Representative Sampling, III: The Current Statistical Literature
Author(s): William Kruskal and Frederick Mosteller
Source: *International Statistical Review / Revue Internationale de Statistique*, Vol. 47, No. 3 (Dec., 1979), pp. 245-265
Published by: International Statistical Institute (ISI)
Stable URL: http://www.jstor.org/stable/1402647
Accessed: 24-08-2017 18:20 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms

International Statistical Institute (ISI) is collaborating with JSTOR to digitize, preserve and extend access to *International Statistical Review / Revue Internationale de Statistique*

JSTOR

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

International Statistical Review, 47 (1979) 245–265 Longman Group Limited/Printed in Great Britain

Representative Sampling, III: the Current Statistical Literature

William Kruskal[1] and Frederick Mosteller[2]
[1] *University of Chicago and* [2] *Harvard University*

Summary

The meanings of 'representative sample' and 'representative sampling' in the statistical literature are classified, illustrated, and discussed. The categories are those of our prior papers, together with three new ones: specific sampling methods, permitting good estimation, and good enough for a particular purpose. The paper closes with a sketch of relevant technical or mathematical directions.

1 Introduction

Our two prior papers on representative sampling, Kruskal and Mosteller 1979a and 1979b, dealt respectively with the concept in non-scientific writing and in the scientific literature outside statistical writing itself. We continue our analysis by discussing representative sampling in statistical literature. That literature is special for our purposes because

(i) it includes a complex and illuminating historical development, and
(ii) it shows that statisticians have not always been as careful and precise as one might have hoped when they deal with a concept that is statistical if anything.

We cannot distinguish sharply between the statistical and other scientific literatures, and we make no claim to such sharpness. Indeed, our second paper in this sequence contains for clarity two or three statistical examples.

Our earlier intent had been to combine in this third paper an historical treatment along with an analysis of current usages of the representative idea. The combined enterprise has become too much for a single paper, and so, with sympathetic editorial counsel, we postpone the historical discussion to a fourth, final paper. Of course, no sharp line separates history from the present, especially since we interpret 'current literature' as anything since 1940. We do not fret about the absence of a sharp line, and we recall that our second paper contained references to the past as far back as Cicero.

In the statistical literature, we shall see examples of all the meanings for 'representative sampling' described earlier:

1. General acclaim for data.
2. Absence of selective forces.
3. Miniature of the population.
4. Typical or ideal case(s).
5. Coverage of the population.
6. Vague term, to be made precise.

To these, we shall here add three more meanings:

7. Representative sampling as a specific sampling method.
8. Representative sampling as permitting good estimation.
9. Representative sampling as good enough for a particular purpose.

and we have a section on linguistic cousins and variations.

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

The survey outlined just above, we call Part 2; we then end this paper with Part 3 on mathematical methods, a sketch of some more technical ideas or directions suggested by earlier discussion of representativeness.

Before turning to the analysis of Part 2, we complete this introduction by extending a little our motivating remarks in the first paper, where we mentioned other writings that resembled our analysis of the meanings of, and concepts behind, a word or phrase. Since then we have learned of further such recent writings, and we give examples. First we cite Lippman (1955), who says that 'Words like liberty, equality, fraternity, justice, have various meanings which reflect the variability of the flux of things. The different meanings are rather like different clothes, each good for a season . . ., none good for all times' (p. 111) and again 'If a term has many diverging definitions, it is better to begin by assuming that it is full of meanings. For none of the main ideas of our civilization has a single meaning' (p. 120).

A lengthy article by Spitzer (1942–43) stresses a point we too have stressed, that the ideas and concepts associated with the word are at least as important as the word itself, even if the word provides focus for a particular investigation. Another article in the philosophical mode, by Carritt (1937), gives clear distinctions among meanings of the word 'good'.

An essay by Lucien Febvre (1973), encourages us by saying that 'It is never a waste of time to study the history of a word. Such journeys, whether short or long, monotonous or varied are always instructive' (p. 219). To illustrate that point, we cite a 1978 article by di Corcia on the word 'bourgeois' and its cousins; di Corcia has in part a modestly statistical approach. He writes: 'How frequently were individuals designated by both *état* – *bourgeois* or otherwise – and *métier*? In a simple random sample of the principals named in Parisian notarial acts of the eighteenth century, only 2.31 per cent were described by such a combination . . .' (p. 227). In a footnote on the next page, he says that 'The *liasses* of the *études* for the eighteenth century at the Minutier Central were entered randomly and a simple random sample made of 951 individuals'.

Michael Stoto has called our attention to an etymological detail that fascinates us in our setting of representative sampling. The Italian word *campione* means both champion and sample – thus embodying the Emersonian sense of the superior specimen. One possible etymology is that *campione* (from Latin *campus*) at first meant fighter, gladiator, etc; from there it extended to the sense of champion, and then extended further to the sense of a sample of merchandise presumably as champion of the whole. We are reminded of the everpresent tension between a sample that typifies and a sample that glorifies, a tension that is with us in daily life, for example, whenever we buy a box of strawberries.

Finally, we mention a beautifully written book of word studies, *Keywords*, by Williams (1976). This book brings us full circle back to our main business for, on pages 222–225, it historically analyzes uses of the word 'representative'.

2 Central analysis

B.1 *Rhetorical usage; general acclaim*

The statistical literature contains its share of vague laudatory use of the expression 'representative sample'. The sense is usually something like this: 'Take my word, without evidence, that my sample will not lead you astray'. This rhetorical usage is questionable, whether in the statistical or the extra-statistical literature.

Our first two examples are perhaps relatively venial for they come from press releases or popular brochures of Federal statistical agencies: 'Sampling is the technique of getting a clear statistical picture of a whole universe . . . from a small representative sample of the whole. The sample must be drawn with mathematical precision.' [From unpaginated brochure, *Census USA/A Thumbnail History of the Nation's Factfinder*, Bureau of the Census, 1974.] 'The

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

price of a representative sample of commodities sold in primary markets of the United States has risen from $100 in 1967 to $122.' [From p. 4 of 'Wholesale Price Index – July 1976' news release of the Bureau of Labor Statistics 12 August 1976.] Indeed, it could be argued that these uses fall under our meaning 6, vague usage, to be made precise. See Section 2.6. In these two examples, however, the making precise comes in *other* documents published by the same agencies.

Next, we exhibit an instructive quotation from a widely used textbook: 'It is just as important to choose samples representative of the population examined as it is to make quite random the taking of the ultimate experimental units. The two procedures are not incompatible, both being integral parts of good sampling designs. Incidentally, the criterion of representativeness makes it imperative that the investigator confine his interpretations to the sampled population, not broadcasting them over populations in general.' [From p. 268 of Snedecor, George W., *Statistical Methods*, Ames, Iowa: Iowa State College Press, 1956.] This quotation has special interest because it clearly separates representativeness from random selection, although many authors – as we shall see – connect the two ideas closely. In addition, Snedecor quite appropriately makes the distinction between sampled and target populations.

Another textbook example shows an ambiguity in *disclaiming* representativeness: '. . . the set of n variables . . . is a subset (not necessarily a representative sample) an infinite universe of such measures.' [From p. 221, at the start of a discussion of image factor analysis, in Harman, Harry H., *Modern Factor Analysis*, Chicago: University of Chicago Press, Third edition, revised, 1976.]

From a survey research textbook, we take a final example: 'Our concern . . . is that the *number* and *kinds* of people in the sample be sufficiently representative of the whole population to enable us to make sound generalizations about that population.' [From p. 24 of Backstrom, Charles H., and Hursh, Gerald D., *Survey Research*, Evanston, Ill.: Northwestern University Press, 1963.] The Backstrom-Hursh book also describes other senses for the idea of representativeness. On pages 2–5, we find both the miniature and the typical concepts. Pages 70–71 recommend checking sample characteristics against census information for distribution of age, sex, income, education, etc.

2.2 *Absence or presence of selective forces*

A second sense of representative sampling is that of the absence of selective forces in the sampling process. (In the same way, the presence of selective forces makes sampling unrepresentative in this sense.) Of course the idea of selective forces itself is ambiguous and requires discussion. Sometimes other words, *e.g.* 'biased', are used for 'selective'. Confusion may sometimes occur; for example, 'biased' has a clear meaning in estimation theory, namely that the expectation of an estimator is not everywhere equal to the parameter being estimated, but 'biased' for a sample is different . . . an estimator based on an unbiased sample might itself be biased and vice-versa.

