

Acknowledgments

THE PRETTY little instances of bumbling and chicanery with which this book is peppered have been gathered widely and not without assistance. Following an appeal of mine through the American Statistical Association, a number of professional statisticians—who, believe me, deplore the misuse of statistics as heartily as anyone alive—sent me items from their own collections. These people, I guess, will be just as glad to remain nameless here. I found valuable specimens in a number of books too, primarily these: *Business Statistics*, by Martin A. Brumbaugh and Lester S. Kellogg; *Gauging Public Opinion*, by Hadley Cantril; *Graphic Presentation*, by Willard Cope Brinton; *Practical Business Statistics*, by Frederick E. Croxton and Dudley J. Cowden; *Basic Statistics*, by George Simpson and Fritz Kafka; and *Elementary Statistical Methods*, by Helen M. Walker.

- Common word
- Uncommon word
- Phrases



Introduction

岳父

"THERE'S a mighty lot of crime around here," said my father-in-law a little while after he moved from Iowa to California. And so there was—in the newspaper he read. It is one that overlooks no crime in its own area and has been known to give more attention to an Iowa murder than was given by the principal daily in the region in which it took place.

My father-in-law's conclusion was statistical in an in-

formal way. It was based on a sample, a remarkably biased one. Like many a more sophisticated statistic it was guilty of semiattachment: It assumed that newspaper space given to crime reporting is a measure of crime rate.

A few winters ago a dozen investigators independently reported figures on antihistamine pills. Each showed that a considerable percentage of colds cleared up after treatment. A great fuss ensued, at least in the advertisements, and a medical-product boom was on. It was based on an eternally springing hope and also on a curious refusal to look past the statistics to a fact that has been known for a long time. As Henry G. Felsen, a humorist and no medical authority, pointed out quite a while ago, proper treatment will cure a cold in seven days, but left to itself a cold will hang on for a week.

So it is with much that you read and hear. Averages and relationships and trends and graphs are not always what they seem. There may be more in them than meets the eye, and there may be a good deal less. 吸引人的

The secret language of statistics, so appealing in a fact-minded culture, is employed to sensationalize, inflate, confuse, and oversimplify. Statistical methods and statistical terms are necessary in reporting the mass data of social and economic trends, business conditions, "opinion" polls, the census. But without writers who use the words with honesty and understanding and readers who know what they mean, the result can only be semantic nonsense.

In popular writing on scientific matters the abused statistic is almost crowding out the picture of the white-jacketed

排擠

大驚小怪
隨之而來

注重事實的文化
聳人聽聞 膨脹

人口普查
語義廢話

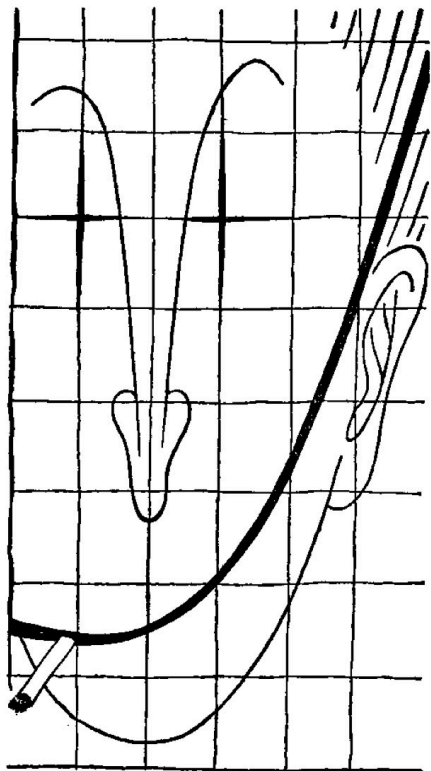
hero laboring overtime without time-and-a-half in an ill-lit laboratory. Like the "little dash of powder, little pot of paint," statistics are making many an important fact "look like what she ain't." A well-wrapped statistic is better than Hitler's "big lie"; it misleads, yet it cannot be pinned on you.

懷舊

This book is a sort of primer in ways to use statistics to deceive. It may seem altogether too much like a manual for swindlers. Perhaps I can justify it in the manner of the retired burglar whose published reminiscences amounted to a graduate course in how to pick a lock and muffle a footfall: The crooks already know these tricks; honest men must learn them in self-defense.

自我
防備





CHAPTER 1

The Sample with
the Built-in Bias

"THE AVERAGE Yaleman, Class of '24," *Time* magazine noted once, commenting on something in the *New York Sun*, "makes \$25,111 a year."

Well, good for him!

