A Justification for the ME of a Single – Proportion CI

Man, oh man, I wish that we were on a semester system instead of quarters. Then, MTH 243 and MTH 244 would be wonderful, 16 – week long classes, and I could spend quality time having you all collect data and really analyzing ***why*** things work the way they do in our stat world.

For instance...when estimating a CI for a proportion, we stated (rather brazenly) that the ME for such a CI is $\sqrt{\frac{\hat{p}\hat{q}}{n}}$. I casually mentioned that the proof of such a fact was here, and we all went about our day. Well, here's the proof, my friends...

This argument, like a few before, can be best viewed as an expectation, or an expected value. This important topic is never given the time it's allowed in MTH 243. Basically, expectation is best defined as "on average, what should happen under a certain set of assumptions?" Well, for a study where we assume that there is a certain proportion of the population, ***p***, has a certain characteristic, we know, since ***p*** is binomial, that $\mu = np$ and $\sigma = \sqrt{npq}$ (actually, you may have forgotten that; not to worry).

Those two parameters are for the entire population, however. We're only viewing a sample proportion, which we call $\hat{p}$. However, assuming that we randomly sample well, we can expect $\hat{p}$ to have the same numerical value as ***p***...and that's written $E\left[\hat{p}\right] = p$.

This should make sense, even without the new notation...on average, we should expect the sample averages to descend onto ***p***...in symbols, that means that $\frac{\mu}{n} = \frac{np}{n} = p$.

Now, that's a pretty gutsy statement; as you remember from class, just because we ***expect*** $\hat{p}$ to equal ***p*** doesn't mean it ever ***will***; it'll just be *within an acceptable margin of error*. In the population, this standard deviation, as mentioned above, is $\sigma = \sqrt{npq}$. Per sample, however, we must adjust this measurement. That is, the expected variation (otherwise known as...you guessed it...the margin of error) will be averaged from this population parameter, as follows (with a little help from algebra II): $E[\sigma] = \frac{\sigma}{n} = \frac{\sqrt{npq}}{n} = \sqrt{\frac{npq}{n^2}} = \sqrt{\frac{pq}{n}}$.

And, since we don't know what ***p*** (or ***q***) are, we use the values of $\hat{p}$ and $\hat{q}$ in these formulas for the mean and standard deviation, and claim that our CI will have a center at $\hat{p}$ and a spread of $\sqrt{\frac{\hat{p}\hat{q}}{n}}$.

And there you have it. What I wouldn't give for 6 more weeks....just 6 more. Think of the damage we could do with that kind of time!