Personally, I must reject all such scales as meaningless and incline to hesitate between adopting a "normal family" (on basis of a standard number of members, irrespective of their characteristics) and a "household" (irrespective of number of members and of their characteristics), the presumption being that in a sufficient random sample the differences either way will tend to cancel out. This may not seem to be a more accurate method than others, but, in all probability, it is just as accurate, and its virtue lies, moreover, in the fact that its limitations are all on the surface instead of being hidden away behind a misleading label. The data of the last Census seem to favor this attitude.[5]

The purpose of budgetary analysis is to discover, allegedly, certain functional relationships, if any, between the varying income and the relative amounts of each type of expenditure. To discover such relationships and to determine them explicitly one must recognize that all laws logically function within limits. One needs not go as far as Hilbert and insist that anything involving an infinity of any kind must be meaningless—in pure mathematics this may be a useful abstraction—but it should be obvious that in all organic laws anything infinite appears a stupid fiction which cannot be argued for except by proceeding to a limit. The behavior of the budgetary items is clearly a biotic phenomenon which fact some of the investigators in our field tend to overlook consistently. If there are any functional relationships in the bud-

[5] L. E. Truesdell, *New Family Statistics for 1930,* J. Am. Statist. Assn., March 1933 (Supplement), pp. 154-8.

getary data these will be found only within definite limits of minimum and maximum, and any contradicting evidence to such laws if found below or above these limits cannot be interpreted as disproving such laws.

We shall make our points clearer by illustrating the above exposition by the so-called Engel's Law (I am referring to the second part of it), incidentally commenting briefly on its validity and demonstrating the details of our method.

It will not be amiss to formulate in a few words the part of Engel's Law (1895) we shall be concerned with in our discussion. Comparing the incomes of laboring families, middle class families, and well-to-do families, Engel conjectured that:

(1) the greater the income, the smaller the percentage of outlay for subsistence (food),

(2) percentage of outlay for clothing is approximately the same, whatever the income,

(3) percentage of outlay for rent, and for fuel and light, is approximately the same, whatever the income,

(4) as income increases in amount, the percentage of outlay for sundries becomes greater.

Most of the investigators incline to accept the first and the last of Engel's propositions, both from the static and dynamic viewpoints. As for myself, I like to consider this law with reference to the following questions:

(1) as incomes increase does the percentage of outlay for food decline and the percentage of outlay for sundries increase?

(2) is this a static law; i. e. in a given place, at a given time, will there be a higher percentage of outlay for sundries and lower percentage of outlay for food with larger incomes, and vice versa for smaller incomes?

(3) does this hold in the dynamic aspect—as incomes increase (in time) do the percentages of outlay for food

decline and those for sundries rise, for short and long time?

(4) is this law reversible, i. e. if incomes decrease do the percentages of outlay for food rise and those for sundries decline, statically and dynamically?

(5) can the percentages of outlay for clothing, rent, and fuel and light be treated as constant, statically and dynamically?

(6) can this law be interperted to mean that when the percentage of outlay for food declines the percentage of outlay for sundries rises, and vice versa, statically and dynamically?

(7) if this law is valid, what is its significance for forecasting?

Let us consider first the problem of limits from a purely abstract viewpoint. We assume for the sake of argument this law to be valid and set up a hypothetical series of incomes with the respective percentages and amounts of outlays for food and for sundries. The following example shows clearly that a limit is eventually reached when the law becomes automatically inoperative.

TABLE I.

Income in $	*% for Food*	*$ for Food*	*% for Sundries*	*$ for Sundries*
Under 900	—	—	—	—
" 1,000	50	500	10	100
" 2,000	45	900	15	300
" 3,000	40	1,200	20	600
" 4,000	35	1,400	25	1,000
" 5,000	30	1,500	30	1,500
" 6,000	25	1,500	35	2,100
" 7,000	20	1,400	40	2,800
" 8,000	15	1,200	45	3,600
" 9,000	10	900	50	4,500
" 10,000	5	500	55	5,500

Aside from demonstrating the inevitableness of limits, this illustration shows also that from purely common sense considerations constancy of interrelationship between variation of percentages for food and percentages for sundries is not feasible. That the absolute amount spent for food cannot decline with increase of income but should constantly keep on rising (though, perhaps, in small amounts), should be clear from common sense, even if we shall consider this amount as stationary after a certain sum is reached and credit the increase to sundries (cooks, maids, travel, eating out, etc.)—yet, even in such cases decline should be out of the question.

Now we can give an illustration of the validity of assumption that the percentages of outlay for rent, clothing, and fuel and light, for convenience of analysis and until proven to be contrary, can be held constant. We have tried this with a variety of data and generally found this to be true.

TABLE II.

COMPARISON OF THE PERCENTAGES OF THE TOTAL FAMILY EXPENDITURE FOR THE DIFFERENT GROUPS OF LIVING COSTS[6]

Item	*Eden's 73 English Budgets 1796*	*Engel's Belgian Data 1853*	*Le Play's Method 100 Budgets 1829-88*	*U.S.D.L. (2562) 1890-1*	*U.S.D.L. (12096) 1918-9*	*Groton, N. Y. (92) 1919*
Food	73	66.9	56.8	41.1	38.2	41.7
Rent	12	7.6	6.8	15.1	13.4	13.1
Clothing	7	14.9	16.5	15.3	16.6	11.3
Fuel and light	5	5.6	4.3	5.9	5.3	6.8
Sundries	3	5.0	15.6	27.7	26.4	27.1

Adding the "rent" and "clothing" items from Table II we obtain: 19.0, 22.5, 23.3, 30.4, 30.0, and 24.4; by adding to these their respective "fuel and light" items we obtain: 24.0, 28.1,

[6] Taken from *Noble*, Cornell University Agricultural Experiment Station Bulletin, # 431, Sept., 1924.