The absence or presence of selective forces as an explication of representativeness occurs widely in the statistical literature. Let us quote first from the classic text on medical statistics by A. Bradford Hill: 'We must . . . consider . . . carefully whether the sample is representative of all such patients and not in any way biased or 'selected' (p. 25). 'By a selected sample . . . [the statistician] denotes a sample which is not representative of the universe of which it is a part. [Note some circularity here! WK-FM]. The selection may have been deliberate . . . for instance, if . . . the treatment of respiratory tuberculosis by . . . surgery were confined to patients with . . . disease in one lung only . . .' (p. 25). 'More often . . . the "selection" is not deliberate but is quite unforeseen or is unrealized . . ., e.g. the number of male births [recorded in the birth column of *The Times*] was 3304 and female births 3034. . . . It is clear that from the point of view of sex ratio the births recorded in *The Times* are unlikely to be representative

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

of the births in the country as a whole. . . . [Perhaps] proud parents are more likely to record their male heirs than their daughters. . . . Whatever the explanation, with such a sample of births . . . one could not generalize . . . with any security' (pp. 26–27). '. . . hospital statistics can . . . rarely be regarded as unselected. The patients are frequently drawn from particular areas and from particular social classes . . . [In] many diseases only those patients who are seriously ill are likely to be taken to hospital' (p. 27). [From Hill, Austin Bradford, *Principles of Medical Statistics*, New York: Oxford, Eighth edition, 1966.]

Non-response in surveys and censuses is often cited as a source of dangerous selection, e.g.: 'A Census form with a mixture of mandatory and optional questions may be tempting . . . but is not . . . recommended . . . those responding could hardly be treated as a representative sample'. [From p. 217 of 'Problems of confidentiality with particular reference to population censuses', T.P. Linehan, *Bulletin of the International Statistical Institute*, **45** (Book 3) (1973) 214–224. Discussion 225–227.]

Vincent Barabba, responding to a seminar query said about selective forces: 'Precinct work, for a political party, would be an example of the worst kind of sampling for drawing inferences about the general population. The person you send out is a very partisan person who is going to people who generally hold similar beliefs. And, if the people don't, the worker is supposed to try to persuade them. But, we want to find people representative of the whole community and, for purposes of sampling, we don't care where they live; that's why we select them randomly. If we had to interview ten people from a particular sampling unit, and we went right down the block and took them in sequence, we would have a biased sample. For example, all ten might be found in one high-rise apartment; thus, no one else, for example those in single-dwelling units, would have a chance to come into the sample. Here's another example of the type of problem that could arise. A survey was conducted in England and after they randomly selected street intersections they interviewed only those houses located on corner lots. Consequently, they ended with a sample of people who lived on the corners of the blocks. Now, in England as in the United States, the corner lot is usually the most expensive lot. So they had built an automatic bias into their sampling design.' [From p. 6 of 'Aspects of the Census Bureau', by Vincent P. Barabba, pp. 1–10 of 'A numerator and denominator for measuring change', Technical Paper 37, U.S. Bureau of the Census, June 1975.] Note the use of 'representative' applied to people, not samples.

Some authors assert that, in order to avoid selective forces, each unit of the population must have an equal probability of entering the sample. That requirement may be confusingly called one of random sampling. For example: 'If the sample is to be representative of the population from which it is drawn, the elements of the sample must have been chosen at random . . . if we say that in choosing elements from the population no one element must be favored over any other element, we express fairly closely what is meant [by a random sample].' [From p. 44 of Johnson, Palmer O. and Jackson, Robert W.B., *Modern Statistical Methods*, Chicago: Rand McNally, 1959.] Aside from the problem of infinite populations, such an approach needs further analysis. First, random sampling in its basic sense requires not only equal probability of appearance in the sample but also independence (for sampling with replacement) or at any rate something specific and symmetric for joint probabilities. To illustrate the difficulty, consider a population of half men and half women; suppose that with probability 0.5 the sample consists of all the men and with probability 0.5 all the women. Then each person is equally likely to appear in the sample, but any particular sample will be wildly unlike the population in sexy ways, although in the long run, balance is retained on the average. The example is extreme for expository force, but less extreme variations are easy to envisage.

Second, sampling in which units have substantially unequal probabilities of appearance can be of great utility under appropriate analysis provided that the sampling probability structure is known. For the population mentioned above, suppose, for example, that the total sample

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

consists of a random sample of ten men from the male stratum together with a random sample of 100 women from the female stratum. Here every sample is most unlike the population in important ways, but it is easy to compensate for that in statistical analysis.

A charming quotation from P. Thionet is germane here: '... frequently ... [I had to deal with a] sample, a by-product of administration for example. The sample was not representative at all, because *distorted*; but individual data were supposed to be excellent, deep verifications being possible. It was desired to make these data comparable with the whole population ...' (p. 282). 'The technique presented here is called *reweighting*. It is used moderately with probability samples ... distorted by non-responses ... so that reweighted statistics remain random variates.... On the contrary ... [I was asked] to reweight non-random samples; our computations had nothing to do with the Probability Calculus, we were "escaping" from sampling theories' (p. 283). '[I] work in a country where trivial concepts relating to Statistical Methods are often misunderstood.... And it is difficult to say to our "inspecteurs des Finances" that they are wrong and have to learn sampling, when in fact they are very clever, very powerful and favorable to applied research' (p. 283). '[Our] samples were "*bad* samples", i.e. too beautiful samples. That is always the case when data are obtained from firms having genuine accounts (well-managed firms), from farms having accounts (accepting the help of some agricultural expert). These firms and farms are by no means representative' (p. 284). [From pp. 282–304 of 'Item analysis and reweighting', by P. Thionet, in Johnson, Norman L., and Smith, Harry, Jr., *New Developments in Survey Sampling*, New York: Wiley-Interscience, 1969.] Reweighting of the kind described by Thionet is, of course, frequently found and is no doubt sometimes effective. It always, however, leads us to ask about other, perhaps quite unknown, factors for which we ought in principle to reweight (or adjust) but cannot because we don't think of them, don't know necessary population information, or run out of resources for analysis.

Another issue is whether selective forces should be regarded as leading to biases in estimators or to increased sampling error. Usually one would lean toward estimation bias as the channel for evils of selectiveness, but the following quotation suggests the opposite: 'The [CPS] sample design ... is based ... on the ... most recent decennial census. As the time increases since the most recent census, the representativeness of the sample declines because of the shifts in population and, although the estimates continue to be unbiased, the sampling errors increase somewhat.' [From pp. 139–140 of 'Revisions in Current Population Survey', by Gary M. Shapiro, *Statistical Reporter*, February 1974.] Whether we go along with that analysis or not depends on what is reasonable to regard as stochastic ... a fascinating topic, but not one for the present discussion.

We end this section by noting an almost paradoxical usage. In most of the examples thus far given, selection is evil, polar to the holy grail of representativeness. There is, however, another domain of discourse in which that grail is sought by means of conscious selection in a purposive attempt to mimic the population. In survey work this is usually called quota sampling, and arguments can certainly be made in its favor, e.g.: 'It may be claimed, with some plausibility, that this purposive method is more likely to give us a sample which is *typical* or representative of the population than a random method ... [but] as the sample becomes larger the random sample becomes more and more representative of the parent, whereas, owing to bias, the purposive sample in general does not.' [From p. 382 of Yule, G. Udney, and Kendall, M.G., *An Introduction to the Theory of Statistics*, New York: Hafner, 1950, 14th Edition.]

We shall in Section 2.7 briefly return to this sense of representative sampling, and we shall discuss similar matters in our historical paper, Part IV of the sequence. An excellent critical analysis is given by Alan Stuart, who writes in part: 'There are several distinct criticisms to be made of representative procedures [like quota sampling]. In the first place, the agreement of even quite a large number of averages or percentages between sample and population in no way guarantees high accuracy in other respects, and very large biases are still possible. ...'

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

[From p. 613, vol. 13 of The *International Encyclopedia of the Social Sciences,* David L. Sills, Editor. Article on Sample Surveys II: 'Nonprobability Sampling' (pp. 612–616). New York: Macmillan and Free Press, 1968. See also The *International Encyclopedia of Statistics,* 1978.

2.3 *Miniature of the population*

The representative sample viewed as a miniature, or perhaps a mirror, of the population is closely related to the concept of non-selectiveness. Not surprisingly, the miniature or mirror terminology occurs in the statistical literature, but nearly always with qualifications. Samples fall short of being miniatures, according to the quotations in our collection, because of (1) selection, (2) sampling fluctuations, and (3) the effective impossibility of resemblance on many traits at once.

The strongest statement of a representative sample as miniature that we have found in the statistical literature appears in a text on psychological statistics: 'For statistical purposes it is necessary that the samples with which we work be what are called *representative samples.* This means that the important characteristics of the population are contained, in their proper proportions, in the sample . . .' (p. 121). '[A fairly large random sample is likely to be representative,] a small approximate "scale model" of the population' (p. 122). [From Senter, R.J., *Analysis of Data,* Glenview, Ill.: Scott, Foresman, 1969.]

Almost as strong is an older definition from a dictionary of statistical terms: '*representative sampling*: Any method of selection which will yield a sample possessing essentially the same characteristics as the larger population or universe from which it is drawn.' [From p. 146 of Kurtz, Albert K., and Edgerton, Harold A., *Statistical Dictionary of Terms and Symbols,* New York: Hafner, 1967, facsimile of the 1939 edition.]

