Journal of Official Statistics
Vol. 2, No. 1, 1986, pp. 71–76
© Statistics Sweden

Letters to the Editor

Letters to the Editor will be confined to discussions of papers which have appeared in the Journal of Official Statistics and of important issues facing the statistical community.

Survey Research Ought to Be Viewed and Practiced as Applied Statistics

Dear Editor,

Statistics, in the sense of a scientific discipline, has been characterized as 'the servant of all sciences' and 'the technology of the scientific method'. These and similar characterizations reflect the fact that statistics has found important applications in many fields of human activity. Statistics should, in fact, play a role in the design and analysis of *any* empirical study, for which the outcome is not deterministic, that is, not (fully) determined by the design.

The role of statistics in empirical studies varies greatly in different fields. In some fields, statistics plays a major role in close conjunction with a good insight into the subject-matter involved. If this role is dominant, it may find expression in the establishment of specialized branches of statistics. Biometry, econometrics, and psychometry are three examples of this. Workers in these specialized branches are typically statisticians by profession, or are at least thoroughly trained in statistics.

But the situation is quite different in other fields. It may even be markedly different: the dominant role is played by the subject-matter discipline involved, while statistics at best plays a minor role.

In my view, we statisticians must not consider these other fields as 'none of our business': On the contrary! Just as a physician observing medical quackery has an obligation to intervene, we statisticians should intervene, in cases of statistical quackery. To be sure, I am far from alone in the statistical community holding this view. I will restrict myself to drawing the readers' attention to the Presidential Address given by Professor John Neter at the 1985 annual meetings of the American Statistical Association.

My past and current observations on empirical studies in the social, behavioral, and liberal arts disciplines – limited to Sweden and the United States – have convinced me that statistical quackery is (or at least ought to be considered) a serious problem in these fields. More specifically, I am concerned about the way in which 'survey research', perhaps the most widely used tool in these disciplines, is being taught and used. The balance of this note will focus on 'survey research'.

There is certainly no universally accepted definition of 'survey' and hence no such definition of 'survey research'. For the purpose of this note, I will be content to mention the following special case: an empirical study, for which the data are collected by interviews with a sample of individuals is a typical survey. An opinion poll is a trivial example.

As to 'survey research', I will quote from the introduction to Glock (1967), an old but still appreciated text. In that introduction, written by the great sociologist P. F. Lazarsfeld, 'survey research' is viewed as "a mode of inquiry which combines a distinct method of data collection with a district form of analysis." The method of data collection is typically "sampling," and the form of analysis is "a special version of multivariate analysis" which allows "the study and interpretation of complex interrelationships among a multiplicity of characteristics."

It would be natural to assume that any 'mode of inquiry' characterized by the use of sampling and multivariate analysis would be generally viewed as 'applied statistics' similar to the specialized branches of applied statistics mentioned in the beginning of this letter. It is true that there is, in the American Statistical Association, a section for 'survey research'. And it is likewise true that there are institutions in some countries which specialize in 'survey research', sometimes described as 'survey research centers' or the like – with professional staff statisticians. But by and large, the statistical profession has in my view, judging from curricula at statistics departments and the contents of the statistical journals, shown insufficient interest in 'survey research'.

The situation is different in some extra-statistical disciplines, and notably in sociology. In what follows, I will focus on 'survey research' in that discipline[1]. The sociological profession has indeed taken a positive interest in 'survey research', which has manifested itself in a variety of ways:

i. many texts on 'survey research' are written by sociologists;
ii. 'survey research' is taught in many departments of sociology;
iii. survey measurement, and especially questionnaire design and interviewing methods, is presented in such courses as a sociological topic; and
iv. papers on 'survey research' methods written by sociologists are relatively frequent in sociological journals.

Thus the impression has grown in broad circles (both academic and extra-academic) that 'survey research' is basically *applied sociology*. This misleading impression is reinforced by some sociologists' use, enthusiastic and often uncritical, of program packages for computerized statistical analysis; it suffices to mention the prevailing LISREL-mania in some sociological circles.

The point I want to make in this letter is that it is high time (but hopefully not too late) for a change. And this change should be profound; the goal must be to make 'survey research' viewed and practiced as *applied statistics*.

This goal has a quantitative and a qualitative aspect:

i. *more* use of statistical theory and methods is called for; and
ii. the theory and methods used must be *today's theory and methods*: it is no longer defensible to rely on old-fashioned cookbooks or to make uncritical uses of program packages for computerized statistical analysis.

We may have a long way to go, with many road blocks. But let's go!

[1] I have observed notable cases in other disciplines as well. It recently came to my attention that a certain professor of education obviously felt qualified to negotiate a program for building up competence in "survey research' in a developing country. Not many competent statisticians would feel qualified for such a task!

