Chapter Six

STATISTICS: MAKING SENSE OF THE NUMBERS

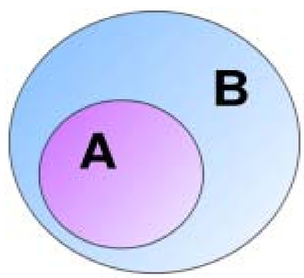
WHAT NUMBERS CAN TELL US

The small, state university I teach at, Eastern Oregon University, like most universities, routinely publishes institutional data. Here's an interesting fact that caught my attention the other day. For the Fall Quarter of 2008, Eastern had 3,666 students. When you breakdown that number on the basis of sex, you discovered something a little surprising. 2,344 of those students were female, while only 1,322 were male.

These statistics tell us a number of things about the university. For one thing, they make good on my claim that Eastern is, indeed, a very small university. But the gender disparity is even more glaring. Why, we might ask, are so many more women attending Eastern than are men? Is this something unique to Eastern, or Oregon, or the Pacific Northwest? A little research will convince you, I think, that Eastern's experience pretty much matches trends nation-wide, and raises empirical questions about why this is happening, and policy questions regarding what to do about it. My point in this chapter is much more modest. I simply want to convince you that statistical data can tell us interesting and important things. And as you may have guessed, statistics often provide good evidence, and that we can understand this evidence through our tool of inference to the best explanation.

SAMPLES AND POPULATIONS

We will use the term *population* as jargon for any sort of a group. A group of people. A group of things like vehicles that get better than 30 mile per gallon. Or groups of very abstract things like depictions of Santa Claus in prime time television. We can use the mathematician's notion of a set to characterize a population. Similarly, we will use the term *sample* as jargon for any part of the group constituting the population. Thus, samples are subsets of the set making up the population. In a familiar Venn diagram, A constitutes the sample, and B the population.



Very often we are interested in samples because we assume that the can tell us something interesting about the population. Although, as we shall see, it was poor statistical procedure, Eastern's gender disparity statistics hinted at something important about the population of contemporary college students. You might well ask why, if we are really interested in the population, why we wouldn't just look at it directly. And the simple answer is one of practicality. It would be too time consuming, expensive, or otherwise impractical to survey the entire population. Thus, we use the sample, which can be examined and described, as a clue about the whole population, which cannot.

Inferences from samples to populations are classic examples of inferences to the best explanation. Our data is the discovery that some sample has an interesting feature or property, and we use this as evidence that the population also has this property. We ask the explanatory question -- why does the sample have *P*? And our hypothesis answers that it has *P* because the population as a whole has *P.*

e_1. Sample has property *P.*
===================
t_0. Population has property *P.*

COULDN'T IT JUST BE A FLUKE?

I hope by now you are almost programmed, when you see an argument like the one directly above, to begin to think of rival explanations. Sure, if the population has *P*, that would be a good explanation of why the sample has *P*. But what else might explain the sample having *P*?

I get home at 6:00 on a Tuesday evening and before I can finish looking at the mail and fixing a martini, the phone has rung three times, all from charitable organizations seeking contributions. I conclude that this Tuesday is a big push for getting money. My sample, those three phone calls, is pretty skimpy. After all, I'm offering a hypothesis about the whole country (state?, county?). Isn't the following rival explanation just as plausible, perhaps more plausible, than my charity full court press theory?

t_1. It's just a coincidence that those three calls were all from charitable organizations.

Or more generally,

t_1. It's just a coincidence that the sample has property *P*.

Modern probability theory has devoted a good deal of time and attention to developing some very sophisticated mathematical tests of how likely it is that a sample will have a given property simply as a matter of random chance. Some of you may be familiar with some to these tests for what is called “statistical significance” from other courses or computer software. Even those of you who hate numbers or math would be well-advised, in my humble opinion, to learn a bit about all of this by taking an introductory statistics course. But that is not my goal in the present context.

Even those of you with the least experience and confidence with mathematics know that the size of the sample matters in important ways. A sample of three calls tells us almost nothing, while a sample of 3,000 can tell us quite a lot. We can confine our discussion to an informal treatment of what statisticians call “statistical significance.” How accurate are our measurements within samples of a given size? A contemporary philosopher of science, Ronald Giere, offers what he calls a “rule of thumb” for answering this question.[1] He offers the following scale for correlating the size of sample with the accuracy of what is being measured.

Sample Size	**Accuracy**
100	±10%
500	±5%
2,000	±2%
10,000	±1%

You might note a couple of things about this little chart. One is how nicely the first digit in the sample size correlates with the accuracy measurement, thus making it pretty darn easy to remember. The other is what economists call the law of diminishing returns. Increasing the sample from 100 to 500 buys you a lot of increased accuracy; increasing it from 2,000 to 10,000 buys you hardly any increased accuracy. You will notice, I predict, that almost all of the polls you read about in the newspapers will have sample sizes around 500. This is because accuracy of about ±5% is all that is needed for most purposes, and it would be very expensive and time consuming to improve that accuracy significantly.