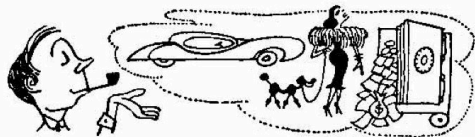
But wait a minute. What does this impressive figure mean? Is it, as it appears to be, evidence that if you send your boy to Yale you won't have to work in your old age and neither will he?

Two things about the figure stand out at first suspicious glance. It is surprisingly precise. It is quite improbably salubrious.

There is small likelihood that the average income of any far-flung group is ever going to be known down to the dollar. It is not particularly probable that you know your

own income for last year so precisely as that unless it was all derived from salary. And \$25,000 incomes are not often all salary; people in that bracket are likely to have well-scattered investments.

Furthermore, this lovely average is undoubtedly calculated from the amounts the Yale men *said* they earned. Even if they had the honor system in New Haven in '24, we cannot be sure that it works so well after a quarter of a century that all these reports are honest ones. Some people when asked their incomes exaggerate out of vanity



虛榮

or optimism. Others minimize, especially, it is to be feared, on income-tax returns; and having done this may hesitate to contradict themselves on any other paper. Who knows what the **revenueurs** may see? It is possible that these two tendencies, to boast and to understate, cancel each other out, but it is unlikely. One tendency may be far stronger than the other, and we do not know which one.

收入者

We have begun then to account for a figure that common sense tells us can hardly represent the truth. Now let us put our finger on the likely source of the biggest error, a source that can produce \$25,111 as the "average income" of some men whose actual average may well be nearer half that amount.

常識

This is the sampling procedure, which is the heart of the greater part of the statistics you meet on all sorts of subjects. Its basis is simple enough, although its **refinements** in practice have led into all sorts of by-ways, some less than respectable. If you have a barrel of beans, some red and some white, there is only one way to find out exactly how many of each color you have: Count 'em. However, you can find out approximately how many are red in much easier fashion by pulling out a handful of beans and counting just those, **figuring that the proportion will be the same all through the barrel**. If your sample is large enough and selected properly, it will represent the whole well enough for most purposes. If it is not, it may be far less accurate than an intelligent guess and have nothing to recommend it but a **spurious** air of scientific precision. It is sad truth that conclusions from such samples, biased or too small or both, lie behind much of what we read or think we know.

虛假的

The report on the Yale men comes from a sample. We can be pretty sure of that because reason tells us that no one can get hold of all the living members of that class of '24. There are bound to be many whose addresses are unknown twenty-five years later.



And, of those whose addresses are known, many will not reply to a questionnaire, particularly a rather personal one. With some kinds of mail questionnaire, a five or ten per cent response is quite high. This one should have done better than that, but nothing like one hundred per cent.

So we find that the income figure is based on a sample composed of all class members whose addresses are known and who replied to the questionnaire. Is this a representative sample? That is, can this group be assumed to be equal in income to the unrepresented group, those who cannot be reached or who do not reply?



Who are the little lost sheep down in the Yale rolls as "address unknown"? Are they the big-income earners—the Wall Street men, the corporation directors, the manufacturing and utility executives? No; the addresses of the rich will not be hard to come by. Many of the most prosperous members of the class can be found through *Who's Who in America* and other reference volumes even if they have neglected to keep in touch with the alumni office. It is a good guess that the lost names are those of

文科學士學位

the men who, twenty-five years or so after becoming Yale bachelors of arts, have not fulfilled any shining promise. They are clerks, mechanics, tramps, unemployed alcoholics, barely surviving writers and artists . . . people of whom it would take half a dozen or more to add up to an income of \$25,111. These men do not so often register at class reunions, if only because they cannot afford the trip.

We are poor little lambs
who have lost our way



被卡住

Who are those who **chucked** the questionnaire into the nearest wastebasket? We cannot be so sure about these, but it is at least a fair guess that many of them are just not making enough money to brag about. They are a little like the fellow who found a note clipped to his first pay check suggesting that he consider the amount of his salary confidential and not material for the interchange of office confidences. "Don't worry," he told the boss. "I'm just as ashamed of it as you are."

It becomes pretty clear that the sample has omitted two groups most likely to depress the average. The \$25,111 figure is beginning to explain itself. If it is a true figure for anything it is one merely for that special group of the class of '24 whose addresses are known and who are willing to stand up and tell how much they earn. Even that requires an assumption that the gentlemen are telling the truth.

Such an assumption is not to be made lightly. Experience from one breed of sampling study, that called market research, suggests that it can hardly ever be made at all. A house-to-house survey purporting to study magazine readership was once made in which a key question was: What magazines does your household read? When the results were tabulated and analyzed it appeared that a great many people loved *Harper's* and not very many read *True Story*. Now there were publishers' figures around at the time that showed very clearly that *True Story* had more millions of circulation than *Harper's* had hundreds of thousands. Perhaps we asked the wrong kind of people, the designers of the survey said to themselves. But no, the questions had been asked in all sorts of neighborhoods all around the country. The only reasonable conclusion then was that a good many of the respondents, as people are called when they answer such questions, had not told the truth. About all the survey had uncovered was snobbery.