Reference

Glock, C. Y. (Ed.) (1967): Survey Research in the Social Sciences. Russel Sage Foundation, New York.

Sincerely,

Tore Dalenius
Brown University
Providence, R. I. 02912
U.S.A.

Reply

Dear Professor Dalenius,

Thank you very much for your letter.

I suppose that a letter of this kind will generate discussions in various parties. However, the editorial board will abstain from any comments at this point. Instead we hereby invite the prospective debaters to reply to your letter. The Journal of Official Statistics is able to publish a limited number of replies.

Sincerely,

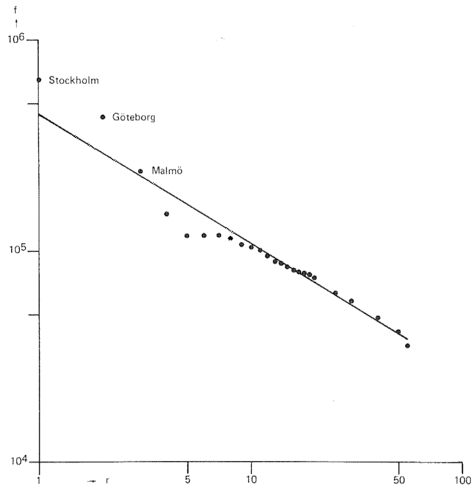
Lars Lyberg
Editor

Zipf's Law

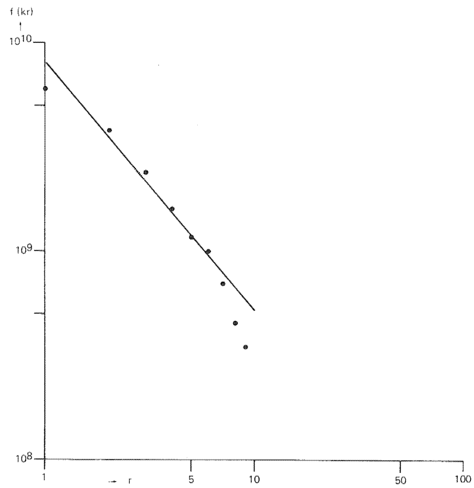
Dear Sir,

Some forty years ago G. Kingsley Zipf published an extensive book entitled "Human Behavior and the Principle of Least Effort; An Introduction to Human Ecology" (Wesley Press Inc., Cambridge, Mass., 1949; Hafner Publ. Co., New York, 1965). He derived a relation between magnitude and rank which refers to collections of numbers in various fields of human activity. I think this relation is not often met in the statistical literature and may possibly be little known, so that it might be worthwhile to throw a spotlight on it.

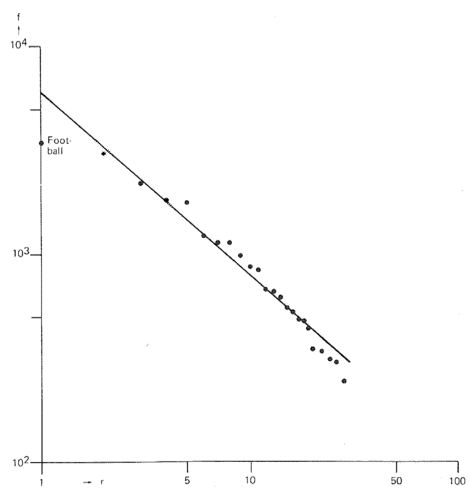
Graph I. Population of the biggest municipalities Sweden 31 - 12 - 1983



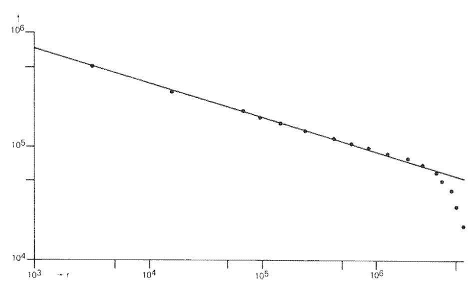
Graph II. Industrial investments Sweden 1982



Graph III. Athletic associations Sweden 1 - 1 - 1983



Graph IV. Income earners Sweden 1982



Consider a collection of quantities f_i (i = 1,2,3,...,n) (e.g., the numbers of inhabitants of the municipalities in a country). They are ranked according to their magnitude; the greatest f_i is given rank $r = 1$. The result is

$$f r^a = \text{constant or } \log f + a \log r = C,$$

where a and C are parameters.

Graphical representation on a double logarithmic scale reveals a straight line, though deviations occur. I give four examples, taken from the statistical data of the 1985 Statistical Abstract of Sweden (Statistisk årsbok), Tables 20, 104, 217 and 436.

It is remarked that in the case of a frequency distribution, which gives less information than a series of separate numbers, the Zipf procedure can be followed and the result is similar to Pareto's Law (see Graph IV).