Something like this is said in a book about Belgian linguistic censuses. After pointing out that sampling often permits more careful investigation than a census, the author says that le relevé partiel '. . . les limite à un échantillon numériquement restraint mais dont les éléments sont sélectionnés de façon à former un petit group parfaitement *représentatif* de l'ensemble à étudier.' [From p. 25 of Lévy, Paul M.G., *La Querelle du Recensement,* Brussels: Institut Belge de Science Politique, 1960.] A similar semicircular approach is taken in the following: 'Representative Sample – A sample, chosen by any means, which exhibits on a small scale the relevant characteristics of the population from which it came; that is, a sample which is, in fact, representative of the population.' [From p. 93 of Freund, John E., and Williams, Frank J., *Dictionary/Outline of Basic Statistics,* New York: McGraw-Hill, 1976.]

Now we turn to discussions of the inevitable impossibility of a perfect miniature. The sampling fluctuation problem is frequently mentioned, e.g., by '. . . samples are imperfect representations of the population in that they exhibit *random fluctuations* about the condition that would represent the exact image of the population.' [From p. 40 of Mandel, John, *The Statistical Analysis of Experimental Data,* New York: Wiley, 1964.] On the other hand, selection may well spoil the miniature: 'In order to yield exact information about a population, a sample would have to be precisely representative: a small-scale image of the population from which it was drawn, with all its frequencies proportional to those of its parent. . . . [Such a sample] can be constructed only after the properties of the population are known. . . . In exploring the unknown, the most we can do is assure ourselves that the sample used is not systematically *un*representative or *biased.*' [From p. 55 of Suits, Daniel B., *Statistics: An Introduction to Quantitative Economic Research,* Chicago: Rand McNally, 1963.]

Some treatments cover both selection and randomness together, e.g.: 'The representativeness of the data, or the degree to which the sample is like the population for which inferences are to be drawn, involves two considerations. First, the sample may differ from the population because the sample has been so drawn that it is biased. In the foregoing example of public polling it is conceivable that a sample of respondents may be chosen such that the sample is

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

quite unlike the population of voters about whom generalizations are to be made. Such an instance might occur if the sample were chosen entirely from telephone owners. . . . Second, the sample from the population may differ because of fluctuations which occur whenever a sample has been taken wholly at random either in a single population or in the subgroups of a stratified population. The fluctuation arising from the sampling process constitutes the framework of statistical inference. Provided the condition of random selection is fulfilled, this fluctuation, called *sampling error*, can readily be taken into account.' [From pp. 105–106 of Wert, James E., Neidt, Charles O., and Ahmann, J. Stanley, *Statistical Methods in Educational and Psychological Research*, New York: Appleton-Century-Crofts, 1954.] The quotation is too stringent in requiring randomness of the sampling, or simple stratification. At least in principle, any probability sampling method permits calculation of sampling fluctuation.

Stephan and McCarthy provide a thoughtful treatment: 'The first aim of most sampling procedures is to obtain a sample . . . that will *represent* the population from which it is selected. In other words . . . results . . . from the sample [should] . . . agree "closely" with results that would have been obtained . . . [by studying] the entire population. . . . This . . . idea has frequently been stated by saying that a sample should be a miniature population or universe.' 'Many different procedures . . . have been suggested and used for obtaining representative samples. Thus it seems fairly obvious that the sample should be scattered evenly throughout the population. . . . The foregoing remarks . . . [read as] though the quality of representativeness were absolute, that is, as though any given sample . . . [is] either entirely representative or entirely non-representative. . . . Actually, this is not the case. In the first place, a sample may represent the population with respect to one characteristic . . . and may not . . . with respect to other[s]. . . . In the second place . . . representativeness is also a matter of degree . . . the term "representative sample", by itself, can never be given a precise meaning. Accordingly we offer the following definition: A *representative sample* is a sample which, for a specified set of variables, resembles the population . . . [in that] certain specified analyses . . . (computation of means, standard deviations, etc. . . .) will yield results . . . within acceptable limits set about the corresponding population values, except that . . . [rarely] the results will fall outside the limits . . . the mere statement or claim that a sample is representative of a population tells us nothing.' [From pp. 31–32 of Stephan, Frederick F., and McCarthy, Philip J., *Sampling Opinions: An Analysis of Survey Procedure*, New York: Wiley, 1958.]

In the above quotation we have omitted for brevity concrete examples. Note how the Stephan-McCarthy approach leads from the miniature idea to one that might at first blush seem quite different: representativeness as permitting satisfactory estimation. We shall return later to discuss that viewpoint.

We end this section with the analysis and metaphor of Gilbert, Light, and Mosteller: 'In discussing both sample surveys and controlled trials, people often suggest that we not take random samples but that we build a small replica of the population, one that will behave like it and thus represent it. . . . When we sample from a population, we would like ideally a sample that is a microcosm or replica or mirror of the target population – the population we want to represent. For example, for a study of types of families, we might note that there are adults who are single, married, widowed, and divorced. We want to stratify our population to take proper account of these four groups and include members from each in the sample. Otherwise, perhaps by chance we would get none or too few in a particular group – such as divorced – to make a reliable analysis. This device of stratifying is widely used to strengthen the sample and bar chance from playing unfortunate tricks. Let us push this example a bit further. Do we want also to take sex of individuals into account? Perhaps, and so perhaps we should also stratify on sex. How about size of immediate family (number of children: zero, one, two, three . . .) – should we not have each family size represented in the study? And region of the country, and size and type of city, and occupation of head of household, and education,

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

and income, and . . . The number of these important variables is rising very rapidly, and worse yet, the number of categories rises even faster. Let us count them. We have four marital statuses, two sexes, say five categories for size of immediate family (by pooling four or over), say four regions of the country, and six sizes and types of city, say five occupation groups, four levels of education, and three levels of income. This gives us in all $4 \times 2 \times 5 \times 4 \times 6 \times 5 \times 4 \times 3 =$ 57,600 different possible types, if we are to mirror the population or have a small microcosm; and one observation per cell may be far from adequate. We thus may need hundreds of thousands of cases! Clearly, this approach is getting too fine for most purposes, and such an investigation will not be carried out except when enormous resources are available. We cannot have a microcosm in most problems. . . . Why isn't the microcosm idea used all the time? The reason is not that stratification doesn't work. Rather it is because we do not have generally a closed system with a few variables (known to affect the responses) having a few levels each, with every individual in a cell being identical. To illustrate, in a grocery store we can think of size versus contents, where size is 1 pound or 5 pounds and contents are a brand of salt or a brand of sugar. Then in a given store we would expect four kinds of packages, and the variation of the packages within a cell might be negligible compared to the differences in size or contents between cells. But in social programs there are always many more variables and so there is not a fixed small number of cells. The microcosm idea will rarely work in a complicated social problem because we always have additional variables that may have important consequences for the outcome.' [From pp. 218–220 of Gilbert, J.P., Light, R.L., and Mosteller, F., 'Assessing social innovations: an empirical base for policy', in Fairley, W.B. and Mosteller, F., *Statistics and Public Policy*, Reading, Massachusetts: Addison-Wesley, 1977.]

2.4 *Representative as typical*

The idea of a representative sample as one made up of typical units of a population is not found frequently in the statistical literature, but it has some strong, if critical, expositors. E.g.: 'A method of selecting a sample often employed in place of random sampling is to choose a sample which is "representative" with respect to certain known characteristics of the population. Thus one might attempt to choose for the sample that single county which agrees most closely with United States averages in respect to such selected characteristics as the per cent of the population engaged in agriculture, the per cent engaged in manufacturing, trade, and other industry groups, the proportion white and the proportion native-born. The same approach is sometimes used in choosing a sample of several counties by the device of first classifying the counties into groups, or strata, and then selecting a "representative" county from each stratum. The method is sometimes applied in sampling business establishments, families, or other units. It differs from stratified random sampling in that the actual selection of the units to include in the sample in each group is done purposively rather than by a random method.' [From pp. 71–72 of Hansen, Morris H., Hurwitz, William N., and Madow, William G., *Sample Survey Methods and Theory*, Vol. **1**, New York: Wiley, 1953.] This sense of 'representative sample' is close to one mode of specifying a particular sampling method, an approach that we shall discuss in a later section.

