A CONFIDENCE INTERVAL is a range of values (i.e. values derived from sample information) that we think covers the true parameter. The figures 36% plus or minus 6% 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). Last time, I suggested that in the media, they report a margin of error that is the equivalent of a "95% confidence" interval in this textbook.

1. SO WHERE DO YOU THINK THEY GOT THE 95% FROM? WHERE HAVE YOU HEARD IT BEFORE?

From the normal! Review pages 355-359 if this is unclear. If we could sample and resample using samples of the same size, in the long run (or infinitely), and if we plotted the outcomes, we would see a normal curve arise. Our statistics from our one sample is one possible outcome – but it is somewhere on the normal curve. We are just uncertain as to where exactly…but we can be 95% confident that the range we gave (based on our statistic) contains the true percentage (the number we really want).

2. Properties of Confidence Intervals

In about 68% of all samples, the sample percentage will be within one standard error of the population percentage. So from the poll outcome, we would say that we were 68% confident that between 33% and 39% of African Americans believe that race relations have improved in Los Angeles (that's ½ of + and - 6%)

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 (this is also the standard margin of error reported in the media) that the percentage is between 30% to 42%

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 the percentage is between 27% and 45%

You can never been 100% confident. There is always the chance that you could have a very bad sample and are nowhere near the true population parameter.

Please note that the parameter is fixed and unchanging. Our sample statistics will change from sample to sample. See the figure on page 385 in your text. If 80 is the parameter, the lines represent confidence intervals for 100 different samples. Notice that a few never "cross" the line.

a. A typical confidence interval has the form "estimated value, plus or minus Z times the SE of the sample". In other words, a confidence interval is 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 z from the normal table. (If the original distribution is normal with an unknown standard deviation, the test statistic is different – but that is an advanced topic not for this class.)

c. The size of the interval will depend on the choice of a confidence level. A lower confidence will give you a narrower interval. A higher confidence will give you a larger interval for the same sample size.

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

e. If your sample size, denoted n, (or number of draws) increases in size, it will reduce the size of your interval. If your n gets smaller, it will increase your interval size.

3. 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%.