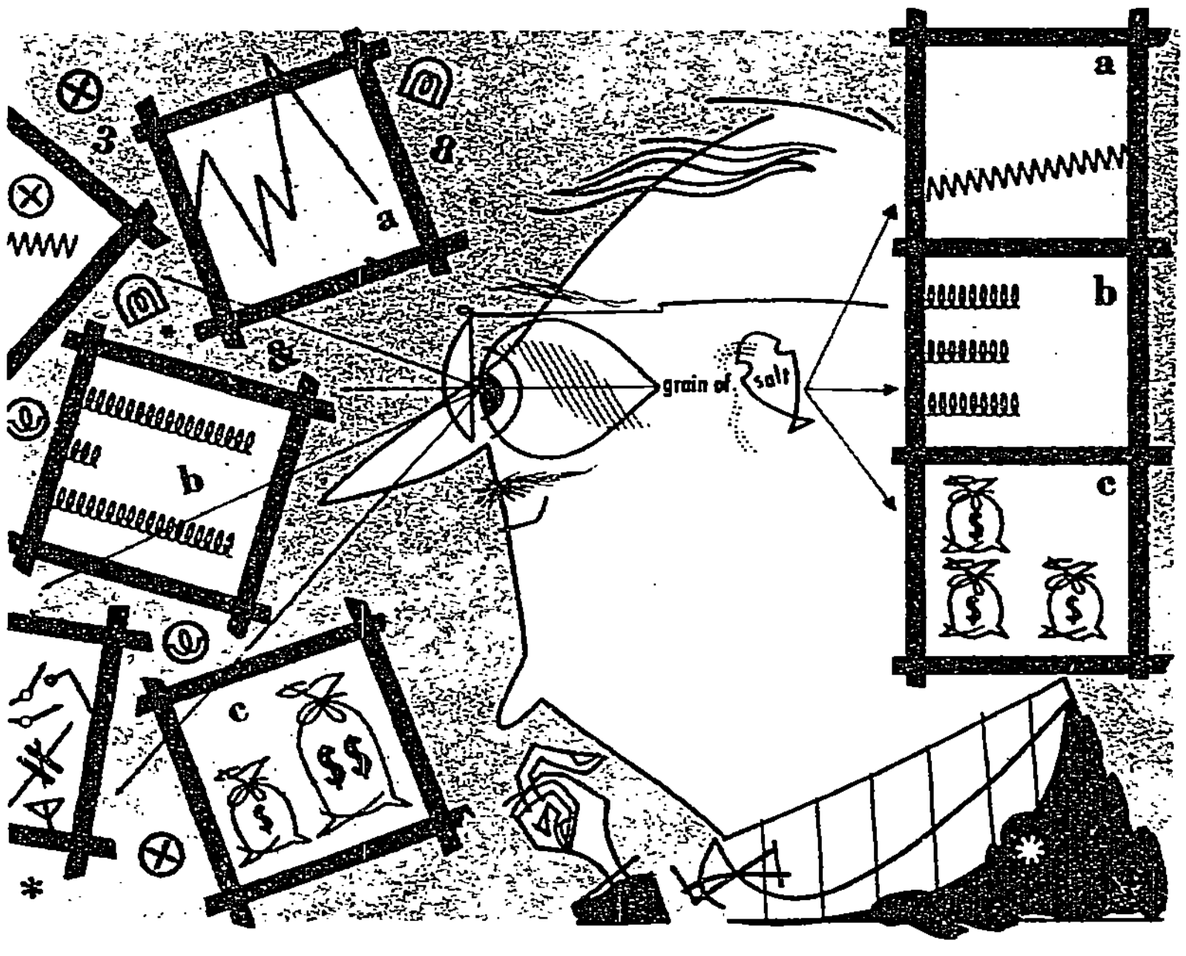
How to Spot Statistical Jokers

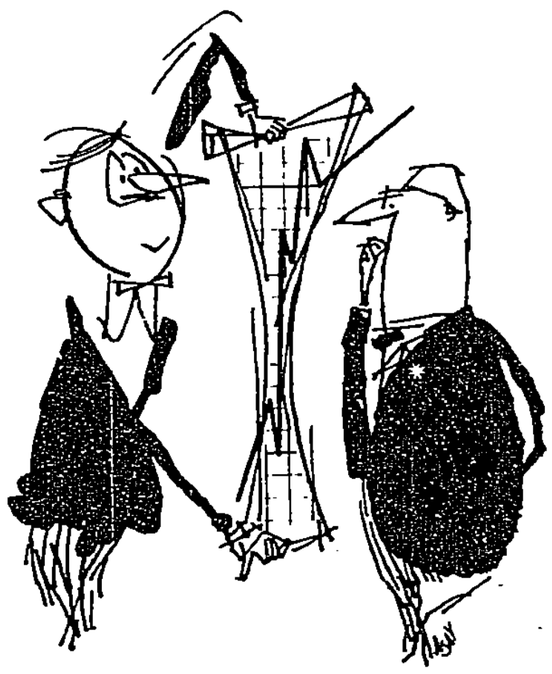
Averages and percentages may look precise, but beware: they ain't necessarily so.

By DARRELL HUFF



Interpreting the flood of statistics—It helps to have 20-20 insight.

Drawings by Ajay



To give a chart a more spectacular appearance, it need only be pulled upward.

ARE wages keeping up with profits and the cost of living? Are you inviting lung cancer when you light a cigarette? Are we in the early stages of a depression without knowing it? Will toothache abandon you if there's fluorine in your drinking water or the latest miracle ingredient in your toothpaste?

Answers to such real and important questions as these lie hidden in the stream of statistics that assails all of us every day. There's no doubt that we have reached the age that H. G. Wells long ago foresaw: "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write."

But what do the figures mean--all those little averages, percentages, percentiles, trends, norms, correlations, significances? What do they say, perhaps? No, it can't be as simple as that. One interpreter will hand down one meaning today and another will come along tomorrow and prove something quite different from the same data.

In the end the consumer of statistics --which is to say each of us--must reach his own decisions as wary as though buying a used car. We can't all be statisticians (or mechanics) but there are many statistical tricks and pitfalls that anyone can readily learn to spot. They are as revealing of phoniness as a worn-out brake pedal under an odometer brazenly reading two thousand miles. For instance—

THE UNSPECIFIED AVERAGE

READ that the average income of employes in the place you work is $4,509 a year and you seem to have been told something precise and meaningful. There's the figure right down to the dollar—and an average is an average, isn't it?