If the ideas of the above quotation be extended to other aspects of specific characteristics than averages, for example to second moments, then we begin to move towards the miniature concept. Yet, as we shall see from the Gini-Galvani story in Part IV, an approach that makes for typicalness of some characteristics – however defined – need not lead to typicalness for other characteristics. This point is made clearly by Hodges and Lehmann: 'Since the purpose behind taking a sample is to get information about the population, it seems natural to select a sample so that it will be representative of the population. For instance, a fruit grower trying to estimate the harvest from a grove of lemon trees might pick out a few trees that he judges to be typical of the grove as a whole, count the fruit on the selected trees, and then assume that the average

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

number of fruit per tree in the whole orchard is about equal to the average for the sampled trees. Again, in trying to forecast an election, we might select counties that have voted like the country as a whole in recent elections, and conduct interviews with voters in these counties. While this method of "purposive" sampling is superficially attractive, it has not worked well in practice. Biases of the selector creep in too easily when the items for the sample are selected purposively by the exercise of judgment. Also, items that are typical of the population in some respects need not be typical in others. For instance, counties that vote like the country on the last election may well not do so on the next, when the issues will be different.' [From pp. 35–36 in Hodges, J.L., Jr., and Lehmann, E.L., *Basic Concepts of Probability and Statistics*, San Francisco: Holden-Day, 1964.]

Kendall and Buckland, in their dictionary, give typicalness as one of their explications when they define 'representative sample': 'In the widest sense, a sample which is representative of a population. Some confusion arises according to whether "representative" is regarded as meaning "selected by some process which gives all samples an equal chance of appearing to represent the population"; or, alternatively, whether it means "typical in respect of certain characteristics, however chosen."

'On the whole, it seems best to confine the word "representative" to samples which turn out to be so, however chosen, rather than apply it to those chosen with the object of being representative.' [From p. 249 in Kendall, Maurice G., and Buckland, William R., *A Dictionary of Statistical Terms*, New York: Hafner, 1960.] The second paragraph is remarkable in its emphasis on the accomplished sample rather than the sampling method. Of course it is difficult to decide whether a specific sample closely resembles the population unless one knows a good bit about the population.

Typicalness is an ambiguous idea: it might mean a sample each of whose members was itself typical, or it might mean a sample typical as a whole. That latter sense is close to the miniature idea; the former sense requires explication of typicalness for a single unit, and if we can do that and find typical units, one might be contented with a sample of size one. (Some defenses of the case study approach make a similar argument.)

The ambiguity is implicit in the following description of past sampling practices: 'In the past . . . it was customary to expect researchers to examine . . . the material or populations they [choose] . . . to study, picking out deliberately what appeared to be the most representative sample. This rule has been . . . largely replaced . . . by random selection. . . .' [From p. 82 of 'Three extensions of sample survey technique: hybrid, nexus, and graduated sampling', by Frederick F. Stephan, pp. 81–104 of Johnson, Norman L., and Smith, Harry, Jr. (Editors), *New Developments in Survey Sampling*, New York: Wiley-Interscience, 1969.] Something like this ambiguity also appears in a 1950 text: 'The National Opinion Research Center has been able to make some accurate national estimates on the basis of a small sample of some 2500 cases distributed in representative areas of the nation. The particular . . . areas . . . are supposed to be representative localities – that is . . . their characteristics resemble the sectional average and the city-size average within each section. The characteristics tested against the section average are . . . [several characteristics, e.g. amount of schooling, rental levels, are described]'. [From pp. 272–273 of Parten, Mildred, *Surveys, Polls, and Samples: Practical Procedures*, New York: Harper, 1950.] Parten goes on to a somewhat puzzling quotation from an NORC document about how coverage of areas is replaced by choice of representative – i.e. typical – cities.

It is also worth noting that in this same book a number of other interpretations of representativeness are found: the rhetorical (p. 107), freedom from selection (p. 327), and identity with random sampling (p. 220).

Finally, we briefly describe Le Play's nineteenth-century budget studies: 'Le Play would seek out what he considered a "typical" or "representative" family which he would presumably

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

define as one in an especially pitiable condition . . . and would probe into every . . . facet of its existence. The findings would be published . . . as representative of working class families in the particular region. . . .' [From p. 3 of Pearl, Robert B., *Methodology of Consumer Expenditure Surveys*, U.S. Bureau of the Census, Working Paper **27**, 1968.] The usage Pearl describes is the mirror image of Emerson's *Representative Men*, in which Plato is the representative philosopher, etc. In both cases, of course, extremes are chosen to help make a point and to arouse sympathy or empathy.

The usage just described seems to us more Pearl's than Le Play's, so we place it in the section of our paper dealing with current usage, rather than in the historical section. We do not find, in our limited reading of Le Play, use of the term 'representative', although the idea and some of the difficulties are certainly there; for example, Le Play says: 'At first it is difficult to believe that a society spread over a vast territory can be comprehended through observation of a small number of families. . . . Man's nature shows infinite diversity. . . . Social constitutions [however] succeed in removing most of the inequalities which the diversity of human nature should produce. The shepherds on the frontier of Europe and Asia present characteristics so identical that it is sufficient to observe one family in order to know them all.' [From p. 469 of Zimmerman and Frampton (1935).] In fact, Le Play seems to have used both average (whatever that means) and extreme families for his detailed case studies. Zimmerman and Frampton (1935) say (p. 170) that: 'Sometimes Le Play took an average family of a typological community and sometimes a typological family of an average community.' 'Typological' seems to mean, or at least to include the meaning, 'extreme'. They also say, in a summary of criticisms of Le Play, that: 'He did not set forth clearly any one method. Sometimes his families were averages, sometimes typical, and sometimes selected cases used to prove a particular manifestation in its extreme forms (typological)' (pp. 65–66). Le Play could not prove that his families were representative of the groups which he studied' (p. 66).

Then, in a summary of responses to those criticisms, Zimmerman and Frampton say 'It is not necessary for an investigator to follow one method of study. . . . Le Play's case histories were systematic studies of certain families [from which] he developed theories somewhat unique and original for their time (p. 67). 'It was not essential . . . to prove that his families were representative. He could hold that most social systems were so omnipresent that a family, even though not typical in all respects, would reflect the [culture's] chief factors of social organization. . . . Further still, the 'typological' method calls for extreme rather than average expressions of a particular social trait. If by using this method Le Play discovered something not ordinarily observed about the . . . society, he should be credited. . . . The method of Le Play may be combined with statistical investigations . . . to satisfy other purposes . . .' (pp. 67–68).

2.5 *Coverage of the population's heterogeneity*

One concept of representativeness for a sample, or for a sampling method, calls for breadth and heterogeneity of the sample members relative to the population. Perhaps the mildest form of that requirement is put forward by Tippett: 'It is obvious that the sample can be representative of the population only if all parts of the population have a chance of being sampled.' [From p. 347 of Tippett, L.H.C., *The Methods of Statistics*, Fourth edition, New York: Dover, 1952.]

A stronger form of the heterogeneity idea is described thus: 'This condition [representativeness] is best achieved by getting the sample from as many parts of the population as possible. For example, if fruit in an orchard is to be judged from a sample of 1000 apples, it is better to take 50 apples from each of 20 well-distributed trees, than to take 250 apples from each of four trees.' [From pp. 31–32 of Wadley, F.M., *Experimental Statistics in Entomology*, Washington, D.C.: U.S. Dept. Agriculture, 1967.] We must add that Wadley prefixes the above quotation

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

by saying that we 'judge representativeness partly by reproducibility; . . . if *sampling error* is low, we believe our samples to be representative'. Wadley also warns about selection, so for him representativeness seems to mean low bias plus low sampling error, perhaps a counsel of perfection.

A notable, although possibly circular, usage under the heterogeneity rubric appears in a discussion of sampling errors for the 1970 U.S. Census: 'The variance calculations were made for a set of representative items rather than for all items tabulated in the census. One of the standard tabulations in the census contained 834 cells including both population and housing items and appeared to cover a reasonable representation of all items being tabulated.' [From p. 73 of Joseph Waksberg, Robert Hanson, and Peter Bounpane, 'Estimation and presentation of sampling errors for sample data from the 1970 U.S. Census', *Bulletin of the International Statistical Institute*, **45** (Book 3) 66–82 (1973).]

That quotation leads naturally to the idea of coverage, a strong form of the heterogeneity requirement, as presented in our earlier papers. A sample exhibits coverage if it contains at least one member from each set of a relevant partition of the population. Thus coverage implies heterogeneity with respect to its partition, but stands in sharp contrast to the miniature idea because coverage carries with it no implication about relative frequencies in the sets of the partition.

A recent survey of sample survey practices, for example, used purposive sampling with an eye to coverage relevant to size of survey, kind of survey taker, mode of data gathering, etc. The report on this meta-survey includes a tentative design for a possible much larger follow-up; the report says: '[A purpose] of the study . . . is to construct a sampling frame from which to select representative surveys from different sectors of the survey-taking population.' [From p. ix of Bailar, Barbara A., and Lanphier, C. Michael, *Development of Survey Methods to Assess Survey Practices*, Washington, D.C.: American Statistical Association, 1978.] This quotation refers to a hypothetical future survey; Bailar and Lanphier do not appear to use the term 'representative' for their own pilot work. It may be ironic that a journalistic description of their work brings in the term twice: '. . . what they [Bailar-Lanphier] eventually concentrated on was the actual conduct of a representative sampling of surveys . . .'; '. . . since the sample is described as representative . . . it may be assumed that the findings are fairly applicable to a broad range of government-sponsored surveying'. [From p. 4 of *Science & Government Report*, 1 Dec. 1977.]

