What Do Successive Frequency Distributions Show?

By Ronald L. Mighell

Comparisons of successive frequency distributions may be made in two distinct ways that are frequently confused. One way is to make comparison against a given set of class intervals covering the relevant range of the data. The result is a fixed frame of reference, a single classification grid. The other way is to compare "corresponding relative parts" of the frequency distributions—the quartiles, deciles, or similar scale orderings. Such scale orderings, in effect, adjust the class intervals for each distribution so that they are keyed to the central tendency of that distribution. This analysis uses data on stature of fathers and their sons to illustrate the confusion that can occur.

Keywords: Frequency distributions, changing frequencies, stature, size classes.

The purpose of this paper is to throw some light on an overlooked corner of economic analysis. Many public and private agencies commonly classify data on size in terms of current period size, and then draw conclusions as to size changes between periods. That this procedure may often lead to erroneous conclusions because it does not follow identical groups through time first caught my attention several years ago in a short study of concentration in farming (*4*). Recently I came across some data from another field that may help us avoid some preconceptions and bring out the essential elements with fewer complications.

THE EXAMPLE

Join me in looking at some venerable numbers on the stature of fathers and their sons (table 1 and fig. 1). These data are from Yule's *An Introduction to the Theory of Statistics* (*7*). He borrowed them from an earlier (1903) paper by Karl Pearson. For simplicity, I have combined some of Yule's classes and rounded his fractional frequencies. Yule used the data to illustrate correlation analysis. I use them to illustrate two ways of comparing successive distributions.

Table 1. Frequency distributions of statures of British fathers and their sons[a]

Stature	Fathers	Sons	Sons as a percentage of fathers
Inches	*Number*	*Number*	*Percent*
57.5-59.5	3	0	---
59.5-61.5	12	4	33
61.5-63.5	50	24	48
63.5-65.5	157	100	64
65.5-67.5	279	237	85
67.5-69.5	295	323	109
69.5-71.5	194	236	122
71.5-73.5	78	105	135
73.5-75.5	10	37	370
75.5-77.5	0	8	---
77.5-79.5	0	4	---
Total	1,078	1,078	100

[a]Adapted from table III, p. 160, G. Udny Yule, *An Introduction to the Theory of Statistics.* Charles Griffin and Co., Limited, London, 1929.

The distributions, as aggregate entities, show that the sons were slightly taller than their fathers. This is perhaps more evident in figure 1, which shows how the distribution of the sons has moved to the right of that for the fathers. Table 1 also shows that the number of sons expressed as a percentage of the fathers for each height class becomes progressively greater as one goes up the size scale. They rise from 33 to 370 percent.

Some observers have inferred from this that the sons of tall fathers show progressively greater increases in stature than the sons of short fathers. There is no question about the arithmetic, but the inference is a mistaken one. The percentages are correct for comparing changes in the contents of the size classes in a given classification scheme, but not for measuring what happened between the two generations in terms of changes in corresponding relative parts of the population distributions. By corresponding relative parts, I mean quartiles, deciles, percentiles, or similar scale orderings. Such orderings refer to the central tendency of each distribution rather than to a single classification grid according to which the two distributions may be defined.

THE EXPLANATION

The explanation of the error in the preceding inference is simple enough, once it is perceived. Technically, the error involves the shape of the frequency distribution and where it is moving. The possibility for such miscon-

