Math 244 Lecture Notes

Chapter 18: Confidence Intervals for 1-Proportion

Overview: Today, we will look at how information about samples (**statistics**) can be used to make predictions about information about larger populations (**parameters**). This will be done using confidence intervals (Pg. 472 to 485)

to A **statistic** is a number that describes a sample. Examples of statistics include $\bar{x}$ (mean), s (standard deviation), and $\hat{p}$ (probability/proportion).

A **parameter** is a number that describes a population. Examples of parameters include μ (mean), σ (standard deviation), and p (probability/proportion)

We often want to know something about an entire population (e.g. what is the average salary for all Americans?) but it's not always possible to sample the entire population. In these cases, we rely on hypothesis tests and confidence intervals.

Example 1. Label the values using the appropriate parameters and/or statistics: It is claimed that the average weight of all cats is 10 lbs with a standard deviation of 3 lbs. Mr. Wherry's cats have an average weight of 15 lbs with a standard deviation of 1 lb.

ANS: All refers to a population. We use parameters to describe the population of all cats. $\mu = 10$ lbs with $\sigma = 3$ lbs. In contrast, Mr. Wherry's cats are merely a sample of all cats. So, $\bar{x} = 15$ lbs and $s = 1$ lb.

Example 2. Label the values using the appropriate parameters and/or statistics: Coca-Cola tells you that the odds of winning a game are 1 in 3. You manage to win 10 times in 25 plays.

Confidence intervals are guessing windows for what should happen for a population based off a sample. In it's most general form, we have that a confidence interval looks like

$$\text{Center} \pm \text{Margin of Error} = \text{Center} \pm (\text{Distribution})^* \, SE.$$

One of the most common normal models that we will use is

$$\hat{P} = N(p, \sqrt{\frac{pq}{n}})$$

where p is the center, N is the normal model telling us to use Z scores, and $\sqrt{\frac{pq}{n}}$ is the standard deviation. Rewriting this information to be based off statistics instead of parameters involving changing p's to $\hat{p}$. So our interval of

$$\text{Center} \pm (\text{Distribution})^* \, SE$$

becomes

Confidence Interval The interval

$$\hat{p} \pm Z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

gives us a guessing range for what we expect p to be in based off the SAMPLE information.

The assumptions are...

(a) Random: These means there is no bias in how the data was collected. No point in doing Stats if you are misrepresenting the population.

(b) Independence/10% Rule: Sample less than 10% of the population or make sure that one outcome does not affect another. If there is dependency, the probability will be changing. This makes predictions much harder to do.

(c) Large n: It takes time for the distribution to look normal. We need ten successes and failures before that happen. Number of successes: $n\hat{p} \geq 10$ and number of failures $n\hat{q} \geq 10$.

Example 3. Using a random sample, an employee at the PCC bookstore finds that 97 out of 112 customers use their credit card to pay for books. Create a 95% confidence interval for what should go on for ALL PCC students.

ANS: Our success rate is $\hat{p} = \frac{x}{n} = \frac{97}{112}$ for those that used a credit card. Our failure rate would be the other 15 people out of 112 that did NOT use a credit card to pay for books. $\hat{q} = \frac{15}{112}$. $n = 112$. To get about 95% percent of the area underneath the normal curve, we need to go out about 2 standard deviations (68/95/99.7 Rule). More accurately, $Z^* \approx 1.96$.

$$\hat{p} \pm Z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \frac{97}{112} \pm 1.96 \sqrt{\frac{\frac{97}{112}\frac{15}{112}}{112}}$$

Typing this in our calculator once with the + and once with the − gives us a Low: 0.803 and High: 0.929. "We are 95% confident that the interval of 80.3% to 92.9% contains the proportion of ALL PCC students that use a credit card at the bookstore."