Yours sincerely,

J.H.C. Lisman
Blauwe Kamerlaan 19
2594 BK's – Gravenhage
The Netherlands

Estimating Variance of a Consumer Price Index and Some Comments on Inference

Dear Sir,
The consumer price index computed by government statistical offices is basically of the form given by Laspeyres. In the form useful for computation it is given by

$$I_t = \sum_{i=1}^{g} w_{oi} r_{ti}, \qquad (1)$$

where w_{oi} ($i=1,2,\dots, g$) are the expenditure weights for the g consumer items and r_{ti} ($i=1,2,\dots, g$) are the corresponding item indices or item price relatives at monthly time period t. The analogue of I_t is the classical formula

$$\sum_{i=1}^{g} p_{oi} q_{oi} \left(\frac{p_{ti}}{p_{oi}}\right) / \sum_{i=1}^{g} p_{oi} q_{oi} =$$

$$= \Sigma\, p_{ti}\, q_{oi} / \Sigma\, p_{oi} q_{oi}.$$

In a paper published by the Organisation for Economic Co-operation and Development in 1980 (OECD, Paris) entitled "Consumer Price Indices," it is noted on p. 9 that the item indices "may be calculated either as relatives of average prices or as relatives of price averages and the two methods give different results. The averages may be weighted or unweighted." Weights are revised annually by some countries. Imputations and adjustments are made in the price indices to take into account disappearing items. Prices are also adjusted for quality changes (see OECD paper again, pp. 9, 32, and 34).

Perhaps because of all these complexities in an index, only in a few countries (e.g. the United States, BLS *Handbook of Methods,* Bulletin 2134–2, 1984, Government Printing Office, Washington D.C.) have attempts been made to look into the statistical aspects of price indices, and to set up sampling procedures on a national basis to collect prices from outlets so as to make possible the estimation of its variance.

Consider the situation when the item indices r_{ti} in (1), with all their *ramified adjustments,* are based on a national probability sample. Assuming that at time period t the errors in w_{oi} can be ignored, we find

$$V(I_t) = \sum_{i=1}^{g} w_{oi}^2 V(r_{ti}) + 2 \sum_{i>j} w_{oi} w_{oj} \operatorname{Cov}(r_{ti}, r_{tj}). \qquad (2)$$

Note that the second term in (2) takes into account the possibility of association or covariation between estimated price relatives of pairs of items. Certainly with a single sample there is no practicable and easy way for estimating $V(r_{ti})$ and $\operatorname{Cov}(r_{ti}, r_{tj})$ that would prove palatable to a government statistical office, even given that the outlet sampling procedure is simple and computer resources are adequate. The technique of interpenetrating samples as developed and used by Mahalanobis and Lahiri provides a good solution to this problem. See D.B. Lahiri's article in *Sankhyā,* Vol. 14 (1954), pp. 264–316.

If a government collects its prices using two *statistically independent* interpenetrating samples of outlets, based on identical sampling procedures for each, then the variance of the index I_t may be estimated without difficulty. Statistical independence is ensured by the replacement of the units of the first sample before the drawing of the second. Observational independence is assured if prices from the two outlet samples are collected by *different* groups of price collectors or *different* parts of the price collecting agency.

Of course all this will cost more. However, considering that observational records of prices paid by consumers are also subject to nonsampling errors, the result may well be worth the extra expenditure. If price observations are entirely free from errors, and no mistakes of any kind are made in the processing of these ideal price data for the computation of an index, then the technique would be almost redundant.

Thus if r_{t1i} and r_{t2i} are estimates of item indices ($i=1,2,\dots, g$) based on two independent interpenetrating samples, made with all relevant adjustments specific to time t, then there will be two indices given by

$$I_{t\alpha} = \sum_{i=1}^{g} w_{oi} r_{t\alpha i}, \qquad \alpha = 1,2$$

the average of which is

$$I_{t\cdot} = \frac{1}{2}(I_{t1} + I_{t2}),$$

with variance

$$V(I_{t\cdot}) = \frac{1}{2} V(I_{t1}) = \frac{1}{2} V(I_{t2}) . \quad (3)$$

Because of the uniformity of sampling procedure in the selection of each independent sample as well as the independence between samples in the work of price collection,

$$V(I_{t1}) = V(I_{t2}) .$$

Furthermore, for $\alpha = 1,2$ an unbiased estimate of $V(I_{t\alpha})$, the expression for which is given by (2), is

$$\hat{V}(I_{t\alpha}) = \frac{1}{2}\sum_{i=1}^{g} w_{oi}^2 (r_{t1i} - r_{t2i})^2 + \sum_{i>j} w_{oi} w_{oj} (r_{t1i} - r_{t2i})(r_{t1j} - r_{t2j}) . \quad (4)$$

Note that in most countries an index is based on hundreds of price observations, and therefore (heuristically speaking) the sampling distribution of $I_{t\alpha}$ for such countries can be expected to be concentrated around its mean or median. Hence the estimated variance given by (4) will reflect this concentration.