That it's not. The word average covers mean, median, and sometimes other things. For some data they're about the same, and for some, including incomes, they are miles apart. The *mean* income of ninety people is their total take divided by ninety. But if the same ninety have a *median* income of $4,509, this indicates that half of them received less than that and half more. A few very high incomes will boost a mean average a good deal and make conditions look much better than they are. They won't affect a median so much, so that is the kind of average most revealing when incomes are in question.

When you see a median income given, as with Bureau of Census figures, you are getting a reasonably good picture of conditions. When you see a mean, you can guess that it has made incomes look a good deal better than they are. And when you see an "average" of unspecified nature, you can suspect that there's a cheery liar of a mean hiding behind it.

THE BIASED SAMPLE

THIS is another highly prevalent statistical joker. Statisticians, who are as able and honest as any other group, spend most of their lives fighting it. It comes in two popular varieties: the sample that is biased from the beginning, and the one that develops its bias as it goes along.

The famous Literary Digest poll remains the archetype of the first. The enormous sample of ten million voters who agreed it would be Landon two to one came from lists of Digest and telephone subscribers. Such a sample is likely to be biased in at least one direction: wealthier than most. Given a year like 1936 when economic status seems to have had a good deal to do with voting, such a sample is badly biased.

The self-biasing sample is even trickier. For an extreme example, imagine what would happen if you sent, to a purely random sample of your fellow-citizens, a questionnaire that included this query: "Do you like to answer questionnaires?" Add up the returns and you'll probably be able to announce that an overwhelming majority of "a typical cross-section of the population" asserts affection for the things. What has happened, of course, is that those whose answer would have been "No" have eliminated themselves from the survey by flinging it into the nearest waste basket.