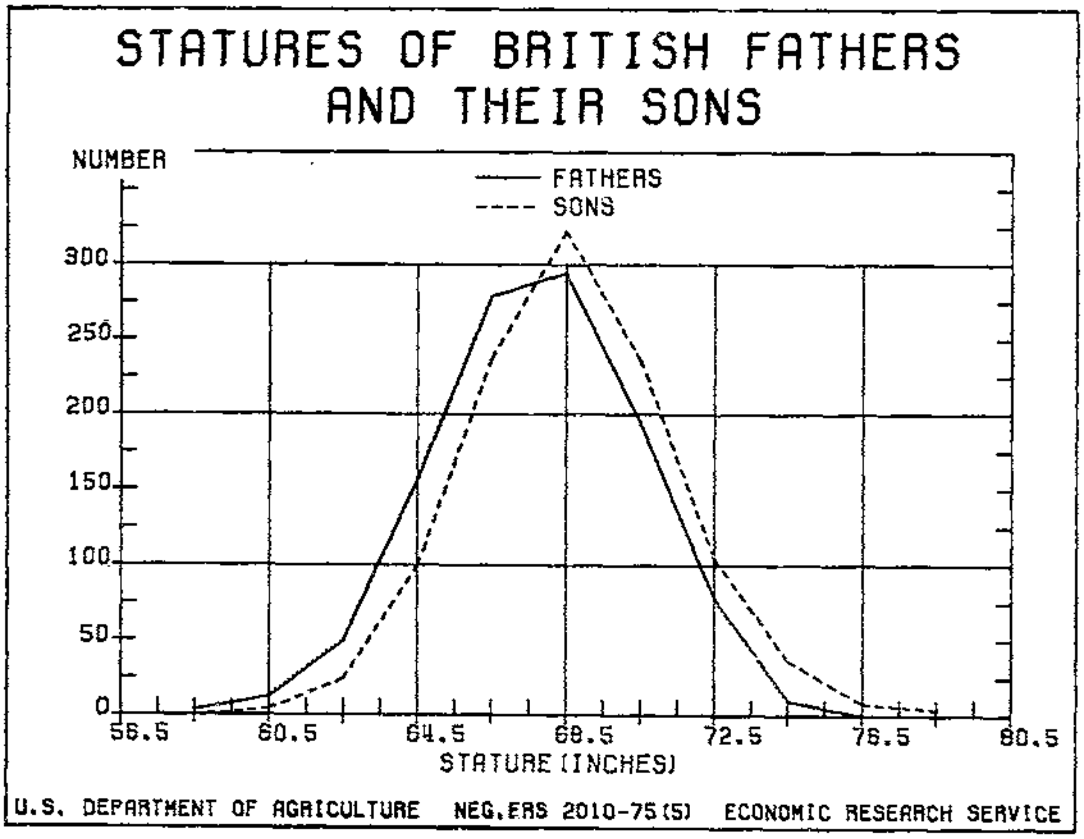

Figure 1

ception is most striking for the bell-shaped normal frequency distribution. But it is also significant for any convex-upward distribution, even in the presence of considerable skewness and distortion.[1]

One clue to understanding is the point of intersection of the frequency curves (fig. 1). Every size class comparison to the left of that point has fewer sons than fathers; every comparison to the right has more. But for each size class of fathers, the corresponding part of the sons' distribution is slightly taller. The aggregate distribution for the sons is about 0.8 inch to the right of that for the fathers; that is, the sons averaged this much taller than the fathers. If the distribution for the fathers is moved over that much and superimposed on the distribution for the sons, the coincidence is very close throughout. Table 2 shows the frequencies as read from the fathers' curve after such a superimposition. Except for minor differences the two curves are highly congruent.

Why did the sons grow taller than their fathers? First, not every son is taller; some are shorter. Yule's original distributions in a correlation table make this clear. There is a scatter, but the verdict of the averages is that the sons are taller. Not much is known about the history of the data, so we can only surmise that the sons' greater stature resulted from favorable background changes in the last part of the 19th century. These probably included improvements in nutrition, disease control, and other environmental factors.

[1] A concave-upward (U-shaped) distribution will give exactly opposite results, but this type of distribution is not common.

Table 2. Frequency distributions of statures of British fathers and their sons and of the fathers' curve as superimposed on the sons' curve[a]

Stature	Fathers	Fathers' curve superimposed[a]	Sons
Inches	*Number*	*Number*	*Number*
57.5-59.5	3	0	0
59.5-61.5	12	8	4
61.5-63.5	50	35	24
63.5-65.5	157	113	100
65.5-67.5	279	235	237
67.5-69.5	295	290	323
69.5-71.5	194	235	236
71.5-73.5	78	119	105
73.5-75.5	10	37	37
75.7-77.5	0	6	8
77.5-79.5	0	0	4
Total	1,078	1,078	1,078

[a]Frequencies shown in column 3 obtained by reading from figure 1 after superimposing the fathers' frequency curve on the sons' curve by moving the fathers' curve to the right 0.8 inch.