In a completely different setting, Harry V. Roberts uses something like the coverage idea as he discusses a possible method of forming committees from a population of people for an activity that requires fairness, the image of fairness, and special abilities. (Roberts has in mind the selection of disciplinary committees from a university's faculty.) He summarizes as follows: '(1) Select a panel of "moderate" size by proportional stratified random sampling. (2) Subdivide the entire panel into non-overlapping committees, trying to make these committees as comparable as possible in whatever respects are deemed relevant to "representativeness". . . . (3) Select actual committees, as needed, by simple random sampling without replacement.' [From p. 20 of Harry V. Roberts, 'Committee selection by statistical sampling', *The American Statistician*, **25**, 18–20, (February, 1971).]

We note that heterogeneity and coverage are relatively unusual usages in the statistical literature. An older meaning, however, had some attention, namely that a sample is more nearly representative as it includes a larger fraction of the population. This meaning is described by Dalenius (1953) in one of the few recent statistical discussions of representativeness, and in Part IV we shall point to the use of this largeness idea by Carroll D. Wright, founder of the U.S. Bureau of Labor Statistics.

We note that the largeness idea appears in at least one totally different setting: 'The writers who supply the comments on the following five poems belonged to an audience of the same

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

type gathered two years later. Only a few of the original numbers remained. The change greatly increases the representative character of these extracts.' [From p. 112 of I.A. Richards, *Practical Criticism*, New York: Harcourt Brace, 1956. First published 1929.]

2.6 *'Representative sampling' as a vague term that is then made precise*

In 'Representative Sampling II' we presented (Section 0) the idea of 'representative sampling' as an initially vague term that would be made precise in the same discussion. We drew analogues with such a term as 'variability,' which might be used vaguely in the opening paragraphs of an exposition, and then be made precise later, say as standard deviation. Indeed one of the examples of this usage in 'Representative Sampling II' came from the statistical literature.

We have found a few examples in the statistical literature of this vagueness followed by precision. A.S.C. Ehrenberg, to choose one, says in his recent textbook that '. . . just any subset of the population will not do. A sample must be more or less representative of the population being studied'. [From p. 284 of Ehrenberg, A.S.C., *Data Reduction/Analysing and Interpreting Statistical Data*, London: Wiley, 1975.] That is surely a vague usage. But then, on p. 286, we find: 'A sample is a selection of objects or measurements taken from a specific population of such items. The aim is to make the results from the sample tell us more or less what we would have found by measuring the whole population. There are various ways of selecting a sample, but only with random sampling is it possible to know how representative the sample results are likely to be'. That begins to be more nearly specific, although we might bridle at the use of 'only with'. Then on p. 288, random sampling is explained in terms of well mixed slips of paper in a hat.

Something like the vague term made precise is used by M.G. Kendall in a descriptive article about the World Fertility Survey. He speaks early on of '. . . nationally representative, internationally comparable, scientifically designed and conducted sample surveys regarding human fertility'. [From p. 379 of M.G. Kendall, 'The World Fertility Survey/An international programme of fertility research', *Bulletin of the International Statistical Institute*, **45** (Book 4) 379–393, discussion 394–401, 1973.] Then on p. 381, Kendall says that the individual country surveys 'will ideally employ a national probability sample . . .', a step along the road to specification. The article makes it clear that later expositions of particular national surveys will define the national probability sampling methods used in each.

We end this section with an apposite quotation from W.E. Deming, published by Howard L. Jones: 'What we select are not representative samples, but probability samples.' [From p. 488 of 'Inadmissible samples and confidence limits', by Howard L. Jones, *Journal of the American Statistical Association*, **53**, 482–490, 1958.]

2.7 *Representative sampling as a specific sampling method*

In the last section we gave examples of an initial vague use of the term 'representative sampling', followed by later specification. Closely allied to that is initial specification of a sampling method as explication of the representative idea.

A quotation from Ehrenberg defines 'representative sampling' simply as random sampling, and Ehrenberg has definitional friends. For example, Herman Chernoff writes: 'A sample will be said to be *representative* of a population if the distribution of the data is precisely that of a random sample with replacement from the population.' [From p. 154 of 'A compromise between bias and variance in the use of nonrepresentative samples', by Herman Chernoff, in Olkin, Ingram and others, (eds) *Contributions to Probability and Statistics/Essays in Honor of Harold Hotelling*, pp. 153–167, Stanford, 1960.]

A second example comes from a different part of the statistical literature: 'We can obtain a satisfactory estimate . . . by taking a representative sample, but we must be certain that the

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

sample is a random one. ... if each citizen is represented by a numbered bean in an urn, a sample of randomized beans can be taken to select a representative random sample of citizens. ...' [From p. 4 of Lewis, Alvin E., *Biostatistics*, New York: Reinhold, 1966.] Perhaps Lewis confuses the exposition somewhat, as we read it, by saying on the next page that '... we intuitively realize that the larger a sample is, the more likely it is to be representative'. Evidently for him the two ideas of a specific sampling plan (random sampling) and of a miniature of the population are intertwined.

C.A.B. Smith finesses the problem by skipping a general definition and giving random sampling as the simplest example of representative sampling: '... we must make sure that the sample is representative of the whole population. The simplest way ... is to choose a set of members of the population *at random*.' [From p. 241 of Smith, Cedric A.B., *Biomathematics*, volume 2, London: Griffin, 1969.] He also requires that the sample be very small compared to the population size. Later on [p. 353] he repeats the idea: 'For any conclusions to be reliable, the sample selected must be representative of the population, e.g. a random sample. More errors are made through using unrepresentative samples than through incorrect methods of analysis.'

So much for random sampling in the strict sense. Some authors use 'representative sampling' to mean 'stratified sampling', itself of course a generalization of random sampling. For example: 'Sometimes ... a more complex form of random sample called the representative or stratified sample gives greater accuracy.' [From p. 101 of Tippett, L.H.C., *Statistics*, 2nd edition, London: Oxford, 1956.]

By far the most frequent of the specific sampling plans, however, is the specialization of stratified sampling in which the subsample sizes from the strata are proportional to the stratum sizes. For example: 'It is of interest that for specified total sample size and with equal σ_i^2's the value of $\sigma_{\bar{X}_s}$ is a minimum if the sample sizes n_i are proportional to M_i, the strata sizes. Such an experimental design is called a *representative sample*.' [From p. 184 of Dixon, Wilfrid, J. and Frank J. Massey, Jr., *Introduction to Statistical Analysis*, New York: McGraw-Hill, 1969.] Or again: 'Quite often ... one knows the ratios N_j/N, while the N_j and the $\sigma_j(V)$ are not known. In such cases it has been an accepted practice to choose the n_j proportional to N_j/N ... [and such a] stratified sample ... is called a *representative sample*. ...' [From pp. 183–184 of Birnbaum, Z.W., *Introduction to Probability and Mathematical Statistics*, New York: Harper, 1962.] Yet again, 'When, in practice, there are indications that individual lots may be stratified in quality, it is of course best to select a "representative" sample, one such that each stratum or subportion of the lot is proportionately represented by a subsample that is selected by a random operation.' [From p. 27 of Dodge, Harold F. and Romig, Harry G., *Sampling Inspection Tables*, New York: Wiley, 1944.]

Cochran, Mosteller, and Tukey, in their study of the first Kinsey Report, discuss representativeness around pp. 313–314. As we read their discussion, they require for representativeness a probability sample in which each population individual has an equal chance of appearing. They give several examples, including one that might be regarded as stratified sampling with subsample sizes proportional to stratum sizes; hence we place the discussion in this section.

Two conditions for representativeness are laid down; the first is that the sampling be probabilistic, and the second that of equiprobability for individuals: '(1) the particular persons in the sample are there by accident, and ... (2) ... each person in the U.S. has *an equal chance* of entering the sample ...' [From p. 314 of Cochran, William G., Mosteller, Frederick, and Tukey, John W., *Statistical Problems of the Kinsey Report on Sexual Behavior in the Human Male*, Washington, D.C.: American Statistical Association, 1954.]

One possible problem with the position just described, as compared, for example, with the requirement of proportional stratification, is that some sampling plans satisfying the Cochran–Mosteller–Tukey condition are intuitively representative in a long-run average sense but have specific realizations that are always wildly unrepresentative. An example is the probability

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

sampling plan described in the section on selective forces, a plan leading either to a sample of all women or all men.

The specific sampling plans thus far discussed in this section have been probabilistic ones. There is another path of usage in which 'representative sampling' means purposive sampling, often quota sampling, so that the samples match the population in some specific characteristics deemed important, but without any explicit random structure. See the discussion at the end of Section 2.2.

