1. 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).

2. 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.

3. Constructing a confidence interval for a population parameter involves five steps:

(1) Find the sample statistic. This is our ESTIMATE of the population parameter. From the example above, Gore got 48% of the vote according to the survey taken the night before the election in Florida.

(2) Compute the standard error for the sample statistic; for simple random samples involving percentages, the standard error is

$$\frac{\sqrt{\text{Sample Size}} \;*\; \text{Standard Deviation of "the box"}}{\text{sample size}} * 100$$

For example, from the Associated Press Poll:

$$\frac{\sqrt{600}\times(1-0)\sqrt{.48*.52}}{600}\approx 2\%$$ and this 2% is the Standard Error for the Percentage (Chapter 20.2)

(3) Find the level of confidence you are interested in from a normal table using the area percentages. For an approximate 95% confidence interval, a z=2.00 will give 95% in the center.

(4) Multiply (2) and (3).

*(2 * 2%)*

(5) Add and subtract (4) from (1). This is your "margin of error" that is, how accurate you believe your statistic is based on the variability of the estimate.

48% + or - 4%

4. Remarks

a. A typical confidence interval has the form "estimated value, plus or minus Z times the SE of the sample". In other words, an estimate plus and minus some multiple of the standard error for the particular statistic. In chapter 21 the statistic used is a percentage.

b. If the original population is normally distributed with a known standard deviation, or if the sample size is "large", then the distribution of the sample percentage is normal, and the appropriate test statistic is thus z from the normal table. (If the original distribution is normal with an unknown standard deviation, the test statistic is different.)

c. Your margin of error will depend on the choice of a confidence level. A lower confidence will give you a smaller margin of error. A higher confidence will give you a larger margin of error.

d. If your standard deviation is small, it is easier to get a more precise fix on the parameter. Your margin of error is smaller for populations with smaller standard errors.

e. If your n increases in size, it will reduce your margin of error. If your n gets smaller, it will increase your margin of error.

5. Interpretations

A. The CORRECT interpretation for a confidence interval is as follows: "We did a procedure of drawing a sample, computing a percentage and a standard error, etc. This procedure will give us a correct interval X% of the time and an incorrect interval 100-X% of the time. We hope this is one of the correct times. Thus, for about X% of all samples, the interval "sample percentage + or - Z standard errors covers the true population percentage.

B. It is WRONG to talk about the chance a particular confidence interval contains the parameter. For example, you can't say "there is a X% chance that the parameter is in the confidence interval" because these confidence intervals vary with samples and the parameter never varies. Any single confidence interval either covers the true parameter or it does not. See page 385 of your text.

C. Another way you might think about this. When you KNOW the TRUE POPULATION PARAMETER, you can make a statement like: there is a 95% CHANCE that the SAMPLE STATISTIC will be in the range of the parameter plus or minus two standard errors.

Example: if you know the parameter is 40% and the SE is 2.5%, then there is a 95% chance that the sample percentage will be in the range of 40% plus or minus 5%.

But when you DO NOT KNOW THE TRUE POPULATION PARAMETER, you are forced to make statements like this: I am 95% confident that the POPULATION PARAMETER is in the range of the statistic plus or minus two standard errors.

Example: if you don't know the parameter but you know the sample statistic is 40% and the SE is 2.5%, then you are 95% confident that the parameter is covered by the range of 40% plus or minus 5%.