What Does The C.I. Really Mean?

Do we have good reason to think that the true value of d is in the 95% C.I. (5461, 5559)? Is this a mathematically justified conclusion?

Suppose you know that the other operator lives in a city, but you don't know which one. You look in an atlas, and find that there are only three cities in Mongolia, with distances from us in kilometres of

5103, 5441, 5690

These distances are accurate to about ± 5 kilometres. What should you think after learning this?

Confidence Interval for a Proportion

We can consider the proportion of units having some characteristic to be the mean of a variable that takes on values 0 and 1. This lets us use the method we've seen to find a confidence interval for the proportion in the population from the sample proportion.

One complication: We've assumed that we know the standard deviation of a single observation. But for Bernoulli $(0-1)$ data, this standard deviation is

$$\sigma \;=\; \sqrt{\pi(1-\pi)}$$

So to know σ, we'd need to know π, in which case we wouldn't need to do any of this.

Fortunately, a fairly good guess at π will be good enough. We can substitute the sample proportion, p, and estimate σ to be $\sqrt{p(1-p)}$.

How To Find a Confidence Interval Of the Kind We've Seen So Far

To find a level C confidence interval for a mean (or proportion, for $0-1$ data), based on n observations:

1. Determine the standard deviation, σ, of a single observation. This might be known beforehand, or for $0-1$ data, be estimated from the sample proportion as $\sqrt{p(1-p)}$.
2. From this, compute the standard deviation of the sample mean (or sample proportion). This is called the *standard error*:
$$SE \;=\; \sigma/\sqrt{n}$$
3. Find the value z^* such that a standard normal variable has probability C of being in the interval $(-z^*, +z^*)$.
4. The confidence interval extends a distance z^*SE below and above the sample mean (or sample proportion).

Example: Election Polls

Reports of election polls with samples sizes of around 1000 often say something like:

> Results are accurate within plus or minus 3%, nineteen times out of twenty.

This is a claim that the interval from the result minus 3% to the result plus %3 is a 95% confidence interval.

Assuming that a simple random sample was used, is this claim correct?

Let's assume the worst case, which is when $\pi = 1/2$, and hence $\sigma = \sqrt{\pi(1-\pi)} = 1/2$. With $n = 1000$, the standard error is $(1/2)/\sqrt{1000} = 0.0158$. For a 95% C.I., $z^* = 1.960$, so the confidence interval extends above and below the sample proportion by $0.0158 \times 1.960 = 0.031$, or 3.1%.