In the end it was found that if you wanted to know what certain people read it was no use asking them. You

列表化的

could learn a good deal more by going to their houses and saying you wanted to buy old magazines and what could be had? Then all you had to do was count the *Yale Reviews* and the *Love Romances*. Even that dubious device, of course, does not tell you what people read, only what they have been exposed to. 暴露

Similarly, the next time you learn from your reading that the average American (you hear a good deal about him these days, most of it faintly improbable) brushes his teeth 1.02 times a day—a figure I have just made up, but it may be as good as anyone else's—ask yourself a question. How can anyone have found out such a thing? Is a woman who has read in countless advertisements that non-brushers are social offenders going to confess to a stranger that she does not brush her teeth regularly? The statistic



may have meaning to one who wants to know only what people say about tooth-brushing but it does not tell a great deal about the frequency with which bristle is applied to incisor.

A river cannot, we are told, rise above its source. Well, it can seem to if there is a pumping station concealed somewhere about. It is equally true that the result of a sampling study is no better than the sample it is based on. By the time the data have been filtered through layers of statistical manipulation and reduced to a decimal-pointed average, the result begins to take on an aura of conviction that a closer look at the sampling would deny.

Does early discovery of cancer save lives? Probably. But of the figures commonly used to prove it the best that can be said is that they don't. These, the records of the Connecticut Tumor Registry, go back to 1935 and appear to show a substantial increase in the five-year survival rate from that year till 1941. Actually those records were begun in 1941, and everything earlier was obtained by tracing back. Many patients had left Connecticut, and whether they had lived or died could not be learned. According to the medical reporter Leonard Engel, the built-in bias thus created is "enough to account for nearly the whole of the claimed improvement in survival rate."

To be worth much, a report based on sampling must use a representative sample, which is one from which every source of bias has been removed. That is where our Yale figure shows its worthlessness. It is also where a great many of the things you can read in newspapers and maga-

zines reveal their inherent lack of meaning.

A psychiatrist reported once that practically everybody is neurotic. Aside from the fact that such use destroys any meaning in the word "neurotic," take a look at the man's sample. That is, whom has the psychiatrist been observing? It turns out that he has reached this edifying conclusion from studying his patients, who are a long, long way from being a sample of the population. If a man were normal, our psychiatrist would never meet him.

操縱

精神科醫生



Give that kind of second look to the things you read, and you can avoid learning a whole lot of things that are not so.

It is worth keeping in mind also that the dependability of a sample can be destroyed just as easily by invisible sources of bias as by these visible ones. That is, even if you can't find a source of demonstrable bias, allow yourself some degree of skepticism about the results as long as there is a possibility of bias somewhere. There always is.

明顯的偏見

總統選舉

The presidential elections in 1948 and 1952 were enough to prove that, if there were any doubt.

For further evidence go back to 1936 and the *Literary Digest's* famed fiasco. The ten million telephone and *Digest* subscribers who assured the editors of the doomed



magazine that it would be Landon 370, Roosevelt 161 came from the list that had accurately predicted the 1932 election. How could there be bias in a list already so tested? There was a bias, of course, as college theses and other post mortems found: People who could afford telephones and magazine subscriptions in 1936 were not a cross section of voters. Economically they were a special kind of people, a sample biased because it was loaded with what turned out to be Republican voters. The sample elected Landon, but the voters thought otherwise.

The basic sample is the kind called "random." It is selected by pure chance from the "universe," a word by which the statistician means the whole of which the

sample is a part. Every tenth name is pulled from a file of index cards. Fifty slips of paper are taken from a hatful. Every twentieth person met on Market Street is interviewed. (But remember that this last is not a sample of the population of the world, or of the United States, or of San Francisco, but only of the people on Market Street at the time. One interviewer for an opinion poll said that she got her people in a railroad station because "all kinds of people can be found in a station." It had to be pointed out to her that mothers of small children, for instance, might be underrepresented there.)

The test of the random sample is this: Does every name or thing in the whole group have an equal chance to be in the sample?

The purely random sample is the only kind that can be examined with entire confidence by means of statistical theory, but there is one thing wrong with it. It is so difficult and expensive to obtain for many uses that sheer cost eliminates it. A more economical substitute, which is almost universally used in such fields as opinion polling and market research, is called stratified random sampling.