COULDN'T THE SAMPLE BE BIASED?

The notion of bias in colloquial speech often conveys a lack of openness, or even prejudice, which counts as a kind of character defect -- he's really biased in his grading against student athletes. I'm biased toward folk and rock music, because it's what I grew up with. Some of you, God forbid, are biased toward hip hop for the same reason. All the notion really means is that people are not equally open -- to giving good grades, appreciating a song as a good one, or noticing that the dishes need to be washed. We need to make sure that our samples are not biased, but equally open to everyone, or everything, in the population.

Statisticians desire randomly selected samples. This is technical jargon that means every single individual in the population has an equal probability of being selected as a member of the sample. My computer can approximate random selection, so it would be relatively easy for me to feed in all of my class rosters for the past five years, randomly select three students from each course, and then query this sample to discover things about my teaching, grading, etc. Not a bad idea, actually.

In the real world, however, technical randomness is often impossible. We only have a couple of days to find out voter sentiment in the upcoming election, and so we phone a sample of 600 likely voters. Obviously this is not a true random sample since every likely voter did not have an equal chance of being selected -- some didn't have phones, some were away on vacation, and some screen their calls. But for practical purposes, if the phone numbers are randomly selected from a master list of likely voters who answer their phones, the information we gather approximates what could be gathered from a technically random sample, and our sample might be characterized as ***practically random.*** Technically random samples are the exception, while what we hope are practically random samples are the rule.

Consider a very famous poll that went spectacularly wrong. The *Literary Digest* had been conducting polls on presidential elections since 1920, and had gotten the winner right in four straight elections, indeed in the 1932 election they got the popular vote within one percent. As the 1936 election approached they once again conducted a massive poll. Take a look at the relevant data.

e_1. The *Literary Digest* mailed out more than 10 million straw vote ballots.
e_2. Their "sample was drawn primarily from automobile registration lists and phone books."[2]
e_3. "Over 2.3 million ballots were returned."[3]
e_4. 55% "straw" voted for Alf Landon, 41% for Roosevelt, and 4% for Lemke.

This led to their conclusion that voters overwhelmingly favored Landon, and their cover story prediction that he would win the election. They made a classic inference from a sample to a population.

e_1. *Literary Digest* sample strongly favors Landon.
============================
t_0. Voters, nationally, strongly favor Landon.

Bad luck for the *Literary Digest*! You, of course, know that Alf Landon never became President. I'll bet a good number of you have never even heard of him before. Roosevelt crushed Landon in the general election 61% to 37%. What went wrong?

The *Digest*'s sample was horribly biased. Not because they were prejudiced or had some axe to grind, but because the way they selected the names and addresses was far from random. Not the technical randomness that we almost never find, but the practical randomness that good polling requires. The clue is in e_2. This was, after all, the height of the great depression. Poor people were much less likely to own a car. And even phones were not then considered necessities, but in a sense, luxuries. Again, poor people were much less likely to have phones. What the *Literary Digest* had unintentionally done is measure the sentiments of relatively wealthy voters, not voters in general.

t_1. Wealthy voters strongly favor[ed] Landon.

It is well-known in political science that wealthier voters tend to vote for Republicans, and less wealthy voters for Democrats. It's hardly surprising, therefore, that a sample of voters biased toward the Republican Party tended to favor the Republican candidate.

There was a second source of bias in the sample that is less well discussed in academic circles. The whole poll depended on what statisticians call the "response rate." The *Literary Digest* sent out a truly amazing number of straw ballots -- more than 10 million. They got a pretty good response, too -- almost a quarter. But we should ask ourselves if there was anything special about those 2.3 million who took the trouble mail their ballots back. It seems reasonable to suppose that they were more educated and politically concerned. So we have a second rival explanation.

> t_2. Better educated and politically concerned voters favored Landon.

And, indeed t_1 and t_2 nicely compliment one another, and suggest a more comprehensive rival.

> t_3. Wealthy voters, as well as better educated and politically concerned voters, favored Landon.

Lest any of you think that all of this concern with polling for presidential elections is a thing of the past, you might well reflect on the historic election that just took place. Here's what professional pollsters were worried about as the 2008 election approached.