An early paper by S.S. Wilks shows the tension between the non-probabilistic concept just mentioned and Wilks' natural desire for randomness somewhere to justify the inferences he wishes to make. He says: 'The newer method of polling is based on a few simple principles of ... "representative" or "stratified" sampling. [This] ... method ... requires that the sample of individuals to be polled be a "properly balanced cross section" of the various important groups of individuals which form the population. [That is,] ... the important groups defined by some "relevant" classification of ... individuals ... are represented in the sample in proportion to the [sizes of the] ... population groups. ... For example, a sample from a population of eligible voters is representative with respect to sex if it contains as many men as women who are at least 21 years old.' [Pp. 262–263 of S.S. Wilks, 'Representative sampling and poll reliability', *Public Opinion Quarterly*, **4**, 261–269, 1940.] Wilks goes on to say that a sample can be made representative with respect to 'any number of mutually exclusive population groups', and that, for national public opinion polls, it has been found practically sufficient to use samples that are representative with respect to (1) geographical region, (2) city size plus rural partition, (3) age, sex, color, and economic status within the groups of (2). Apparently the representativeness required under (3) is marginal, not joint.

It is difficult to achieve representativeness, especially for economic status, because the population distribution of economic status is poorly known. Thus so-called check data (automobile ownership, voting in last election, etc.) are used as proxies. If the sample is found unrepresentative, for example with too high a proportion of automobile owners, 'steps are taken to reallocate the sampling with respect to economic status [p. 264]'. The procedure is one of trial and error.

Then comes the bridge to statistical orthodoxy, a bridge in several parts: '... it is assumed that the distribution of opinion in the sample will approximate that in the population [as] ... the distribution of "check data" in the sample approximates that in the population. The basis for this assumption is that the [classification] ... into subgroups is ... "relevant" ..., that is, ... it yields subgroups within which opinion is ... more homogeneous than ... if all groups were thrown together, and furthermore that the sampling is representative with respect to the subgroup and is random within each subgroup' [p. 264].

There are three main ideas in this quotation. First, the idea of homogeneity within subgroups, the fundamental idea behind stratified sampling; second, that subgroups (i.e., strata) are sampled proportionately to their sizes; and third, that sampling is random within subgroups. The list is needed by Wilks for the confidence limit procedures[1] he treats later in the article, although the earlier approach to get a 'balanced cross section' precludes randomness in the ordinary sense.

Then Wilks says that: 'The strongest reason, perhaps, for making a sample representative with respect to important population subgroups is to increase the assurance that the "yes" percentages ... actually cluster around the true "yes" percentage in the population. In other words, representativeness decreases the risk of getting a biased sample' [pp. 265–266]. The explanation seems circular, but even Homer nods.

An excellent way to end this section is the following quotation from Leslie Kish's important

[1] Wilks here gives procedures based on simple random sampling rather than stratified sampling, presumably for simplicity of exposition, although he does not say so.

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

textbook. It helps explain why we have found relatively few uses of 'representative sampling' in the contemporary statistical journal literature: '*Representative sampling* is a term easier to avoid because it is disappearing from the technical vocabulary. At different times it has been used for random sampling, proportionate sampling, quota sampling, and purposive sampling. In general, if often denotes the aims of representing a population well with a sample; and this is the sense of the terms *population sampling* and *survey sampling* in our vocabulary.' [From p. 26 of Kish, Leslie, *Survey Sampling*, New York: Wiley, 1965.]

2.8 *Representative sampling as permitting good estimation*

We had earlier thought that the statistical literature would contain many characterizations of representative sampling as permitting virtuous, or at least satisfactory, estimation of population characteristics. Vague as that is, it still is a fundamental idea of statistical inference. Virtue of estimation might be explicated in various ways, perhaps most generally by identifiability, i.e. by requiring that such and such a parameter be a functional of the sample's distribution. That is a requirement at once weak and strong. Or virtue might be described in terms of little or no bias, in terms of low sampling error, or in yet other terms.

Of course these ideas, except perhaps for identifiability, have appeared in many of our quotations, but explicit mention of good estimation seems to be infrequent. It half-appears in the quotation from Stephan and McCarthy in Section 2.3 on miniatures, and we find a more explicit version in another text on surveys: 'A sampling method, if it is to provide a sample representative of the population, must be such that all characteristics of the population, including that of variability among units of the population, are reflected in the sample as closely as the size of the sample will permit, so that reliable estimates of the population characters can be formed from the sample.' [From p. 1 of Sukhatme, Pandurang V., *Sampling Theory of Surveys With Applications*, Ames, Iowa: Iowa State U. Press, 1954.]

2.9 *Representative as good enough for a particular purpose*

An infrequent but noteworthy meaning of 'representative sample' is 'good enough for our purposes', a meaning that might well not require a properly executed probability sample of the population. For example, suppose that physicians in a region believed that practically no patients having a certain kind of burn were admitted to hospitals. Then if a substantial sample of admissions with burns had 50 per cent response, of which 15 per cent had the special kind of burn, that would show us that the population has a serious percentage of such cases. Clearing up the non-response could only enlarge the result. Similarly, if the physicians thought that all patients with a particular kind of burn developed a particular symptom, but a sample showed that a number did not, that would be good enough to settle the particular issue.

Thus any sampling showing that something thought to be rare or absent occurs with some frequency, or that something thought to be universal is not so, may be good enough for the purpose of correcting our views.

If one wanted to know whether kangaroo temperature ran about 10°, 100°, or 1000°F, a sample of one would do nicely. Thus for order-of-magnitude estimation a small, poorly drawn, or otherwise inadequate sample may also be good enough and hence adequately representative in this sense.

Although authors may sometimes have this meaning of representative sample in mind, we have not seen it clearly stated in the literature. In a letter to one of the authors (WHK), John R. Platt writes about the purpose of an article: 'It was not our intention to do this particular job – as we tried to emphasize in our opening paragraphs. We were not trying to propose a complete list of "Nobel-prize caliber" social scientists, or to show the complete branching trees of pre- and post-influences. Our experience and correspondence indicates that maybe 80 per cent

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

overlap is all one can expect between such lists made by small teams with different knowledge and interests – which is one reason for relying so heavily on *Intl. Encycl. of Soc. Sci.*, as an established authority, even if it should also need correction or might have different evaluation in many instances. No: our purpose was just to get a sufficiently representative sample of these contributions and social inventions and innovations to answer fairly accurately our primary question, which is, where do such innovations come from? and after that, how long do they take to get generally accepted or implemented? From our correspondence, I have the feeling that your list, or anyone else's, even if they differed from ours in 20–50 per cent of the names, would still have almost the same answers to these questions. But maybe it is time for all these questions to be examined much more carefully by more complete studies that can go far beyond these initial explorations.' As the last sentence shows, this sort of sampling may still leave misgivings.

A second example that may illustrate the point is: 'People ask us what kind of a sample we get. Well, our objective is to get as close to a 100 per cent complete coverage as is possible and, in fact, we are generally successful in making contact with about 90 per cent of all the occupied housing units in a city. The contact rate will go higher in certain more stable neighborhoods; it may be lower in poorer neighborhoods and those with a fast turnover, but we've never seen it below about 65 per cent. Those who are expert in statistics and sampling tell us that this should be a very representative sample.' [From p. 54 of 'Profiles of small urban area change,' by Richard Hanel; pp. 53–76 of 'A numerator and denominator for measuring change', Technical Paper 37, U.S. Bureau of the Census, June 1975.]

In a third example, the author seems to mean by 'representative data' those that will not be seriously, perhaps deliberately, misleading. Edgar S. Dunn, Jr., writes, 'The professional statistician tends to see the problem of entity representation largely as a problem of how to draw observations from an environmental universe so that the sampled set provides a reliable representation of all the possible observations bounded by some conceptual entity [p. 133]. The second misuse [related to personal privacy] involves data about individuals that is accidentally misrepresentative and that . . . unjustly impinges upon his normal rights and privileges . . . [p. 189]. Apart from accidental misrepresentation or the deliberate misuse of representative data about individuals, the practice of the *deliberate misrepresentation* of data with the *intent to misinform* is widespread in our society. [p. 189]. [Image manipulation] has the effect of concentrating economic and political power in the hands of those . . . with sufficient resources to command representative data and the sophistication to filter out misrepresentative data' [p. 189]. [From Dunn, Edgar S., Jr., *Social Information Processing and Statistical Systems – Change and Reform*, New York: Wiley, 1974.]