To get this stratified sample you divide your universe into several groups in proportion to their known prevalence. And right there your trouble can begin: Your information about their proportion may not be correct. You instruct your interviewers to see to it that they talk to so many Negroes and such-and-such a percentage of people in each of several income brackets, to a specified number of farmers, and so on. All the while the group must be

消除

divided equally between persons over forty and under forty years of age.

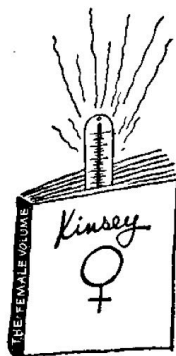
That sounds fine—but what happens? On the question of Negro or white the interviewer will judge correctly most of the time. On income he will make more mistakes. As to farmers—how do you classify a man who farms part time and works in the city too? Even the question of age can pose some problems which are most easily settled by choosing only respondents who obviously are well under or well over forty. In that case the sample will be biased by the virtual absence of the late-thirties and early-forties age groups. You can't win.

On top of all this, how do you get a random sample within the stratification? The obvious thing is to start with a list of everybody and go after names chosen from it at random: but that is too expensive. So you go into the streets—and bias your sample against stay-at-homes. You go from door to door by day—and miss most of the employed people. You switch to evening interviews—and neglect the movie-goers and night-clubbers.

The operation of a poll comes down in the end to a running battle against sources of bias, and this battle is conducted all the time by all the reputable polling organizations. What the reader of the reports must remember is that the battle is never won. No conclusion that "sixty-seven per cent of the American people are against" something or other should be read without the lingering question, Sixty-seven per cent of which American people?

So with Dr. Alfred C. Kinsey's "female volume." The

分層



problem, as with anything based on sampling, is how to read it (or a popular summary of it) without learning too much that is not necessarily so. There are at least three levels of sampling involved. Dr. Kinsey's samples of the population (one level) are far from random ones and may not be particularly representative, but they are enormous samples by comparison with anything done in his field before and his figures must be accepted as revealing and important if not necessarily on the nose. It is possibly more important to remember that any questionnaire is only a sample (another level) of the possible questions and that the answer the lady gives is no more than a sample (third level) of her attitudes and experiences on each question.

The kind of people who make up the interviewing staff can shade the result in an interesting fashion. Some years ago, during the war, the National Opinion Research Center sent out two staffs of interviewers to ask three questions of five hundred Negroes in a Southern city. White interviewers made up one staff, Negro the other.

One question was, "Would Negroes be treated better or worse here if the Japanese conquered the U.S.A.?" Negro interviewers reported that nine per cent of those they asked said "better." White interviewers found only two per cent of such responses. And while Negro interviewers found only twenty-five per cent who thought Negroes would be treated worse, white interviewers turned up forty-five per cent.

When "Nazis" was substituted for "Japanese" in the question, the results were similar.

The third question probed attitudes that might be based on feelings revealed by the first two. "Do you think it is more important to concentrate on beating the Axis, or to make democracy work better here at home?" "Beat Axis" was the reply of thirty-nine per cent, according to the Negro interviewers; of sixty-two per cent, according to the white.

Here is bias introduced by unknown factors. It seems likely that the most effective factor was a tendency that must always be allowed for in reading poll results, a desire to give a pleasing answer. Would it be any wonder if, when answering a question with connotations of disloyalty in wartime, a Southern Negro would tell a white man what

黑人

軸心國

內涵

sounded good rather than what he actually believed? It is also possible that the different groups of interviewers chose different kinds of people to talk to.

In any case the results are obviously so biased as to be worthless. You can judge for yourself how many other poll-based conclusions are just as biased, just as worthless—but with no check available to show them up.



You have pretty fair evidence to go on if you suspect that polls in general are biased in one specific direction, the direction of the *Literary Digest* error. This bias is toward the person with more money, more education, more information and alertness, better appearance, more conventional behavior, and more settled habits than the average of the population he is chosen to represent. 慣例行為

You can easily see what produces this. Let us say that you are an interviewer assigned to a street corner, with one interview to get. You spot two men who seem to fit the category you must complete: over forty, Negro, urban. One is in clean overalls, decently patched, neat. The other is dirty and he looks surly. With a job to get done, you approach the more likely-looking fellow, and your colleagues all over the country are making similar decisions.

Some of the strongest feeling against public-opinion polls is found in liberal or left-wing circles, where it is rather commonly believed that polls are generally rigged. 被操縱的 Behind this view is the fact that poll results so often fail to square with the opinions and desires of those whose thinking is not in the conservative direction. Polls, they point out, seem to elect Republicans even when voters shortly thereafter do otherwise.

Actually, as we have seen, it is not necessary that a poll be rigged—that is, that the results be deliberately twisted in order to create a false impression. The tendency of the sample to be biased in this consistent direction can rig it automatically.