> The number of people who have dropped their landlines in favor of cell phones isn't really all that large. The problem is, that number is growing like crazy.
>
> "We were all scared to death in 2004, because we had a close race and the cell phone-only problem was already with us then," says Scott Keeter, the head of surveys at the Pew Research Center.
>
> Exit polls for the 2004 presidential election found some 7 percent of voters said they used their cell phones exclusively. Keeter says that number is likely to grow to nearly 15 percent by the 2008 election.
>
> Pollsters have learned quite a bit about the cell phone-only users they do call. They are most likely to be under 30, unmarried, renters, making less than $30,000 a year, and are slightly more likely to be black or Hispanic, says Keeter"

> The fact that they are very different is the potential problem for polling," Keeter says.
>
> He adds, "It suggests that if there are enough of them, and you are missing them in your landline surveys, then your polls will have a bias because of that."[4]

NAOMI ORESKES' STUDY

There is an interesting segment in Al Gore's movie, *An Inconvenient Truth*, where he cites a scholarly study of peer-reviewed articles on climate change.

> A University of California at San Diego scientist, Dr. Naomi Oreskes, published in *Science* magazine a massive study of every peer-reviewed science journal article on global warming from the previous 10 years. She and her team selected a large random sample of 928 articles representing almost 10% of that total, and carefully analyzed how many of the articles agreed or disagreed with the prevailing consensus view. About a quarter of the articles in the sample dealt with aspects of global warming that did not involve any discussion of the central elements of the consensus. Of the three-quarters that did address these main points, the percentage that disagreed with the consensus? Zero.[5]

Here we have, a little bit second hand, an incredibly interesting, and potentially quite important, sample. The argument leaves the conclusion unstated, but still quite obvious -- almost all natural scientists publishing on climate change endorse the consensus view about climate change.

e_1. In a sample of 928 peer-reviewed articles dealing with climate change, zero percent disagreed with the consensus view.

==============================

t_0. Virtually all peer-reviewed research on climate change endorses the consensus view.

Mr. Gore is quite right that Dr. Oreske published a short, but very influential, article, “Beyond the Ivory Tower: The Scientific Consensus on Climate Change,” in a prestigious journal, *Science*, in December of 2004.[6] She begins by reminding her readers that policy makers and the mass media often suggest that great scientific uncertainty about “anthropogenic” climate change, but states flatly, “[t]his is not the case.”[7]

In defense of her thesis, she offers a fairly elaborate study she has conducted. She offers a working definition of what she will call “the consensus view,” from reports by the Intergovernmental Panel on Climate Change.

> Human activities ... are modifying the concentration of atmospheric constituents ... that absorb or scatter radiant energy. ... [M]ost of the observed warming over the last 50 years is likely to have been due to the increase in greenhouse gas concentrations.[8]

Notice the challenge she faces. She is making a claim about a very large, and not that well-defined, population -- science (“great scientific uncertainty”). To make matters worse, policy makers and the media dispute her claim.

Her first move is to more carefully define the population she is interested in. She utilizes a standard reference tool in the natural sciences, the Institute for Scientific Information data base. In this data base authors are asked to identify certain “key words,” really topics, that their articles address.

Professor Oreske searched for the key word "climate change." Her team then considered 928 articles. The original article says nothing about how they were selected, and indeed conveys the impression that there were a total of 928 that she found. Mr. Gore, however, says that they were randomly selected, and constitute 10% of the articles.

Obviously not every article is going to explicitly endorse or disagree with the consensus view, so Oreske and her team had to read and "code" the articles. They broke them down into six categories.

> The 928 papers were divided into six categories: explicit endorsement of the consensus position, evaluation of impacts, mitigation proposals, methods, paleoclimate analysis, and rejection of the consensus position. Of all the papers, 75% fell into the first three categories, either explicitly or implicitly accepting the consensus view; 25% dealt with methods or paleoclimate, taking no position on current anthropogenic climate change. Remarkably, none of the papers disagreed with the consensus position.[9]

She is also quite candid that a certain amount of judgment and editing of the sample was required.

> Some abstracts were deleted from our analysis because, although the authors had put "climate change" in their key words, the paper was not about climate change.[10]

So, what do we, none of us trained climate scientists, think of Professor Oreske's evidence? We possess the tools to make some sort of evaluation.

We have a fair amount of data that is being offered as evidence.

e_1. Definition of the "consensus view."
e_2. ISI data base
e_3. Key word: climate change
e_4. 928 articles

e_5. Some articles did not really address climate change, and were removed.
e_6. Six potential categories.
e_7. 75% "implicitly or explicitly" endorsed the consensus view.
e_8. 25% took no stand.
e_9. Not one article disagreed with the consensus view.

============================

t_0. Almost all scientists working and publishing on climate change endorse the consensus view.

RIVAL EXPLANATIONS OF THE SAMPLE

Perhaps it was just a fluke that all 928 articles either endorsed the consensus view, or took no position on it. Certainly this sort of mathematical coincidence is possible.

t_1. It was a fluke that the 928 articles showed no skepticism about the consensus view; the ISI data base contained many articles that were.