2.10 *Linguistic cousins of representativeness; other variations*

We describe in this section two further brief topics related to representativeness. First, we have found a few examples in which one starts with the sample and then asks of what population it is representative. For example: 'In medical research, investigators are usually faced with existing samples and do not have the opportunity to take a probability sample from the universe about which they want to draw conclusions. For example, in a study on cancer of the breast, usually a group of women who have cancer of the breast and are in a particular hospital at a given point in time compose the sample studied. The question then becomes, "From what statistical universe is this a representative sample?" In order to have confidence in the results of studies based on this kind of a sample, as many pertinent characteristics in the sample as possible should be tested against known characteristics in the target population. The more nearly the sample corresponds to the universe in known characteristics, the more confidence an investigator can have in drawing conclusions about the unknown characteristics.' [From p. 6 of Schor, Stanley, *Fundamentals of Biostatistics*, New York: Putnam's, 1978.] Or: 'The basis of all . . .

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

scientific inferences from empirical material is the assumption that a series of observations is a representative sample from a population ... [In] medicine one seldom starts with drawing a sample from a population [but rather with]... a series of observations resulting from numerous external conditions. ... A relationship between population and sample must be established [and in such medical cases]... this can be only reached by searching for that population from which the survey in question can be a representative sample.' [From p. 235 of 'Use of non-representative surveys for etiological problems', by S. Koller, pp. 235–246 of Johnson, Norman L., and Smith, Harry, Jr. (eds), *New Developments in Survey Sampling*, New York: Wiley/Interscience, 1969.] Of course the problem of handling a sample of unknown provenance is always difficult, in medicine, geology, economics, or anywhere. It seems to us, however, that the problem is far from understood and that definitions do not as such help with it.

Second, some later authors have harked back to the old term 'representative method' which will be discussed in our forthcoming historical Part IV. For example: 'When it is desired to obtain information as to the characteristics of some large population, such as the inhabitants of a country, the fir trees of a district, a consignment of articles delivered by a factory, etc. it is often practically impossible to observe or measure every individual in the whole population. The method generally used in such situations is known as the *representative method*: a sample of individuals is selected for observation, and it is endeavored to make the sample as representative as possible of the total population. The observed characteristics of the sample are then used to form estimates of the unknown characteristics of the total population.' [From p. 331 of Cramér, Harald, *Mathematical Methods of Statistics*, Princeton: Princeton University Press, 1946.] And: 'In the mathematical statistics literature one very frequently speaks of an investigation by the *representative method*, but this notion is not defined precisely. By the term representative method, we shall mean a random sampling method, where the sample may or may not be simple.' [From p. 504 of Fisz, Marek, *Probability Theory and Mathematical Statistics*, New York: Wiley, 1963.] The term 'representative method' is important in the history of statistics, and it does appeal at first to the intuition. Nonetheless, it contains such circularities and ambiguities that we do not hope for its revival.

3 Mathematical methods

3.1 *Introduction*

Our papers have described and illustrated several senses of representative sampling. By and large the discussion has not been in a mathematical or technical mode, although some of the senses discussed could lead – have led – to relevant mathematical developments, for example the idea of stratified sampling and the consequent questions of design and analysis.

In this section, we sketch briefly, without any intent of completeness, a few technical or mathematical directions related to senses of the expression 'representative sampling'.

3.2 *Measuring representativeness*

Given two populations or two probability distributions on the same space, what are sensible ways of defining or measuring the agreement between them? That is, what are sensible ways of defining how representative each is of the other? The other side of that question is to ask for sensible ways of measuring or defining how much the distributions *differ*; what might be a reasonable distance function for pairs of distributions? Whether we measure how full a cup is, or how empty, is not of the essence. Much depends on the nature of our interests and thus on the loss function associated with each situation. We sketch a few possibilities.

The simplest example may be that of two populations made up of unstructured categories numbered $i = 1, 2, ..., I$, with associated probabilities p_i and q_i respectively; $\Sigma p_i = \Sigma q_i = 1$. One might say that these populations represent each other perfectly if $p_i = q_i$ for all i. When

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

some $p_i \neq q_i$, the extent of disagreement may be approached on the basis of many summary functions, for example,

$$\Sigma \mid p_i - q_i \mid,\ \Sigma \mid p_i - q_i \mid /(p_i + q_i),\ \Sigma(p_i - q_i)^2,$$

with the possibility of all kinds of weighting or other complications, for example, cross-deviation terms.

An approach that seems at first different is to consider the worst set, the maximum of $\mid \sum_A p_i - \sum_A q_i \mid$ where A ranges over all subsets of the integers $1, 2, \ldots, I$. This maximum disagreement in probability turns out to be just half of $\Sigma \mid p_i - q_i \mid$, the so-called L_1 distance.

Choice within this over-rich array of possibilities presumably depends upon the purposes of an investigation, and on such considerations as (1) relevance, (2) clear and useful interpretability, and (3) mathematical manipulability. A discussion is given by Goodman and Kruskal (1959), pp. 143–147.

From unstructured finite discrete distributions, one is naturally led to general discrete distributions, perhaps with structure (linear ordering, even with numerically valued categories), and to other kinds of distributions, e.g. continuous distributions on the line, multivariate distributions, etc. A few relevant references are Adhikari and Joshi (1956), Kuipers and Niederreiter (1974), and Rao (1949).

Beyond direct population comparisons, one is led to sampling considerations and to inferential questions. In the simple categorical case, one might consider multinomial sampling and ask about the degree to which its results represent a given population, known or hypothetical. Thus one would think about, say, the distribution of $\Sigma \mid p_i - \hat{p}_i \mid$, where $\hat{p}_i$ is the observed proportion of counts in category i. A great variety of derivations, manipulations, and approximations arises here. For populations with two categories, see, for example, Goodman and Kruskal (1963, 1972) and Chapman (1978). For more general cases, see the discussion of goodness of fit by David (1978).

Indeed with only a bit of imperialism we might extend our scope to most of statistics. After all, interest in representativeness may be special to some particular aspect of a distribution, the mean, say, or the variance, etc. Then nearly all statistical procedures may be regarded as questions about representativeness of particular kinds. We refrain from such expansiveness.

3.3 *The Sobel-Huyett approach*

Sobel and Huyett (1958) have a special way of looking at the representativeness of a sample with respect to an unstructured finite categorical distribution. Since this is one of the few technical papers in the recent literature that deals explicitly with representativeness, we present its central idea. In the notation of 3.2, Sobel and Huyett look at the inequalities

$$\mid p_i - \hat{p}_1 \mid \leqq \beta_i, \quad i = 1, \ldots, I,$$

for given positive numbers β_i. If the inequalities are all satisfied, then the sample is said to be representative (relative to the β_i) of the population. The paper deals primarily with the sample size required to achieve a given probability that a multinomial sample will be representative, and we give two examples from the Sobel-Huyett results.

Example. With $I = 2$, each $p_i = 0.5$, and each $\beta_i = 0.2$, to achieve probability 0.75 that both inequalities hold, a sample of three works and so does a sample of six, but *not* samples of four or five; samples of nine or more work. This is a simple case because the chance event of interest may be written $0.3 \leqq \hat{p}_i \leqq 0.7$ and the calculations or binomial table entries are easy.

Example. With $I = 4$, each $p_i = 0.25$, and each $\beta_i = 0.1$, then to achieve probability 0.75 that all the inequalities hold, the minimum sample size required is 52. This does not imply that

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

53 will work, but there is a sample size, n, beyond which all sizes work. The value of n is not known to us.

Sobel and Huyett provide extensive tables for several such examples, and they extend the treatment of representativeness.

N.D. Neff and J.I. Naus have carried out unpublished work on an analogous problem for a population taken as a continuous distribution on the line. Without loss of generality take it to be the uniform distribution on [0, 1] and consider a random sample of size n. Let $M_{y,p}$ be the number of observations in the interval $[y, y+p]$. Then, for given α and β, how large must n be so that

$$\Pr\{\mathrm{Max}\mid n^{-1}M_{y,p}-p\mid \leqq \beta\} \geqq \alpha$$

where the maximum is over $0 \leqq y \leqq 1-p$?

Example. With $p = 0.1$, $\beta = 0.3$, and $\alpha = 0.95$, we want probability 0.95 that no interval of population probability 0.1 has a sample frequency deviating more than 0.3 from 0.1 The least n that works is 13; 14 does not work, but 15 and all greater values do.

Note the similarity of this approach to that of the Kolmogorov–Smirnov test. For Neff and Naus, the interval is of constant length; for Kolmogorov-Smirnov, the intervals are all half-infinite to one side.

3.4 *Outcome variables and comparison variables*

We pointed out in 3.2 that one might focus on representativeness only as regards a single aspect of a distribution, say the mean. Means can disagree wildly while probabilities look much the same (move a tiny probability far to the side), or conversely means may be the same while probabilities look very different (e.g. two normal distributions with zero mean and very different variances).

Yet more to our present point may be the relationship between outcome variables and comparison variables. We shall see in Part IV how Kiaer and others examine representativeness by comparing a sample with a population on variables known for the population (e.g. age, sex, marital status) although the purpose of the sampling was to examine other variables *not* known for all the population (e.g. assets, physical disability, blood pressure). We call the second set *outcome* variables, and the first *comparison* variables.

