

Counting on Inaccuracy in the US Census

Los Angeles Times (October 16, 1997)

Any kindergartner knows how to count. So why can't the U.S. Congress and the Census Bureau agree on how to count the number of people in the country?

Statisticians say counting noses isn't the best way to come up with the right number - at least not for a population as large and as varied as that of the United States. Far more accurate, they say, is a statistical method called sampling.

Yet Republicans in congress have blocked the latest effort to test sampling for use in the 2000 census, calling it a "risky statistical scheme". The Constitution calls for an "actual enumeration" of the population every 10 years. "To enumerate means to count one to one," said Rep. J. Dennis Hastert (R - Ill.).

So what's wrong with counting? In short, it doesn't work, say mathematicians.

"It would be nice to count everyone," said Temple University's John Allen Paulos, author of "Innumeracy: Mathematical Illiteracy and Its Consequences." But people who are transient, suspicious of government, or marginal, often get overlooked. "So you can choose to forget about them, or you can estimate them in various ways," he said.

Paulos said the widespread belief that it's possible to count each and every person is "absurd... Even if it were exact, people would die by the time you finished counting. And there's always some error, even in the most countable groups like Iowa farmers. There'll always be a few crazy uncles hiding in the attic."

Sampling, mathematicians say, simply acknowledges the inaccuracies built into the system. Statistical techniques allow them to determine how big the error is and to correct for it.

"People in Washington don't understand that any count has some uncertainty around it," said Janet Norwood, a statistician with the National Urban League.

Sampling is controversial because a lot depends on the census- - everything from funds for local programs to district boundaries, and therefore congressional seats. The conventional wisdom is that Democrats would gain seats if sampling were used- - because the people undercounted tend to be less affluent and members of minority groups, who vote Democratic. "But there's no basis in fact for that," Norwood said.

Even though sampling has not been used in the census, it is ubiquitous in everyday life- - a standard way of counting and checking results. "You go to the doctor for a blood test,"

said statistician John Rolph of USC 's Marshall School of Business. "Does he drain all the blood out of your body? No. He takes a sample."

It's impossible to test every drop of blood or taste every spoonful of soup or audit every line in a budget or crash test every car. If the first and second spoonful of soup are too salty, you can assume- - within a certain range of uncertainty- - that the entire pot of soup is salty. If one section of a company's annual report is fraught with errors, you can assume that the rest requires further investigation.

A Problem With Size

Of course, counting everyone in the United States wouldn't be a problem if the nation were the size of a classroom of students.

"If I want to know how many students are in my class, I can count noses," said statistician and lawyer Mary Gray, a professor at American University. "But the university has trouble knowing how many students it has." Some don't pay their bills, some are part-time, some don't have their papers in order, she said.

The problem gets worse as the numbers get bigger. It has been known for decades that the census makes mistakes, but until the most recent count, the bureau kept getting better at correcting them, Rolph

said. In the 1990 census, 4 million to 5 million people fell between the cracks, according to the American Statistical Assn.

"We count some people more than once, and others not at all," Gray said. "What we know for certain is that the error is not zero. And we've made no effort to account for that."

Many students are counted twice- - once at school and once at home- - even though the census bureau instructs parents not to count children living away from home. "They do it anyway," Gray said.

Far greater numbers of people are not counted at all.

"There is a growing group of people who are less easy to reach than in the past," said Rudolph, who last year chaired a panel on census sampling for the association. That group includes the homeless, some immigrants, people in nontraditional living arrangements, and people suspicious of government.

The problem has gotten so bad, he said, that the Census Bureau will be forced to do things differently.

Under the current system, the Census Bureau compiles an address list for everyone in the country and sends out questionnaires. About two thirds of those questionnaires are returned, according to experts.

"That leaves [the bureau] in a quandary," Norwood said. "They can try to keep sending out people to households that have not responded. It takes time. It's very high cost. And they still can't get everyone."

The cost, in fact, is rapidly getting out of control; it increased enormously per person during the last census, Rolph said. The bulk of that money, he said, goes to "following up the non-responses".

And following up creates further biases. Census takers "follow up basically until they run out of time and money," Rolph said. "When you do that, you get those who are easiest to count."

Random Samples

To help overcome the problem, statisticians have recommended that the Census Bureau use sampling to get an accurate estimate of the missing people. In a simplified scenario, the census would list all the households that didn't respond to the questionnaire, Rolph said, then take a random sample of, say, 10%.

"Then we would invest a lot of time and effort to get every one of those 10%. Then we can say, that 10% is representative of the other 90%."

Of course, sampling also introduces errors. They are

preferable, for two reasons, to the errors created by counting, say mathematicians. First, sampling errors are generally not as biased as counting errors. Sampling errors would apply equally to blacks, Latinos and middle-class whites, said Gray, while counting errors apply mostly to minorities and poor people.

Second, the errors introduced by sampling can be precisely calculated. For example, if 2,000 people out of 200 million were sampled, the error is roughly 2%.

"With a carefully drawn sample, we can calculate what the sampling error is," Rolph said. Moreover, sampling would allow for better supervision of census takers, Rolph said. Now, some census takers indulge in practices such as "curbstoning", he said.

Say a census taker has a quota of 100 people to count on a given day. "How do you know whether they really went out and talked to them or looked up at the house from a curbstone and made up a fictitious set of occupants?" Rolph said.

Sampling concentrates the resources of the census takers on the smaller, more focused groups, where census takers can be supervised more closely.

A good sample needs to be both random and representative of the whole. If the pot of soup is hot on the bottom but still

cold on top, and you take a spoonful from the top, you won't get an accurate idea of the temperature of the soup.

Sampling techniques generally take these problems into account, adjusting samples to make sure that all parts of the pot- - or all segments of the population- - are accurately represented.

The size of the pot doesn't matter much. It's counterintuitive, said Paulos, but if you sample 50 people at random from the entire United States, you will get almost as good a result as if you sampled 50 people out of a large graduating class.

It is hard to find any opposition to sampling among mathematicians.

"That's not to say you couldn't find a conservative statistician who would oppose sampling," Gray said. "But it would be like a scientist who says there's no relationship between smoking and lung cancer. There's absolutely no scientific basis for it."



Just for Laughter...

The Japanese eat very little fat and suffer fewer heart attacks than the British or the Americans.

On the other hand, the French eat a lot of fat and also suffer fewer heart attacks than the British or the Americans.

The Japanese drink very little red wine and suffer fewer heart attacks than the British or the Americans.

The Italians drink excessive amounts of red wine and also suffer fewer heart attacks than the British or the Americans.

Conclusion: Eat and drink whatever you like. It's speaking English that kills you.