EXPANDING AND CONTRACTING SECTORS

Some analysts have referred to the two parts of the distribution indicated above as the "expanding and contracting" sectors.[2] This phrasing is intriguing but may be misleading, because it defines sector in terms of the classification scheme instead of in terms of corresponding relative parts of the distributions of successive populations. What the definition does can be seen by imagining a freight train a mile long that passes through a tunnel of the same length. Suppose we call the part of the train emerging from the tunnel sector A and the part in the tunnel sector B. As the train moves ahead, sector A is continually expanding and sector B is contracting. We are defining sector in terms of a single classification scheme. The train itself and its makeup remain unchanged. No harm is done, because we know how sector is defined and the continuity of the train is understood. But if we should define sector in terms of the composition of the train itself, and call the first half of the train, sector A, and the second half, sector B, we would have another situation.

Most instances of successive distributions are less closely connected than the cars in a train, and there is more opportunity for confusing the two kinds of sectors.

TWO WAYS OF COMPARING FREQUENCIES

So long as we consider a single distribution at one time or place, no problem arises because the two ways are essentially identical. But the moment we compare one distribution with another, two different ways appear. Do we want to compare parts of the same classification grid or do we want to compare corresponding relative parts of the populations? We may wish to do either or both. What we must not do is mix them up and draw wrong conclusions. Each approach has a valid use as long as it is kept separate.

For example, distributions like those in table 1 would be useful to a manufacturer of readymade clothing. He would like to know how many suits might sell in each size class for sons and for fathers separately. His operations depend on information keyed to such distributions. Similarly, a farm machinery manufacturer plans his projections of tractor and machine sizes against information on changes in farm sizes, because these affect the demand schedules he tries to meet.

But many situations involve comparisons between corresponding relative parts of successive populations. Persons using analysis-of-income data make such comparisons, and income specialists are accustomed to making quartile and decile comparisons between different population groups and between different nations. From such measurements they can form some idea of relative concentrations of income among populations widely different in income scale.

[2] Nikolitch was apparently the first to use these terms in his analysis of changes in sizes of farms (*5*).

In opening his discussion about the expanding and contracting sectors of American agriculture, Nikolitch had this to say:

> The total number of farms in American agriculture has been declining for more than 20 years. Yet the number of larger farms is increasing while the number of smaller farms is decreasing rapidly. Thus American agriculture has an expanding sector of large farms and a contracting sector of relatively small farms (*5*, p. 1).

Nikolitch recognized that ". . . the concentration of production in the expanding sector of agriculture is due in large measure to the increase in the number of such farms . . ." brought about by smaller farms increasing in size and moving up the size scale. But he sometimes seemed to forget this relationship; and many readers have misinterpreted the findings, assuming that the large farms are increasing in size more rapidly than the small farms.

In comparing changes in relative size distribution of commercial farms in the United States from 1959 to 1964, I found that, despite a considerable drop in numbers of farms and increase in size of farms, the relative size distribution had not changed greatly (*4*). The truth seems to be that, on the average, farms in all size groups as of any given base period have been growing larger at roughly the same rate, in the same sense that the British sons in our illustration were growing taller in all groups. Similar findings for 1939 and 1964 are presented by Charles L. Schultz in his analyses of census data (*6*). The fact that there has been a dramatic reduction in the total number of American farms in the last 30 years means, of course, that farming is absolutely more concentrated and in fewer hands. But it does not mean that relative concentration within the distribution has changed appreciably. The top 20 percent of the farms may still produce about the same percentage of the farm output.

The useful series on number of farms by value of sales classes is sometimes misinterpreted in the same way (*2*). For example, consider the two pie charts in figure 2, which appeared in the July 1975 *Agricultural Outlook* (*1*, p. 21). The accompanying discussion says that "rapid changes are taking place in the structure of our Nation's agriculture, as evidenced . . ." by changes in the distribution of farm numbers and gross and net income measures by value of sales classes. It goes on to point out that large increases in numbers and sales in the larger sales classes took place between 1960 and 1974. But it neglects to say that the greater part of the change is due to inflation and it takes no account of the probability that the two distributions if compared by percentiles are probably not much different.