From (3) and (4) it will be found that an unbiased estimate of $V(I_{t\cdot})$ is

$$\hat{V}(I_{t\cdot}) = \frac{1}{4}\sum_{i=1}^{g} w_{oi}^2 (r_{t1i} - r_{t2i})^2 + \frac{1}{2}\sum_{i>j} w_{oi} w_{oj} (r_{t1i} - r_{t2i})(r_{t1j} - r_{t2j}) = \frac{1}{4}\{\sum_{i=1}^{g} w_{oi}(r_{t1i} - r_{t2i})\}^2 = \frac{1}{4}(I_{t1} - I_{t2})^2; \quad (5)$$

which is a simple formula. Note that nonsampling errors in price data (which include all errors at all stages up to computation of the index) are automatically built into I_{t1} and I_{t2} and therefore into $\hat{V}(I_{t\cdot})$. Hence the standard error $\frac{1}{2}|I_{t1} - I_{t2}|$ also measures the effects of all such errors. Considering that a consumer price index is used for a variety of purposes, including wage adjustments, this property of the technique, which provides for the assessment of both sampling and nonsampling errors, is a desirable one.

Furthermore, following the argument initiated by "Student" in 1908 *(Biometrika,* Vol. 6, p. 13), we find

$$P\{I_{t1} < I_{tM} < I_{t2}\} = \frac{1}{2}, \quad (6)$$

where I_{tM} is the median of the distribution of the index of which I_{t1} and I_{t2} are members. In view of the remarks immediately following (4), which imply the closeness of the mean to the median, the probability statement (6) is of practical significance.

To generalize, if routine price collection is organized in k interpenetrating samples of an overall size permitted by the budgeted funds, then arguing as above there will be k indices $I_{t\alpha}$ $(\alpha = 1,2,\ldots,k)$ the mean of which is

$$I_{t\cdot} = \sum_{\alpha=1}^{k} I_{t\alpha} / k ,$$

with estimated variance

$$\hat{V}(I_{t\cdot}) = \sum_{\alpha=1}^{k} (I_{t\alpha} - I_{t\cdot})^2 / \{k(k-1)\} .$$

We have also that

$$P\{I_{t(\text{least})} < I_{tM} < I_{t(\text{greatest})}\} = 1 - (\tfrac{1}{2})^{k-1} \quad (7)$$

When $k = 5$, (7) shows that the probability of the median index lying between the least and the greatest index will be $1 - (\frac{1}{2})^4 = 0.9375$.

In the situation when sampling and nonsampling errors in the base year expenditure weights w_{oi} $(i = 1,2,\ldots, g)$ are admitted, a *tractable* solution is still *possible* if two statistically independent interpenetrating samples are used for estimating them. Then there would be available pairs of expenditure weights $w_{o\alpha i}$ $(\alpha = 1, 2; i = 1,2,\ldots,g)$. With the price information from the two interpenetrating samples of outlets, an unbiased estimator of the variance of the mean of the four possible price indices can be obtained. However, the derivation of this variance is very laborious and thus the foregoing simple case is to be preferred.

In retrospect it can be seen that F.Y. Edgeworth was really the originator of the basic ideas on the sampling variance of price index numbers. In his second memorandum of 1889 to the British Association for the Advancement of Science, he derived, under each of two assumptions, formulae for the modulus of an index number defined as a weighted average of price relatives where the weights and the corresponding price relatives are assumed to be independent random variables. (Our modern equivalent term for this exercise would be "derivation of standard error.") He illustrated his theory with two interesting examples on the computation of this modulus, one of which was suggestively captioned "tests of accurate measurement." (For easy reference see pp. 227 and 304–321 of Edgeworth's *Papers Relating to Political Economy,* Vol. 1 published by The Royal Economic Society, London in 1925.) I point this out because, since 1950, writers on the sampling variance of price index numbers appear to have overlooked this part of his work.

My concern with the statistical aspects of price index numbers started in Burma in 1952 when I read a paper to the Burma Research Society on the subject at a conference organized by the late John Sydenham Furnivall, the principal founder of the Society. He was interested in the whole range of colonial economics and wrote widely on the subject with deep insight and much humane concern. Of some ideological relevance to this letter is the article on "The organisation of consumption" which he contributed to the *Economic Journal,* Vol. 20, pp. 23–30, in 1910 when Edgeworth was its editor.

Yours truly,

John C. Koop
3201 Clark Avenue
Raleigh, NC, 27607
U.S.A.