Single Proportion CI Exercises

Pages	Suggested Reading
319	Section 8.1.2 (only the first three paragraphs of it, though…ending with "*…is used with these intervals*.")
331 - 335	Sections 8.4 (notation note…your book notes the sample proportion as ***p'*** …I was trained to call it $\hat{p}$. "p hat", as I call it, seems to be more ubiquitous, but hey…whatever.

Pages	Problems
343	1 (note: when the "error Bound" is mentioned, your authors simply mean "Margin of Error"), 10, 11, 16, 18, 20, 24 – 26

Addenda Videos
• Justification for the Margin of Error around $\hat{p}$ • Derivation of the Sample Size Calculation (2 parts)

E1. In the Harris Interactive report titled, "***Video Game Addiction***", this claim appears: "*In [the] US, 8.5% of youth gamers (ages 8 to 18) can be classified as pathological or clinically 'addicted' to playing video games*. " This result is (ostensibly) based on the fact that 100 out of the 1178 eight – to – eighteen year olds surveyed displayed such addictive behaviors.

a. Where did Harris get their "8.5%" figure from?

b. What should have accompanied the 8.5% point estimate? Hint: use 95% confidence, like the rest of the world usually does. ☺

c. Does this CI allow us to conclude (with 95% confidence) that more than 8% of video – playing youngsters display behavioral signs that indicated addiction?

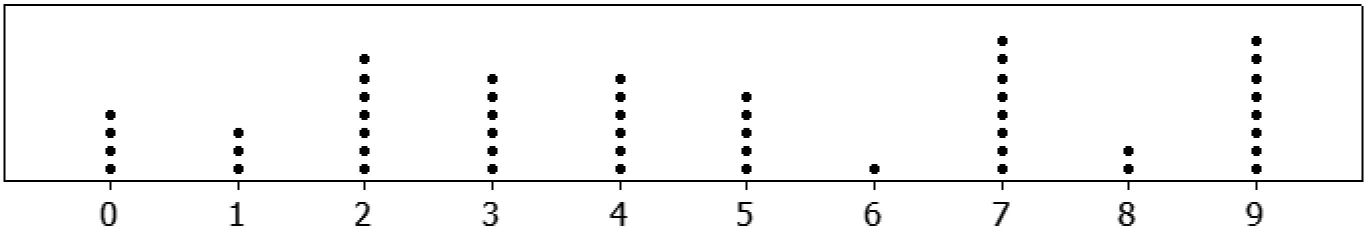
d. Does this CI allow us to conclude (with 95% confidence) that more than 5% of video – playing youngsters display behavioral signs that indicated addiction?

e. Does this CI allow us to conclude (with 95% confidence) that fewer than 9% of video – playing youngsters display behavioral signs that indicated addiction?

f. Does this CI allow us to conclude (with 95% confidence) that fewer than 11% of video – playing youngsters display behavioral signs that indicated addiction?

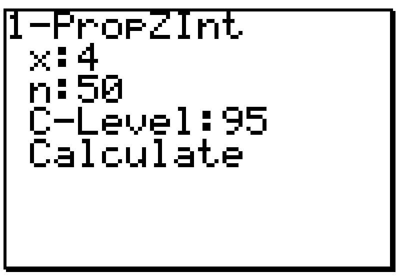
g. For part c.…suppose you really, really, really wanted (for whatever reason) to say, "more than 8% display behavioral signs that indicated addiction." What's the only legitimate statistical method to achieve this?

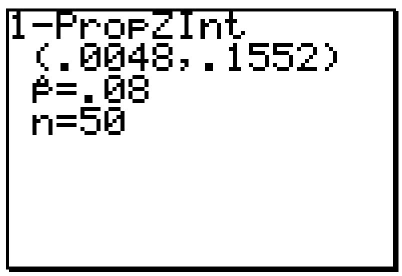
Suppose you were testing a randomizer for a website programmer (I've had to do this occasionally). You gain a copy of the randomization code and run it a few times. Here are the results you get:



First, a bit of explanation: this randomizer contains 10 digits (0 through 9). Assuming that it's working properly, each of the 10 digits should appear approximately 10% of time.

E2. For each of the ten digits (0 through 9), construct the 95% CI of its "percentage of appearance", based on the number of appearances shown in the dotplot. I'll do the one for 0 to get you started (remember to press [STAT] [◄] [ALPHA] [MATH] to get **1-PropZInt)**:





So, I'm 95% confident that the digit 0 should appear between 0.5% and 15.5% of the time.

E3. Now, assuming the randomizer is working properly, each digit should appear around 10% of the time. Of course, we wouldn't expect them to always appear 10% of the time…which is why we created the "wings" in E1. So long as 10% is within the "wings" of the 95% CI, we conclude that there is no reason to believe that the digit is not appearing 10% of time. For example, the digit 0's CI contains 10%, so I feel 95% sure that it's appearing with correct frequency.

Check each of your CIs…does any digit appear too frequently or too infrequently?

E4. Good work! Now, there's one thing left to consider. In this example, we created ten 95% confidence intervals. I want us all to remember what "95% confidence" means: it means, if we were to properly, randomly sample data 100 times, then, in 95 of those times, we'd correctly capture the parameter we were trying to measure (in this case, the proportion of appearance of the digits). That means that we'll miss that parameter 5% of the time, even if we do the sampling perfectly, we can still be off due to random chance. In our old probability lingo:

P(we trap the true parameter) = 0.95

So, here's the bad news…since we did ten 95% CIs in the previous problems, we're not at 95% confidence anymore. This is a joint probability. So, here's where we're at now:

P(we trapped the true parameter of each of the ten digits) = P(we trapped 0's parameter AND 1's AND 2's AND 3's AND 4's AND 5's AND 6's AND 7's AND 8's AND 9's)

Remember what "and" means in probability? For independent events (which these are) it means to multiply the individual, independent probabilities. Each of those probabilities are 95% (since they're all 95% CIs). So, at what confidence were we operating when we formed ten 95% CIs independently (called the "family wise confidence level")?