CONCLUSIONS

Thus, two distinct ways exist to compare successive frequency distributions. Each stands on its own feet; each serves a different purpose. One considers changes from the viewpoint of a single frame of reference, the

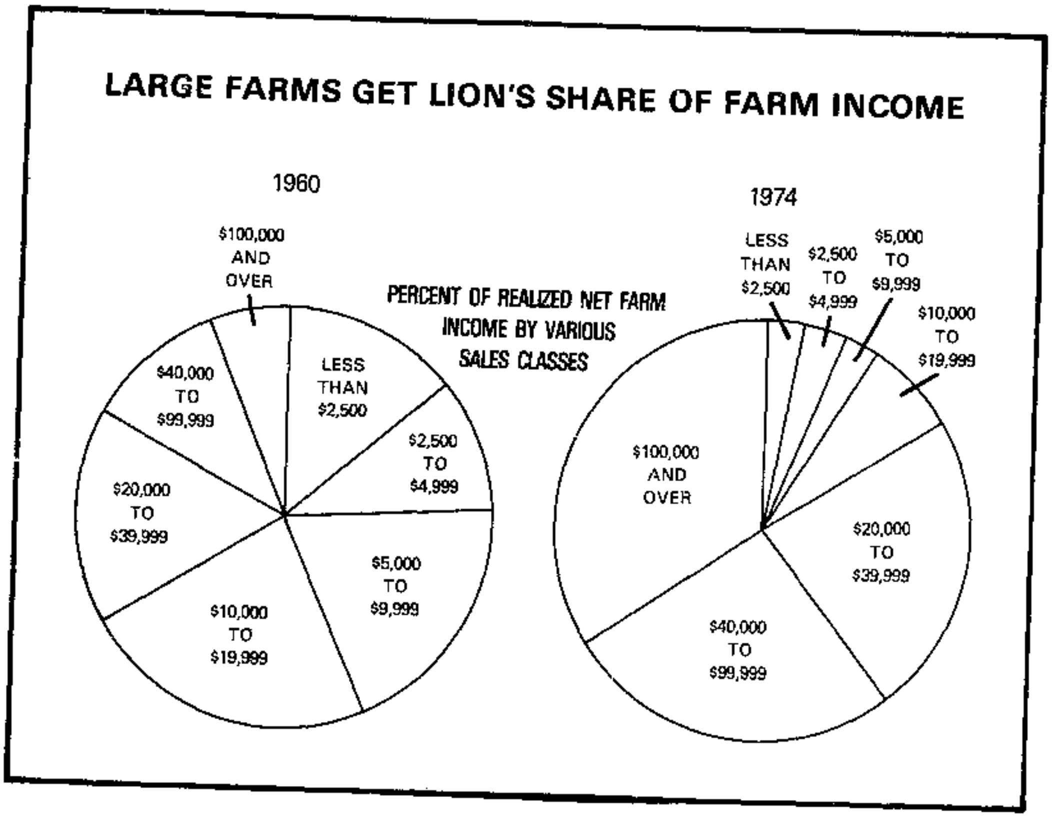

Figure 2

fixed class interval scheme of classification. The other considers changes between corresponding relative parts (quartiles, deciles, etc.) of each frequency distribution.

The examples presented illustrate the purpose and proper use of each comparison. The findings have significance in the interpretation of changes in farm size, farm income, and similar problems.

REFERENCES

(1) U.S. Department of Agriculture. *Agricultural Outlook.* Econ. Res. Serv., AO-2, July 1975.

(2) ______________. *Farm Income Statistics.* Econ. Res. Serv., Stat. Bul. 547, July 1975.

(3) U.S. Department of Agriculture. *1974 Handbook of Agricultural Charts.* Econ. Res. Serv., Agr. Handbook 477, October 1974.

(4) Mighell, Ronald L. "Concentration in Farming and Transition Bias." *J. Farm Econ.* 51: 1114-18, December 1969.

(5) Nikolitch, Radoje. *The Expanding and Contracting Sectors of American Agriculture.* U.S. Dept. Agr., Econ. Res. Serv., Agr. Econ. Rpt. 74, May 1965.

(6) Schultze, Charles L. "The Distribution of Farm Subsidies: Who Gets the Benefits? A Staff Paper, The Brookings Inst., 1971.

(7) Yule, G. Udny. *An Introduction to the Theory of Statistics.* Charles Griffin and Co., Ltd., London, 1929.