MODULE 5 EXAMPLES

TAs: Nihal Reddy
Email: nireddy@ucsd.edu
OH: Thursdays 6-7pm

Slide Credits: Kira Fleischer

PROBLEM #: KEY TOPICS FROM PROBLEM

Problem setup and description.

Question

Key notes from readings/lectures needed to answer the question

Solution: written with as much detail as we expect you to give on your homework sets

PROBLEM 1: SAMPLING BIAS

Shawn is the host of a podcast and he wanted to know how many people like his show. He decided that he would do this through an online questionnaire. During his podcasts, he asked his listeners to fill out the questionnaires which ask them whether they like Shawn’s show. **Please discuss possible biases in this sample.**

Solution: He will only get responses from those people who listen to his podcast and won’t get any opinions from those who don’t like his show.

PROBLEM 2: SAMPLE PROPORTIONS, STANDARD ERROR

A random sample of 100 parents of children in an elementary school were asked whether they spend a fair amount of time with their kids over the weekend. About 47% of the parents said that they are too busy with work to spend time playing with their children.

(a) Explain what $\hat{p}$ means in this context.

$\hat{p}$ refers to the sample proportion and is an estimate of the true population proportion p

Solution: $\hat{p} = 0.47$ means that 47% of the 100 parents in our sample said they don't have time to play with their children. This is an estimate of p, the true proportion of all parents of the children in this elementary school who would say that they don't have time to do so.

PROBLEM 2: SAMPLE PROPORTIONS, STANDARD ERROR

A random sample of 100 parents of children in an elementary school were asked whether they spend a fair amount of time with their kids over the weekend. About 47% of the parents said that they are too busy with work to spend time playing with their children.

(b) Calculate the standard error of $\hat{p}$.

$$SE(\hat{p}) = \sqrt{\frac{p*(1-p)}{n}} \approx \sqrt{\frac{\hat{p}*(1-\hat{p})}{n}}$$

Solution: $SE(\hat{p}) = \sqrt{\frac{0.47*0.53}{100}} = 0.050$

PROBLEM 2: SAMPLE PROPORTIONS, STANDARD ERROR

A random sample of 100 parents of children in an elementary school were asked whether they spend a fair amount of time with their kids over the weekend. About 47% of the parents said that they are too busy with work to spend time playing with their children.

(c) Explain what the standard error means in this context.

The standard error is an estimate of the amount of the variation in the sample proportion $\hat{p}$ we expect to see from sample to sample when we perform this same sampling procedure several times.

Solution: The standard error is the estimate of the amount of the variation in the sample proportion we expect to see from sample to sample when we ask 100 parents whether they spend time with their children over the weekend.

PROBLEM 3: CENTRAL LIMIT THEOREM

According to a survey, around 46% of 300 students said that they do plan on travelling over the summer. Assume that this sample was taken randomly. **Is it reasonable to use a normal model for the distribution of the sample proportion? Explain.**

We can use a normal model for the distribution of the sample proportion if we meet the two conditions for the Central Limit Theorem (CLT):

1. Independence: observations in the sample must be independent
   - The most common way for observations to be considered independent is if they come from a simple random sample
2. Success-Failure Condition: the sample size must be sufficiently large
   - Satisfy these inequalities: $np \geq 10,$ and $n(1-p) \geq 10$

Solution: Since this sample is random, we have satisfied the condition for independence. Now we check $300 * 0.46 = 138 \geq 10$, and $300 * 0.54 = 162 \geq 10$. Therefore the success-failure condition is also met. Thus it is safe to apply CLT and use a normal model for the distribution of the sample proportion.

PROBLEM 4: INTERPRETING CONFIDENCE INTERVALS

In a survey where parents were asked about whether they spend a fair amount of time with their children over the weekend, a 95% confidence interval is constructed to be [44%, 50%]. **Interpret the interval in this context.**

Solution: If we were to repeat this procedure many times, then approximately 95% of the confidence intervals we construct would include the true proportion of parents who spend a fair amount of time with their children over the weekend.

