Journal of

UNDERGRADUATE MATHEMATICS

CONTENTS

VOLUME 7
NUMBER 2

SEPTEMBER
1975

(see back cover)

PROPOSALS FOR RESEARCH

The following is a list of questions which are suitable for undergraduate research and whose solutions are on the level of papers sublished in this journal. Solutions may be submitted by individuals or by departments and should be mailed to: Journal of Undergraduate Mathematics, Guilford College, Greensboro, North Carolina, 27410.

WHAT FUNCTIONS GENERATE A PRICE INDEX?

Edward T. Ordman

INTRODUCTION. Goverments and other agencies frequently desire to assign *index numbers* to various phenomena. For instance, if the level of prices in 1967 is taken as 100, the average level of prices in 1970, as given by the consumer price index, was 116.3 (a 16.3 percent rise); by 1973 it had risen to 133.1. Notice that since the 1970-73 rise started from a base of 116.3, it was a rise of 14.4 percent, not 16.8 percent. While the Consumer Price Index and many other indices of changes over time, there are also indices of differences with respect to place. E.g., if prices in the *average* city of 100, a city with high (respectively, low) prices may have an average level of 115 (or 85).

While one is correct in expecting that a *price index function* is usually fairly easy to write down (it is simply some kind of weighted average), it is not true that any plausible-looking function one can write down is useable as an index. This proposal is intended to point out one or two pitfalls and encourage you to try to list requirements for such a function sufficently numerous so that you may narrow down the list of eligible functions.

1. AN EXAMPLE. Since most index numbers are expressed as percentages, it may seem reasonable to arrive at them by averaging percentages. Suppose we wish to determine, as between two cities A and B, which has higher prices and by how much. We could try to find out prices of a list of items $I_1, I_2, \cdots, I_n$ in the two cities (call the prices in the first city x_{11}, $x_{12}, \cdots, x_{1n}$ and those in the second city $x_{21}, x_{22}, \cdots, x_{2n}$), find what percent of x_{1i} is represented by x_{2i}, and average the percentages. That is, set $f(A,B) = (\Sigma 100x_{2i}/x_{1i})/n$ (sum for $i = 1, 2, \cdots, n$). Of course, if there are more than two cities, we can define $f(x,Y)$ analogously for any pair of cities X and Y. In fact, this has been done in some real-life applications [2].

The above formula, unhappily, produces results no mathematician is likely to be willing to tolerate. Suppose that there are only two items I_1 and I_2, with prices $1.00 and $2.00 respectively in city A and $2.00 and $1.00 respectively in city B. We find that city B prices are 200 percent and 50 percent of the city A prices, for an *average* of 125 percent (that is, $f(A,B) = 125$): clearly, prices are 25 percent higher in city B. Unfortunately, we also find $f(B,A) = 125$: prices are 25 percent higher in city A.

2. AN EXERCISE. Using the same formula, find a combination of prices for two items so that prices in city B are 50 percent higher than in city A, prices in city C are 50 percent higher than in city B, but prices in city C are only 60 percent higher than in city A. How much freedom is there in assigning f(A,B), f(B,C), f(A,C), f(C,A), f(B,A), and f(C,B)? Are they all independent, or do some determine the others?

3. DEFINING THE PROBLEM. For concreteness, let us define a *price index function* as a real-valued function of 2n real variables. (There is nothing iron-clad about this definition. If you do not like it, make your own). Treat the 2n variables as two lists of n prices (or other real numbers) each -- typically write f(X,Y) where $X = (x_1, x_2, \cdots, x_n)$ and $Y = (y_1, y_2, \cdots, y_n)$. Mathematically it is immaterial whether the two price lists represent prices at different times or prices in different places.

Now consider the sorts of pathologies we discussed above. It seems reasonable to require that our function satisfy some laws - let us call a few (why not?) the reflexsive, symmetric, and transitive laws. The first two are $f(X,X) = 100$ and $f(X,Y)/100 = 100/f(X,Y)$; the reader may decide an appropriate phrasing for the third. Are these laws sufficient to characterize what we instinctively want a price index function to be? Should we have more (or fewer) laws? Is there beginning to be any evidence that we should pick some other framework for our study (perhaps functions of a single n-tuple instead of a pair of n-tuples?).

PROPOSED THEOREM. *If* f *is a reflexive, symmetric, transitive price index function of* 2n *variables, there is a function* g *of* n *variables such that for all* X *and* Y *(n-tuples),* $f(X,Y) = 100g(Y)/g(X)$.

For example, let $g(x_1, x_2, \cdots, x_n) = x_1 + x_2 + \cdots + x_n)/n$, and let $f(X,Y) = 100g(Y)/g(X)$. If your definitions are reasonable, this is almost certainly a well-behaved index function. For which functions g is f, defined this way, an index function?