This approach of checking that comparison variables match roughly and then analyzing outcome variables is frequent in medicine, sociology, public policy, and many other fields in which census-like information permits comparison of a sample with the population on some – typically background – variables.

We note in passing that authors rarely find the comparison so frightful that they are deterred from analyzing the outcome variables in the simplest possible way. That remark is not pure irony, for there is an obvious selective force: authors who are deterred may be less likely to publish.

Underlying such comparisons is a real threat; the author fears that the comparison variables influence the outcome variables, or reflect a common influence, and thus analysis of the outcome variables is on a firmer base if the sample resembles the population in the comparison variables. Analogous considerations arise when two or more samples correspond to different treatments, and we worry about how much the samples may differ intrinsically in ways that will be confounded with treatment. (In both situations, completely unobserved variables might show us large differences between sample and population, or between two samples; avoidance of that danger is one motivation for probabilistic sampling procedures.)

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

A natural step is to examine the relationships between the comparison and the outcome variables, and to make some covariance-like adjustment to handle any disparity. That may not be as simple as it seems, and in many circumstances a bias towards over-adjustment is present. See Campbell (1978) p. 303.

Of course the comparison variables may have no relationship with the outcome ones. Disagreement of the sample and population on the comparison variables then does not by itself disturb representativeness of the sample with respect to the outcome variables; but neither would agreement support representativeness! The analogous point obtains for two or more samples with different treatments.

It is worth mentioning too that comparisons of a sample with a census should be made carefully because censuses themselves are subject to biasing errors of non-response and measurement.

3.5 *Representativeness in estimation*

We end with an example showing how representativeness may be relative to an estimator as well as to sampling method. Suppose we are interested in heights of individuals, heights measured by an instrument with normal additive error structure, mean zero, and variance σ^2. Furthermore, individuals' true heights are taken to vary among themselves normally with variance $a^2\sigma^2$ (a is known), and the error of measurement and true height deviation are supposed independent.

Two individuals are drawn at random from the population and have measured heights x_1 and x_2. A good estimator of the first individual's height is x_1, but a better estimator is

$$x_1^* = \frac{x_1\,(1+a^2)+x_2}{2+a^2};$$

see Mosteller (1948).

If $a = 0$, so that all true heights are the same, then $x_1^* = (x_1+x_2)/2$, and x_2 is, so to speak, just as representative of individual 1 as is x_1. Conversely, as a^2 becomes very large, x_1^* approaches x_1: the fresh information from measuring x_2 becomes useless in terms of x_1.

Perhaps $2/(2+a^2)$ is a reasonable measure of the representativeness of individual 2 for individual 1, since it is the proportion of weight given x_2 relative to the weight when $a = 0$. Thus if $a = 1$, the representativeness measure is 2/3. More realistically in the case of height, σ might be $\frac{1}{8}$ inch and $a\sigma$ about 2 inches, giving $a = 16$ and representativeness index $2/(2+256) = 1/129$, a rather small number. Thus representativeness for a specific purpose might depend in part on the statistics used.

We note that the ideas of this subsection are related to relatively recent investigations of empirical Bayes methods and of Stein-James estimation.

Acknowledgments

We express our gratitude to those who have suggested examples, commented on the analysis of this paper, and otherwise provided stimulus and aid: Keith Baker, Tore Dalenius, W.E. Deming, John Emerson, Miriam Gasko-Green, Eric Hamp, David C. Hoaglin, Patricia Hume, Joseph I. Naus, Anita Parunak, Nancy Romanowicz, I.R. Savage, Susan Shott, Stephen Stigler, Michael Stoto, Ronald Thisted, Christine Waternaux. We also appreciate the generous aid of others at various stages in the preparation of these papers: Colin Atkinson, Vincent Barabba, Petter Jakob Bjerve, Howard Frazier, Katherine Godfrey, Joseph Ingelfinger, Tanja Kazan, Margaret E. Martin, Ann F. S. Mitchell, Marjorie Olson, Harry V. Roberts, David L. Sills, Keith Soper, Robert E. Streeter, Judith M. Tanur, Lawrence A. Thibodeau, Charles Turner, Cleo Youtz.

The preparation of this paper was aided by National Science Foundation grants NSF SOC 72-05228 AO4 to the University of Chicago and NSF SOC 75-15702 AO1 to Harvard University.

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms

References

Adhikari, B.P. and Joshi, D.D. (1956). Distance-discrimination et résumé exhaustif. *Publications de l'Institut de Statistique de l'Université de Paris*, **5**, 57–74.

Campbell, Donald T. (1978). Experimental design III. Quasi-experimental design, in *International Encyclopedia of Statistics*, edited by William H. Kruskal and Judith M. Tanur, New York: Free Press, 299–305.

Carritt, E.F. (1937). An ambiguity of the word 'good'. *Proceedings of the British Academy*, **23**, 51–80.

Chapman, Douglas G. (1978). Counted data, in *International Encyclopedia of Statistics*, edited by William H. Kruskal and Judith M. Tanur, New York: Free Press, 76–87.

Dalenius, T. (1953). Representativitet, *Statistisk Tidskrift*, Statistiska Centralbyrån, Stockholm. Swedish, with English summary, 274–284.

David, H.T. (1978). Goodness of fit, in *International Encyclopedia of Statistics* edited by William H. Kruskal and Judith Tanur, New York: Free Press, 399–409.

di Corcia, Joseph (1978). *Bourg, bourgeois, bourgeois de Paris* from the eleventh to the eighteenth century. *Journal of Modern History*, **50**, 207–233.

Febvre, Lucien (1973). *Civilisation:* evolution of a word and group of ideas, in his *A New Kind of History and Other Essays*, edited by Peter Burke. New York: Harper, 219–257.

Goodman, Leo A. and Kruskal, William H. (1959). Measures of association for cross classifications. II: Further discussion and references. *Journal of the American Statistical Association*, **54**, 123–163.

Goodman, Leo A. and Kruskal, William H. (1963). Measures of association for cross classifications. III: Approximate sampling theory. *Journal of the American Statistical Association*, **58**, 525–540.

Goodman, Leo A. and Kruskal, William H. (1972). Measures of association for cross classifications. IV: Simplification of asymptotic variances. *Journal of the American Statistical Association*, **67**, 415–421.

Kruskal, W.H. and Mosteller, F. (1979a). Representative sampling, I: Nonscientific literature. *International Statistical Review*, **47**, 13–24.

Kruskal, W.H. and Mosteller, F. (1979b). Representative sampling, II: Scientific literature, excluding statistics. *International Statistical Review*, **47**, 111–127.

Kuipers, L. and Niederreiter, H. (1974). *Uniform Distribution of Sequences*, New York: Wiley.

Lippman, Walter (1955). *Essays on the Public Philosophy*, New York: Little, Brown. Pagination of the quotations is from the 1956 paperback edition, New York: Mentor Books, New American Library.

Mosteller, Frederick (1948). On pooling data. *Journal of the American Statistical Association*, **43**, 231–242.

Rao, C.R. (1949). On the distance between two populations. *Sankhyā*, **9**, 246–248.

Sobel, Milton and Huyett, Marilyn J. (1958). Nonparametric definition of the representativeness of a sample – with tables. *The Bell System Technical Journal*, **37**, 135–161.

Spitzer, Leo (1942–43). *Milieu* and *ambiance*: an essay in historical semantics. *Philosophy and Phenomenological Research*, **3**, 1–42, 169–218, 380.

Williams, Raymond (1976). *Keywords/A Vocabulary of Culture and Society*, New York: Oxford.

Zimmerman, Carle C. and Frampton, Merle E. (1935). *Family and Society*, New York: Van Nostrand. Contains, beginning on p. 361, an abridged translation of Frédéric Le Play's *Les Ouvriers Européens*, **1**, 1879. The abridgement is based on a translation into English by Samuel Dupertuis. See also: Lazarsfeld, Paul F. (1961). Notes on the history of quantification in sociology – trends, sources, and problems. *Isis*, **52**, 277–333. Reprinted in *Studies in the History of Statistics and Probability*, Vol. II (1977), edited by Maurice Kendall and R.L. Plackett. London: Griffin. Also reprinted in Woolf, Harry (1961). *Quantification: A History of the Meaning of Measurement in the Natural and Social Sciences*. Indianapolis: Bobbs-Merrill.

Résumé

Les différents sens dans la littérature statistique, de: 'échantillon représentatif' et 'sondage représentatif', ont été classés, illustrés et discutés. Les catégories retenues sont celles de nos précédents articles, ainsi que trois nouvelles: méthodes spécifiques de sondage, méthodes permettant de bonnes estimations, méthodes suffisamment bonnes pour un objectif particulier. L'article s'achève sur un aperçu des orientations techniques ou mathématiques correspondantes.

This content downloaded from 128.62.51.103 on Thu, 24 Aug 2017 18:20:40 UTC
All use subject to http://about.jstor.org/terms