Elements of Statistics

Lecture Notes # 7

Today

- different types of bias
- Probability sampling method and simple random sampling

1 Biased Sampling

Definition 1.1 *A sampling method that produces results that systematically differ from the truth about the population is said to be* ***biased****.*

By systematically it means it is not different from truth on occasion. It differs in a kind of systemic way. More examples later on will further reveal what is a biased sampling method.

Say, you are a vegetable wholesaler and a truckload of potatoes has just been brought in. It may be convenient to sample a few buckets of potatoes taken from the top of the load, 'cause it will be pretty hard to get to the potatoes at the bottom. This is called a convenience sampling. Convenience samples are generally biased. If you want to examine the average weight of potato and suppose all the heavier ones are sitting at the bottom.Then convenience sampling might lead to inaccuracy.

Definition 1.2 *A* ***convenience sample*** *is a sample consisting of units of the population that are easily accessible.*

Another example is that magazines are always including surveys in their issues asking its readers to respond on how they think the magazine itself. So the results of such a survey would only reflect the opinions of those volunteer to respond. Think about who are more likely to respond. They must be those who have a strong opinion about the magazine, either quite happy about it, or quite unhappy about it. So this kind of volunteer sampling would be biased.

Definition 1.3 *A* ***volunteer sample*** *is a sample consisting of units of the population that chose to respond.*

Another kind of bias (from http://skepdic.com/selectionbias.html). There is once a famous work done by Alfred C. Kinsey. His study shows that 10% of the population is gay. And this data is widely cited in mass media and all sorts of scientific publication. However, many studies later on indicate that the percentage of adults who describe themselves as Gay is much less than 10%. And some even found the figure to be somewhat between 1% and 2%. And the truth is that he interviewed only white men, and these respondents were disproportionately from lower socioeconomic classes. This is an example of the so-called *selection bias.*

Definition 1.4 ***Selection bias*** *is the systematic tendency on the part of the sampling procedure to exclude or include a certain type of unit.*

Yet another kind. Say a survey was conducted by mail, and anyone who would like to take the survey should fill out the questionnaire and send it back by mail. What are the problems? Many people might choose not to respond. Say, people pretty busy all day long won't bother to respond. Those who don't find the survey of any interest might not respond as well. This is called a *nonresponse bias.*

Definition 1.5 *is the distortion that can arise because a large number of units selected for the sample do not respond or refuse to respond, and these nonresponders have a tendency to be different from the responders.*

For the above example, what do you think might increase the return rate? Send gifts. Self-addressed stamped envelop.

Still another example, think about I am conducting a course evaluation survey by distributing questionnaires. And there is one question like,

> Please comment regarding the Instructor of this course.
> (organized, enthusiastic, attentive to students' needs, etc.)

What would be your response? Please write down in a piece of paper with your name on it and hand it to me. Try to be quick.

Indeed, this question is phrased in such a way that it suggest the desired answer. And this is also a kind of bias, called *response bias.*

Definition 1.6 *is the distortion that can arise because the wording of a question and the behavior of the interviewer can affect the response received.*

In real situation different kinds of bias might be interweaving. Refer to the textbook Page 90 Example 2.6.

2 Simple Random Sampling

The primary reason for convenience sample, volunteer sample, or selection bias, are that people are allowed to make choices, and it is really difficult for people to be impartial. Most people would agree that using a coin toss to select one person out of two would be a *fair* selection method. Why this is thought as fair? Each person has the same the chance to be selected. So this gives the basic idea of **probability sampling method.** If the right sampling methods are used, even a relatively small sample could accurately reflect the responses of a large population.

Definition 2.1 *A sampling method that gives each unit in the population a known non-zero chance of being selected is called a* ***probability sampling method****.*

And, the simplest way of probability sampling method is called a **simple random sample**. A **simple random sample of size n** is a sample n units selected in such a way that every possible sample of the **given size n** has the same chance of being selected as any other sample of the same size.

Coin flipping is a simple random sampling method for population size of 2. For any larger size we can imagine it is like mix N (N is the size of population) indistinguishable balls in the basket, each having a label representing each unit in the population. But the usual practice is to generate random numbers in order to select a sample. We can either get them from a table or generate using a fancy calculator or computer.