It was a self-biasing sample rather like this hypothetical one that led to a news story saying, "The average Yale-man, Class of '24, makes $25,111 a year." Aside from being ridiculously precise, this average is boosted by being just what you think it is—a mean rather than a median. On top of that, it has built into it the self-biasing quality common to many mail surveys. That is, it has taken into account only those members of the class whose addresses are still known and who are sufficiently proud of their incomes to reveal them to their classmates. Those poor little lambs who have lost their economic way are not added in.

THE IMPROBABLY PRECISE FIGURE

THIS is another giveaway worth watching for. Anything based on a survey or other form of sampling procedure has a known statistical error in it, along with other and unknown possibilities of error. Spurious precision is a way of covering this up. If the Yale average had any meaning at all, it would more properly have been expressed as "about $25,000," since at best it could not have been considered accurate down to that last $111. But it's a good deal more impressive the other way, isn't it?

Likewise such a popular item as the I. Q. A child who is scored 110 is not necessarily brighter, even within the limited meaning of such a test, than the one rated at 108. The average error (called "probable error") of the Stanford-Binet, one of the best of such tests, is given as 3 per cent.



Surveys of college graduates' incomes are likely to conceal non-reporting failures.

THE INADEQUATE SAMPLE

ALL sorts of fascinating chicanery is produced by this means. Flip a coin six times and what will you get? (I just tried it and got heads five times.) You may get a fifty-fifty result, and you may not. But flip it a thousand times and you can depend on getting approximately the normal expectation of half heads. Test a new toothpaste on six people and you may find a "remarkable reduction in dental caries!" even though the stuff is worthless. If you don't get the result you want, test another group. By the time you've done this a few times you're practically sure to get a favorable result.

Watch out, then, for reports relying on a tiny sample unless accompanied by a statistician's statement of significance. Good evidence on some things can be amassed with a handful of cases but more often it takes hundreds to mean anything. A polio vaccine was once tested on 450 children before anyone stopped to consider this factor. The incidence of paralytic polio is so low that meaningful results can be obtained only when the cases are in the tens of thousands, a fact taken into account in the 1954 work.

CORRELATIONS

THESE produce a great part of our statistical reading matter. They become misleading when cause-and-effect are read into them without evidence. College graduates make more money than other people. True, apparently. Therefore, if you send your son to college you'll increase his chances of making money. Not proved. Maybe the same traits (ambition, brains, capital) that make a boy likely to go to college are the ones that would have made him more prosperous than most folks anyway.

Therein lies one of the biggest questions about the research reports on correlation *(Continued on Page 43)*

DARRELL HUFF is a specialist in reading between statistical lines. He is author of the book, "How to Lie With Statistics."

The New York Times
Published: August 22, 1954
Copyright © The New York Times

How to Spot Statistical Jokers

(Continued from Page 13)

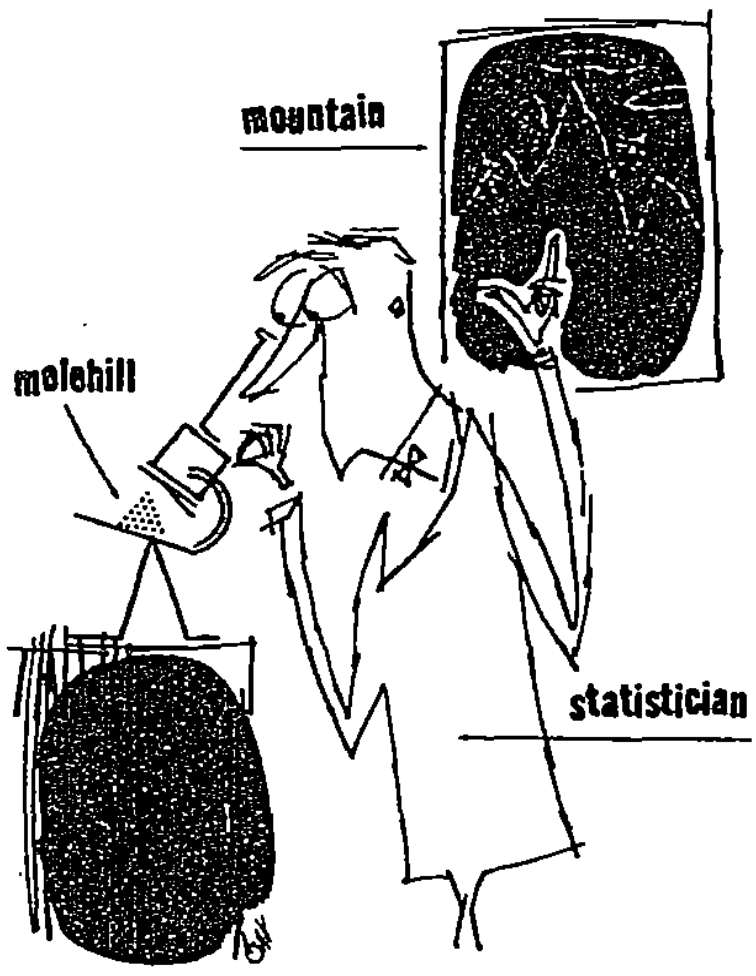
between smoking and death from heart disease or lung cancer. There seems to be good evidence that the two go together—but has the one necessarily caused the other? Still to be eliminated are many possibilities including this one: may not smoking appeal most strongly to persons of a temperament that predisposes of these ailments?