4. MEANS. In the above paragraph we tried to characterize index function by establishing general laws that restricted the functions to be considered. Let us now approach the problem the other way, starting from examples and generalizing. By a *weighted arithmetic mean* we mean a function of the form $m(x_1, x_2, \cdots, x_n) = \Sigma a_i x_i/n$ for fixed a_i, i = 1 to n. $f(X,Y) = 100\, m(Y)/m(X)$ may be a good example of an index function; another class of examples uses a weighted geometric mean. Would a harmonic mean work? Is there any general characterization of what sorts of *means* work? (Do they *all*, whatever that means?).

5. DOES THE APPLICATION MAKE SENSE? Now return to the motivating application - say, a simple price index. (Take, for instance, the comsumer price index or the Dow-Jones average). Oftem texts on the subject talk of the *weight* of an item in the index. (For instance, when the price of sugar quintupled in 1974, its relative impact on the consumer price index quintupled, creating a debate as to whether this was accurate - did typical families allot five times as much money to buying sugar as previously?). In a *general* mean or a *general* price index function, what might the *weight* of an item mean? This writer frankly does not know, in any very general setting, but it might have to do with the relative change in the index produced by a small percentage change in the price of the item - perhaps something like a partial derivative. Could one devise an index which assumes that some items are necessities and others are in some sense optional - where the weight of the necessities increases as prices increase, but the weight of

the *optional items* remains constant or even shrinks as their prices (or total price levels?) rise? Could one define *weights* so that a condition on weights could be used to characterize the different sorts of means? (Is there some such theorem as: *if* f(X) *is a mean and the weight of each variable is constant, then* f(X) *is a weighted geometric mean?*). Could this technique be used to more dramatically restrict the functions usable as price index functions, in the presence of suitable hypotheses about consumer behavior? These may be quite difficult problems.

6. THE MISSING DATA PROBLEM. One practical problem that is of critical importance in real life seems to be regularly neglected in elementary courses: *some of the needed prices often cannot be obtained.* One research project presently under way at the University of Kentucky involves gathering 4,200 prices per month. In principle, index numbers are derived from a simple function of 8,400 variables (in practice, a ratio of two weighted arithmetic means). Unfortunately, in an average month, about 15 percent (over 600) prices cannot be obtained - the item is out of season, out of stock, the store does not carry it any more, or perhaps no price is marked and the sales people cannot agree.

There is apparently very little theory available as to how to handle missing data problems. Many *obvious* techniques can be ruled out by means of the theory suggested in the sections above. For instance, one might say something like *disregard each pair with at least one member missing.* If in time 1 item A cost $2 and item 2 cost $2, in time 2 item 1 is unavailable and item 2 is $3, and in time 3 prices are $4 and $3, what should the index be in time 2? Remember that we want the increase determined by a direct comparison of times 1 and 3 to square with 1-to-2 and 2-to-3 increases (transitive law), and that in many practical cases it may be necessary to announce the index for period 2 before period 3 prices become available. Of course, some problems here are practical or even political rather than mathematical in nature. One may or may not want the index to reflect the existence of *shortages*; one may want to compute subindices (e.g., the cost of food as a part of the consumer price index, or the cost of fuel) and have the main index consistent with the subindices as well as with individual items. It is possible that the choice of the original index function may determine how much difficulty these problems of application create.

7. LITERATURE. It may be worth indicating a few points where a literature search might begin. By far the most important source for practical discussions is [1], especially Chapters 10 through 13; it gives many references to papers showing actual government practice. [4] is a recent theoretical (and highly mathematical) article whose references may be used as a start to searching recent theoretical literature. [3] and [5] are intermediate between practice and theory, and rather closer to the approach taken here.

REFERENCES

1. U.S. Department of Labor, Bureau of Labor Statistics, *BLS Handbook of Methods*, Bulletin 1711 (U.S. Government Printing Office, Washington, D.C., 1971).
2. Florida, Department of Administration, *Florida Price Level Index 1973*, A report to the Florida Legislature, (Dr. James C. Simmons, Florida State University, Project Director) (Florida, July, 1974).
3. Doris P. Rothwell, *Use of varying seasonal weights in price index construction*, J. American Statistical Assoc., 53 (1958), 66-77.

4. P. A. Samuelson and S. Swamy, *Invariant economic index numbers and canonical duality: survey and synthesis*, Amer. Economic Review, 64 (1974), 566-593.

5. Allen D. Searle, *Toward comprehensive measurement of prices*, Monthly Labor Review, 94 (1971), 9-22.

Edward T. Ordman
Research Associate
Office of Business Development and Government Serivces
College of Business and Economics
University of Kentucky