Example 4. You try: Suppose that Sasha can successful throw a snowball through a tire swing 60 out of 75 times in a recent sample. Using this information, estimate her population success rate by making a 95% confidence interval. [Do on scratch paper]

NOTE: We are confident in the intervals ability to capture the true probability. 95% of the time that we create intervals like this using the same sample size, we should capture the true value for p. This tells us that 5% of the time we will be wrong. Error in Stats is unavoidable. We can only hope to understand what potential for error exists.

A little terminology:

The **critical value** is denoted by z^*. It's the z-score corresponding to a specified confidence level.
The **standard error** is our guess for the SD based off the sample. $SE = \sqrt{\frac{\hat{p}\hat{q}}{n}}$
The **margin of error** is given by $MOE = z^* SE(\hat{p})$, or $ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$.
The **confidence interval** for the true value of a proportion is determined by:

best point estimate (center) $\pm$ margin of error i.e. $\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Example 5. Let's examine some current data. As of today, Gallup Daily estimates, with 95% certainty, that 29.1% of US employees are actively engaged at work give or take 3%. Identify all key terms

ANS: We have an estimate of 29.1% for our center. We call this our best point estimate. In other words, $\hat{p} = 0.291$.

The "give or take" tells us that we are looking at the wiggly room. This is exactly what the Margin of Error does. So, $MoE = 0.03$.

They are using 95% confidence, so our $C = 95\%$. This gives us a Z^*, or Critical Value, of 1.96.

NOTE: We could make a confidence interval by taking the center and adding and subtracting the MoE. So, our 95% CI here would look like 0.291 ± 0.03. The critical value is built in the MoE.

Example 6. How many people did Gallup survey?

ANS: I'll set this up, but double-check my solving. The MoE is given by $ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$. We plug in our success rate $\hat{p} = 0.291$, failure rate $\hat{q} = 1 - 0.291 = 0.709$, critical value 1.96, and MoE of 0.03 to get...

$$0.03 = 1.96 \sqrt{\frac{0.291 * 0.709}{n}}.$$

Solving for n (process on pg. 482 if you need to review), the number of samples, gives us

$$n = \frac{1.96^2 * 0.291 * 0.709}{0.03^2} \approx 881 \text{ people.}$$

Okay, but what about $C = 90\%$? What's Z^*?

Here's the general process...

- Place the area C in the center of a normal curve.
- Determine the area in BOTH tails, α. Then, determine the area in the left-tail.
- Use the table to get the z-score OR InvNorm(area). Switch the sign.

Example 7. Find Z^* for $C = 95\%$.

ANS: We start with 95% in the center of our curve. This leaves $\alpha = 1 - 0.95 = 0.05$ for both the remaining tails. Or, we have $\alpha = 0.05/2 = 0.025$ in just the left tail. For today, just use this online calculator:
`http://stattrek.com/online-calculator/normal.aspx`
Type your $\alpha/2$ in the "Cumulative Probability" spot and hit calculate. Then, take the "Standard Score" but write it as a positive. For example, I typed in 0.025 in that second space, hit enter, and I got a $Z^* = 1.960$. Reload between uses.

Example 8. Find Z^* for $C = 90\%$ and $C = 99\%$

ANS: You should be getting $Z^* = 1.645$ and $Z^* = 2.576$ respectively. Try it out.

NEXT STEPS:

(a) Log into MyStatLab and do the HW for Ch. 18 OR do it from the book.

(b) Watch the Week 1 videos in MyStatLab. You can sign up for a two-week free trial. I highly recommend doing that. In particular, I'd go to "Homework", "Week 1 Multimedia", and finally "Step-by-Step Example: A Confidence Interval for a Proportion".

(c) Watch some videos I made here:
   - General Idea: `https://youtu.be/QRlWfODRfLo`
   - TI-83/84 help: `https://youtu.be/A2fqAlmm3W0`
   - TI-89 help:`https://youtu.be/kXZCMrErE4s`

(d) Look up videos on YouTube. For example:
`https://youtu.be/Vgkyn39DdZM`
is okay even though the terminology doesn't perfectly match up.