Statistics 10 Lecture 16 The Accuracy of Percentages (Chapter 21.1-21.3)

1. Overview

In Chapters 17-20, the parameters were known, many samples are taken, and then we estimate the chance of a specific outcome. In Chapter 21 we begin STATISTICAL INFERENCE, here the parameters are unknown, and we draw conclusions from sample outcomes to make guesses about the value of the parameters. We are given a single sample and then ask the question -- what did the "box" that generated the sample outcome look like?

In this chapter we will examine CONFIDENCE INTERVALS (21.2) for estimating the value of the population parameter. The confidence intervals are based on the sampling distribution of statistics from Chapter 20.1 An important point to remember, population parameters, although unknown, are fixed (they do not change). It was the OUTCOME (statistic from a sample) that was random. Randomness, that is either your data comes from a random sample or from a randomized experiment, is an important prerequisite.

2. A Real Life Example

Up until now, we have information on the "box", but in reality, we almost never do. So, statisticians substitute statistics from the sample and "pretend" that they actually have parameter values. In other words, statistics substitute for parameters. Here is an election example. On November 6, 2000 the night before the election, pollsters had given Gore 48% of the vote in Florida to Bush's 46%. The margin of error was 4%. Approximately 600 voters were randomly sampled and then polled.

3. Calculating the "Margin of Error"

From the sample, let's treat votes for Gore as a "1" and votes for everyone else as a "0" we can calculate a SD ((1-0) $\sqrt{.48 * .52}$ *which is about .5) and then the SE of a percentage. For the poll the standard error is* $\sqrt{600}$ *x .5 or about 12.25 voters or 2.04% (12.25/600 times 100 to make it a percentage) of the sample. They rounded it to about 2 percentage points. And then multiplied the 2 percent by 2 and then reported "+ or - 4 percentage points". What are they doing?*

4. Confidence Interval Basics (21.2)

A CONFIDENCE INTERVAL is a range of values (i.e. values derived from sample information) which we think covers the true parameter. The Associated Press reported that the "margin of error" was about 4% for the likely voters in Florida. This suggests a range around the sample statistic of 52% to 44% for Gore on the night before the election. This interval is supposed to covers Gore's true share of the vote. This is about plus or minus 2 Standard Errors and is the way the media expresses results from polls. What they are saying is that they were "95% confident that the interval 44% to 52% covers the true percentage of the vote for Gore in Florida".

It was NOT CLEAR enough to say that Gore would win from this result…just because Bush might have as little as 42% (and lose) or as much as 50%. It's interesting to note that as of today (this could change) Gore actually wound up with 48.84%…so a little more than predicted but 300 votes less than Bush. By the way, Bush has 48.85% as of today.

The figures 48% plus or minus 4% are confidence intervals for the population percentage and they are calculated from sample percentages and sample standard deviations. Up until now, we've been in a situation where we know exactly what the "box" looks like, now we don't, but we have samples which can reveal "the truth" (i.e. the parameter).

5. Properties of Confidence Intervals

In about 68% of all samples, the sample percentage will be within one standard error of the population percentage. From the poll, we would say that we were 68% confident that Gore's percentage of the vote is in the interval 46% to 50% In about 95% of all samples, the sample percentage will be within two standard errors of the population percentage. From the poll, we would say that we were 95% confident that Gore's percentage of the vote is in the interval 44% to 52% In about 99% of all samples, the sample percentage will be within three standard errors of the population percentage. From the poll, we would say that we were 99% confident that Gore's percentage of the vote is in the interval 42% to 54% You can never been 100% confident. There is always the chance that you could have a very bad sample and know nothing about the true population parameter.