Which is not to say, of course, that the published conclusions are false or that smoking is harmless. But it remains true that although the deaths are real, the cause is still a matter of speculation. This cause-and-effect relationship may turn out to be a true one or it may some day join the lovely correlation once established between cancer and milk drinking

It was shown that cancer was increasing in New England, Minnesota, Wisconsin, and Switzerland, where a lot of milk is produced and consumed, but remaining rare in Ceylon, where milk is scarce. Milk-drinking English women were shown to have some kinds of cancer eighteen times as frequently as Japanese women, who seldom drank milk. It sounded pretty significant but all it really meant was that areas in one group have populations with longer life spans than in the other. Cancer, predominantly a disease of middle life, has little chance to attack where people die young.

THE GEE-WHIZ GRAPHS

WATCHING these go by is a sport to be recommended to anyone interested in looking a phony statistic in the eye. The shape of a graph is supposed to give a quick picture of some statistical fact. Don't bet on it. Chart a trend in graphic fashion and usually you will get a modest-looking rise; things like wages and costs and sales and populations generally change slowly. To give the indicating line a more spectacular appearance you have only to stretch your chart upward, putting more space between the lines.



Inadequate samples can be used to produce all sorts of fascinating chicanery.

The trouble with this is that it produces a very tall and thin chart, a suspicious-looking thing and awkward on a printed page. So chop off the bottom, leaving only the area in which the moving line is found. The result is that any slight rise becomes a hop from the bottom to the top. 1 per cent can look like a hundred or a thousand. Moral: In reading any graph in which the bottom line isn't zero, look not so much at what's there as what's not



I. Q. tests are not necessarily proof that Tot A is more brilliant than Tot B.

Even more deceptive is an unfortunately common version of the pictorial graph in which values are represented by little men or cows or moneybags. If a hundred dollars is portrayed by one moneybag and four hundred by four just like it, then you have a proper and honest graph. But it is all too usual to represent that second quantity by a single moneybag, too, making it four times as tall as the first.

What's wrong with that? Well, a bag four times as high is also four times as wide and so it covers sixteen times as much area on the page. And since it is a picture of a three-dimensional object it actually indicates a bag four times four times four times as big as the first one. An increase of four to one has been made to look like sixty-four to one, as fine a piece of exaggeration as you could ask for.

SEMANTIC TRICKS

BLEND these with statistics and you have the phoniest figures of them all. The formula is: when you can't prove what you want to prove, demonstrate something else and then pretend they're the same thing.

To peddle a nostrum as a cold cure, have an "independent laboratory" testify to how many unspecified germs it will kill in a test tube. To prove young car drivers dangerous, give the figures showing them involved in more accidents than elderly ones; to hang responsibility on older drivers, give figures on accidents per million miles driven. To exaggerate dangers of train travel cite "fatalities chargeable to railroads" but don't mention that more than 90 per cent who died were either riding the rods or in automobiles that met trains at crossings.

If you'll read with an eye alert to statistical mishandling, you'll come upon many splendid specimens to add to these. You won't learn any more by reading with a skeptical eye, but you'll do something that is perhaps many times better. You'll avoid learning a remarkable lot that isn't so.

The New York Times
Published: August 22, 1954
Copyright © The New York Times