Such a rival is logically possible. But I want to insist, however, that it is very improbable. Remember Giere's "rule of thumb?" He tells us that for random samples, the margin of error is a direct function of the size of the sample. Samples of 500 are accurate to about ±5% and samples of 2,000 are accurate to about ±2%. That means that Professor Oreskes' sample has an accuracy of, conservatively, ±4%. For a statistician adopting a 95% confidence level, there is only a 5% chance that the population falls outside of the ±4% margin of error. Could it happen, yes. Is it likely at all, no.[11]

Much more interesting rivals will have to do with the problem of bias, either intentional, or more likely, unintentional. I suspect that some of you have already wondered if there might be a bias in the ISI data base. Maybe they only list "green" articles. Again, the following rival explanation is possible.

> t_2. The ISI data base is biased in favor of the consensus view.

A very different sort of bias is possible because of Oreskes' methodology. It is highly unlikely that most of the articles in sample came right out and said where they stood on the consensus view. Indeed, she tells us that some of the endorsement was implicit. That must mean that her team had to "code" or otherwise interpret that article's intention and subsequent endorsement or non-endorsement. Perhaps her team was so unconsciously wedded to the consensus view that they misinterpreted many of the articles as endorsing, or taking no stand, when in fact the authors of those articles intended a rejection of the consensus view. Thus, another possible rival explanation focuses on the coding of the articles.

> t_3. Oreskey, because of her biases, misinterpreted many of the articles as favorable, or neutral, when in fact the authors were arguing against the consensus view.

A final rival explanation centers on the possible bias of the entire scientific community. One might argue, as some have in defense of "creation science," that there is a kind of conspiracy that effectively censors articles that challenge the consensus view (not just of climate change, but any accepted scientific theory) from being published in peer-reviewed journals in the first place. Here the rival does not really challenge the population of peer-reviewed publications, but rather the implied attitude of endorsement by working scientists.

> t_4. Respectable scientists arguing against the consensus view cannot get their articles published in peer reviewed journals.

THE BEST EXPLANATION?

In the case of the rival focusing on statistical fluke, I could argue against its plausibility by focusing on its mathematical improbability. No such technique exists for dealing with the rivals t_2, t_3, and t_4. Nevertheless, I want to argue that they are all implausible, at least when compared to the original explanation that there exists practically universal endorsement of the consensus view among trained climate scientists.

Consider first the journal that Oreskes' article appeared in, *Science*. The journal is one of the most highly respected academic journals in the world. They have a huge interest in policing themselves, since their name is on the cover of every article they publish. We saw in an earlier chapter that a journal like *The Supreme Court Economic Review*, takes the responsibility so seriously that it will publicly retract an article, even when this means embarrassing themselves, and a prominent scholar and public figure.

Next, we must face the charge that the Institute for Scientific Information is somehow biased. Again, we are dealing with a very prestigious and widely used reference tool, that is now operated by a for profit corporation. The ISI has a huge stake, both its reputation, but also financial, in being regarded absolutely trustworthy. Thus, they too, can be expected to police themselves.

The same may be argued for Professor Oreskes herself. She is a highly respected scholar, educator, and university administrator. Her own professional reputation is on the line. She would be insane not to carefully ensure the accuracy of an article in a major journal that was guaranteed to be read and debated by a wide audience of scientists, and indeed those outside of the sciences.

Finally we come to perhaps the most serious of the charges in our rivals. Perhaps all of climate science is biased against critics of the consensus view. As I said, above, these sorts of conscious or unconscious conspiracy theories are offered by critics of natural selection. I want to concede that something like that can happen, and the history of science tells us that it has happened on occasion. In a way, the criticism of Semmelweis' theory by skeptics of the entrenched generation had shades of this mechanism. But with all this conceded I have to tell you that this sort of thing is very, very rare. Most natural scientists respect the need for skepticism from their peers. Studies challenging the consensus view, in one sense have a better chance of being published, if for no other reason than that they are saying something new. Furthermore, we live in the age of information. Much more is being published, and many more venues for peer-reviewed academic publishing exist now. Thus, the fact that the ICI data base did not include even one skeptical defense leads me to believe that that there just aren't many skeptics out there, at least not within mainstream climate science.

---

[1] Ronald Giere, *Explaining Science*
[2] Peverill Squire, "Why the 1936 *Literary Digest* Poll Failed" http://marioguerrero.info/154/readings/Squire.pdf
[3] *Ibid.*
[4] NPR http://www.npr.org/templates/story/story.php?storyId=14863373

---

[5] Gore, p. 262.

[6] http://www.sciencemag.org/cgi/content/full/306/5702/1686

[7] *Ibid.*

[8] *Ibid.*

[9] *Ibid.*

[10] *Ibid.*

[11] The rule of thumb really only works when there is close to a 50/50 split in the original population. When it is very lopsided as it appears to be in the ISI data base, the margin of error becomes much, much, smaller.