E5. So, that's the chance that we're sure that each of the ten digits is behaving the way we said they were behaving (the "confidence" in our results). What's the chance we're wrong in at least one, simply due to random chance, and not carelessness nor bad data collection (called the "family wise error rate")?

E6. What confidence would you need to be achieving in order to keep your family wise confidence level at least at 95%?

So, you can see that comparing multiple independently sampled datasets causes quite a problem with the confidence in your answer. Fortunately, there are tools to deal with just this, and we'll see some of them in the coming weeks!

Answers.

E1.

- **a.** They constructed the point estimate $\hat{p}$.
- **b.** It needs a margin of error! But…***why***? Remember why this is so vitally important?
- **c.** Nope…since 8% is within the interval, and there are some percentages below it, (i.e., 7%), we can't say "more than 8%".
- **d.** Yep!
- **e.** Nope!
- **f.** Yep!
- **g.** Well, the problem with part c's 8% is that it's within the ME for the CI. You need to shrink the ME down until it's less than 0.5% (see why?). Now…how can you legitimately do this?

E2. I think you can handle these!

E3. These too!

E4. $(0.95)^{10} \approx 60\%$ confidence. Yikes.

E5. 1 – 60% = 40%. It's ***almost*** as bad as a coin flip!

E6. Since I don't; know what the confidence is, I'll call it ***C***. Using the ideas from E4, we can solve $(C)^{10} = 0.95$. I get around 99.5% confidence. However, as you will see, a confidence level this high can raise another problem…but we'll

Single Proportion CI Quizzes

Quiz 1.

1. (**3 points**) Suppose a candidate is planning a poll and wants to estimate the percentage of voters that support him, with a ME of no more than 2.5% and 95% confidence. How large a sample is needed?

2. (**3 points**) Now, suppose this candidate randomly polls 6000 folks (well above the necessary number), and finds that, of these, 3201 support him. Construct the 95% CI for his voter support percentage.

3. (**1 point**) Based on your previous CI, can your candidate be 95% confident that the majority (that is, more than 50%) of potential voters support him?

4. (**3 points**) Why or why not? Make sure you use the number 50% (or 0.5) in your answer.

<u>Quiz 2.</u>

Suppose a certain CI (based on a certain sample size, ***n***) has a MOE of 8%. Increasing the sample size would, clearly, decrease the MOE.

1. (**2 points**) What would you have to do to the sample size to shrink the MOE from 8% to 4%? I know I didn't give you an actual number for the sample size; I'm looking for an answer like "it needs to be doubled" or "it needs to be squared" …something along those lines.

2. (**2 points**) What would you have to do to the sample size to shrink the MOE from 8% to 2%?

3. (**2 points**) What would you have to do to the sample size to shrink the MOE from 8% to 1%?

Now, any study like this one has a "response rate". Simply put, this is the percentage of those sampled that actually follow through with the entire study well enough so that the researcher can get data from them. From this source (http://www.gallup.com/poll/7510/looking-closely-survey-response-rates.aspx), it seems that, as of a few years ago, a 40% response rate wasn't unheard of. So, if you ran a sample size calculation and it told you that you needed 1000 subjects, only 400 of them (on average) would complete the study.

4. (**2 points**) If you had a response rate of 40%, but needed to actually have 1000 subjects at the end of your study, how many potential subjects would you have had to sample up front?

And you can see how this gets worse as the response rates go further and further down.

Now, in and of itself, low response rates aren't bad – we can do pretty good statistics on small samples. But there can be implications of low response rates (sometimes called "nonresponse bias")that can negatively impact the final results of a study.

5. (**2 points**) Read the article linked above and give one example where nonresponse bias can have negative consequences on the outcome of study.

Quiz 3.

1. **(1 point)** The size of the Margin of Error (MOE) of a single proportion CI is dependent upon the value of the sample proportion (AKA, $\hat{p}$). **True** or **false**?

2. **(1 point)** The size of the MOE of a single proportion CI is dependent upon the sample size ***n***. **True** or **false**?

3. **(1 point)** The size of the MOE of a single proportion CI is dependent upon the population size (***N***). **True** or **false**?

4. **(1 point)** The size of the MOE of a single proportion CI is dependent upon the confidence level (usually notated as "1 - α"…we'll explain in class). **True** or **false**?

5. **(1 point)** If a study gives you only $\hat{p}$, you could correctly construct a 95% CI around that $\hat{p}$ value . **True** or **false**?

6. **(2 points)** Refer back to the previous question…if you answered "true", then construct the CI (make up a value for your $\hat{p}$, and tell me what it is). If you answered "false", explain why you can't.

7. **(1 point)** If a study gives you $\hat{p}$ and the sample size used, you could correctly construct a 95% CI around that $\hat{p}$ value . **True** or **false**?

8. **(2 points)** Refer back to the previous question…if you answered "true", then construct the CI (make up a value for your $\hat{p}$ and ***n***, and, again, tell me what they are). If you answered "false", explain why you can't.