27.6, 36.3, 35.3, and 31.2. It seems justifiable to assume these items in their summation to be a constant factor in time analysis. That they are constant for static analysis will be shown later. But it may be mentioned in passing that taking the data from Noble's Table 19 (average percentages of expenditure of items of cost of living of 518 families in New York City, by income groups)[1] and adding up our "constant factor" we get 35.9 for the lowest income group and 36.4 for the highest.

One illustration more. Below are the figures taken from the U. S. B. L., 18th annual report, 1904, p. 101.

TABLE III.

Classified Income	*Rent*	*Fuel*	*Light*	*Food*	*Clothing*	*Sundries*
Under $ 200	16.93	6.69	1.27	50.85	8.68	15.58
" 300	18.02	6.09	1.13	47.33	8.66	18.77
" 400	18.61	5.97	1.14	48.09	10.02	16.09
" 500	18.57	5.54	1.12	46.88	11.39	16.50
" 600	18.43	5.09	1.12	46.16	11.98	17.20
" 700	18.48	4.65	1.12	43.48	12.88	19.39
" 800	18.17	4.14	1.12	41.44	13.50	21.63
" 900	17.07	3.87	1.10	41.37	13.57	23.02
" 1,000	17.58	3.85	1.11	39.90	14.35	23.21
" 1,100	17.53	3.77	1.16	38.79	15.06	23.69
" 1,200	16.59	3.63	1.08	37.68	14.89	26.13
1,200 and over	17.40	3.85	1.18	36.45	15.72	25.40

The "constant factor" taken at the lowest and highest incomes is found to be 33.57 and 36.15 respectively. The examination of the table from the point of view of finding a law, or functional relationship, reveals such phenomenon for the range of incomes from $500 to $1,200, inclusive. We shall proceed to examine the data included in these limits in accordance with our method.

We find the "constant factor" for $500 income to be 36.62 and for $1,200 income, 36.19.

[1] Op. cit.

We assumed a straight line relationship and computed simple coefficients of correlation between:

(1) incomes and percentages of outlay for food
(2) incomes and percentages of outlay for sundries
(3) percentages of outlay for food and those for sundries

We want to stress in this connectoin that to us the coefficient of correlation means a measure of relationship which is already empirically established, not a proof of such relationship. We used L. P. Ayres[8] formula which we found convenient for computing purposes. To avoid a fictitious correlation between incomes and the percentages of outlay for food and for sundries, we have divided the income column by a constant. To facilitate computation we have likewise divided the "percentages for food" and the "percentages for sundries" columns by constants.

In making a summary comment on Engel's law, I would like to stress the following points from a purely methodological viewpoint. There seems to be definite evidence that in a given place, at a given time, the law holds consistently within certain limits. For very low income groups some other law may hold, or no law at all, and as to how extremely large incomes are spent we do not know. From the dynamic aspect, the law appears to have been working from the time of the French revolution up to the beginning of the present depression (much evidence could be cited to support this fairly well known fact, e.g. works of Schmoller, Rogers, D'Avenel, U. S. B. L. S. Bulletins, etc.). However, the study of W. A. Berridge (The need for a new survey of family budgets and buying habits, N. Y. Times, May 10, 1931, and "The Annalist," July 17, 1931) seems to indicate that from the secular standpoint this law is not immediatly reversible, for with the shrinking incomes we observe a definite decline in the outlays

[8] *J. Educ. Research*, I, March-June, 1920.

for all items, including food, except for the outlay for sundries which appears to be almost stationary.

That percentages of outlay for clothing, rent, and fuel and light, can be added up and treated as a constant factor both statically and dynamically with rising incomes we can be reasonably certain of; what will happen with decreasing incomes in time analysis we are not ready to say. However, it must be borne in mind that even with the rising incomes the relationship between the percentages of outlay for food and for sundries need not be perfect as one may be led to think from their high individual coefficients of correlation with income in the example given above.

As to a practical application of the budgetary analysis to forecasting, I shall venture to say that in a socially planned society (if such society is workable), the study of itemized expenditures may prove invaluable. In other societies it may be used to forecast some sort of consumption indices—if these will be successfully computed they will undoubtedly help to flatten the curve of business cycles to an appreciable degree. As to how to develop these indices, I have no suggestion to make just now, except that it must be on basis of extension of a crude analysis similar to one offered here, and application of probability technique, properly based on psychological and historical findings. All I hope to have made clear in this paper is that the subject is very difficult, and that an analysis offered here is sufficient as a first step.

In conclusion, I must stress my indebtedness to Professors J. D. Black, J. A. Schumpeter, and C. C. Zimmerman for advice and suggestions. I am also grateful to Professor Zimmerman for the materials he let me examine. But above all I am indebted to Professor W. L. Crum from whom my point of view and method of attack are wholly derived; anything of value that I may have said in this paper is due to him.

John D. Black