OR

We are 95% confident that the true proportion of parents who spend a fair amount of time with their children over the weekend is between 44% and 50%.

PROBLEM 5: INTERPRETING CONFIDENCE INTERVALS

In a Statistics class, the professor is interested in whether students want to have quizzes biweekly to better understand the material. She took a random sample and the 95% confidence interval she constructed for the proportion that would like biweekly quizzes is 20% ± 3%. **Interpret whether the following statements are true. Explain why.**

(a) We are 95% confident that the true population proportion is between 17% and 23%.

Solution: Yes, this is a correct way to interpret confidence intervals.

(b) The true population proportion p will be in [17%, 23%] with probability 0.95.

Solution: False. Once a confidence interval is constructed, there is no randomness anymore, so we cannot talk about probability/chance.

(c) If we repeatedly take a random sample and construct a CI with this confidence procedure, 95% of the intervals would contain the true proportion.

Solution: Yes, this is a correct way to interpret confidence intervals.

(d) 95% of all random samples of students will show that 21% of the students want to have biweekly quizzes.

Solution: False. Different samples will give different confidence intervals; 95% is a proportion of all intervals constructed in this confidence procedure, but not of the random samples.

PROBLEM 6: MARGIN OF ERROR

In constructing a confidence interval, the margin of error (MOE) is defined as $z* \times$ SE. You are interested in what percentage of undergraduate students are planning on going to graduate school. A survey showed that 30% have this plan based on a random sample of 200 students. Now, you'd like to repeat the study interviewing n random undergraduate students. **Find n if you'd like to obtain a 99% CI with a 5% MOE.**

$MOE = z^* * SE = z^* * \sqrt{\frac{\hat{p}*(1-\hat{p})}{n}}$; z^* for 99% CI is 2.58

Solution: Using the value of z^*associated with a 99% CI ($z^* = 2.58$), we get

$$0.05 = 2.58 * \sqrt{\frac{0.3*0.7}{n}}$$

Solving for n, we get $n = \frac{2.58^2(0.3)(0.7)}{0.05^2} = 559.14$. Hence, we would need to sample 560 undergraduate students.

PROBLEM 7: CONSTRUCTING CI'S

A research firm wants to find the proportion of people in large cities who have a cell phone. They took a random sample of 400 people and 382 of them said they do own a cell phone.

(a) What is the standard error?

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}*(1-\hat{p})}{n}}$$

Solution: $\hat{p} = \frac{382}{400} = 0.955$

$$SE(\hat{p}) = \sqrt{\frac{0.955*0.045}{400}} = 0.010$$

PROBLEM 7: CONSTRUCTING CI'S

A research firm wants to find the proportion of people in large cities who have a cell phone. They took a random sample of 400 people and 382 of them said they do own a cell phone.

(b) Check the conditions for constructing an 85% confidence interval.

To construct a confidence interval, we are inherently assuming the sample can be modeled with a normal distribution. Thus, we must satisfy the conditions of the CLT:

1. Independence: is there a random sample?

2. Success-Failure: $np \geq 10$, and $n(1-p) \geq 10$

Solution: Since this is a random sample, the independence condition of the CLT is met. The S-F condition is also met, since $400 * 0.955 = 382 \geq 10$, and $400 * 0.045 = 18 \geq 10$

PROBLEM 7: CONSTRUCTING CI'S

A research firm wants to find the proportion of people in large cities who have a cell phone. They took a random sample of 400 people and 382 of them said they do own a cell phone.

(c) If you were to construct an 85% interval, what is the value of z*?

You can find the value of z* in several wa

- Using Python, R, or other coding languages
- Using Excel (=NORM.S.INV(1 - 0.075))
- Using a graphing calculator
- Using the z-score table

```
[1] from scipy.stats import norm

# Calculate the z* value for an 85% confidence interval
confidence_level = 0.85
alpha = 1 - confidence_level  # Total area in both tails
z_star = norm.ppf(1 - alpha / 2)  # Z* for two-tailed test

z_star

1.4395314709384563
```

Solution: (explain the method you used) z*=1.44