MODULE 6 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: DEFINING HYPOTHESES FOR TESTING

Consider the following situations for which we would like to conduct rigorous hypothesis testing.

Write the corresponding null and alternative hypotheses:

(a) A factory is interested in whether the light bulbs they make are effective or not. They want to know if the percentage of the bulbs that are defective is below 5%.

Null hypothesis (H_0): often represents a skeptical perspective, a claim to be tested, or a perspective of "no difference".

Alternative hypothesis (H_A): represents an alternative claim under consideration and generally represents a new or stronger perspective.

Solution: Let p be the proportion of light bulbs they make that are not effective. We test H_0: $p = 0.05$ against H_A: $p < 0.05$.

PROBLEM 1: DEFINING HYPOTHESES FOR TESTING

Consider the following situations for which we would like to conduct rigorous hypothesis testing.

Write the corresponding null and alternative hypotheses:

Null hypothesis (H_0): often represents a skeptical perspective, a claim to be tested, or a perspective of "no difference".

Alternative hypothesis (H_A): represents an alternative claim under consideration and generally represents a new or stronger perspective.

(b) A town wants to know if the proportion of residents who support a piece of legislation is 68% or not.

Solution: Let p be the proportion of residents that support the law. We test $H_0: p = 0.68$ against $H_A: p \neq 0.68$.

PROBLEM 2: DRAWING CONCLUSIONS FROM HYPOTHESIS TESTS

In 2021, a research team in China was interested in whether birth rate had increased compared with that in 2018. They conducted a hypothesis test and computed that the p-value is 0.056.

(a) Do you think they should reject or fail to reject the null hypothesis?

p-value $> \alpha = 0.05$: fail to reject the null hypothesis

p-value $< \alpha = 0.05$: reject the null hypothesis

Solution: Since the p-value is greater than $\alpha = 0.05$, they should fail to reject the null hypothesis.

(b) Is it reasonable to conclude that the birth rate has increased since 2018? Explain.

p-value $> \alpha = 0.05 \Longrightarrow$ fail to reject the null hypothesis $\Longrightarrow$ we do not have strong evidence to make a conclusion

p-value $< \alpha = 0.05 \Longrightarrow$ reject the null hypothesis $\Longrightarrow$ we have convincing evidence to make a conclusion

Solution: Since we failed to reject the null hypothesis, we don't have strong evidence to conclude that birth rate has increased since 2018.

PROBLEM 3: CONSTRUCTING A CONFIDENCE INTERVAL

You are interested in conducting a hypothesis test based on a 95% confidence interval to see if the unemployment rate in a large city is 22% or not. You took a random sample of 60 people and found that 10 people are unemployed.

(a) Check the conditions required in order to proceed with the test.

To use a hypothesis test based on a confidence interval, we assume that $\hat{p}$ is approximately normally distributed. The sampling distribution (distribution of $\hat{p}$) is approximately normal if the conditions of the CLT are met:

1. Observations are independent: satisfied if the sample is taken randomly

2. Sample size is sufficiently large (aka Success-Failure condition): satisfied if $np \geq 10$ and $n(1-p) \geq 10$ (typically, we use $\hat{p}$ in place of p since we do not know the true proportion p)

Solution: This sample is taken randomly, so the independence condition is met. Also, $n\hat{p} = 10 \geq 10$, and $n(1-\hat{p}) = 50 \geq 10$, so the S-F condition is met.

PROBLEM 3: CONSTRUCTING A CONFIDENCE INTERVAL

You are interested in conducting a hypothesis test based on a 95% confidence interval to see if the unemployment rate in a large city is 22% or not. You took a random sample of 60 people and found that 10 people are unemployed.

(b) Compute the 95% confidence interval, interpret it in the context, and make a conclusion of whether you should reject or fail to reject the null hypothesis.

Confidence interval for population proportion: $\hat{p} \pm z^* \cdot SE_{\hat{p}}$

$SE_{\hat{p}} = \sqrt{\frac{\hat{p}*(1-\hat{p})}{n}}$; $z^* = 1.96$ (for 95% CI)

Solution: $SE_{\hat{p}} = \sqrt{\frac{\hat{p}*(1-\hat{p})}{n}} = \sqrt{\frac{0.167*0.833}{60}} = 0.048$. Thus, the 95% confidence interval is represented as: $\left(0.167 - (1.96 * 0.048), 0.167 + (1.96 * 0.048)\right) = (0.073, 0.261)$.

We are 95% confident that the proportion of people that are unemployed in the city is in between 7.3% and 26.1%. Since 22% is within this interval, we fail to reject the null hypothesis and cannot conclude that the unemployment rate is not 22% in this large city (at 95% confidence level).

PROBLEM 4: EVALUATING A HYPOTHESIS TEST

We are interested in whether the proportion of college students who think they are overweight is 35% or not. A random sample of 120 students were chosen and 40 of them said they think that they are overweight. A student conducted the following hypothesis test with some errors.

Could you identify the mistakes they made?

1. They identified the null and alternative hypotheses as $H_0: \hat{p} = 0.35$ and $H_A: \hat{p} > 0.35$.

2. They checked the conditions for using the CLT approximation: Random sample; $n\hat{p} = 40 \geq 10$

3. They calculated $\hat{p} = \frac{40}{120} = 0.33$; $SE = \sqrt{\frac{0.33*0.67}{120}} = 0.043$; $z = \frac{0.35-0.33}{0.043} = 0.465$; p-value: $P(Z > 0.465) = 0.32$

4. They concluded that we fail to reject the null hypothesis at significance level $\alpha = 0.05$, and we don't have strong evidence against the null hypothesis.

PROBLEM 4: EVALUATING A HYPOTHESIS TEST

We are interested in whether the proportion of college students who think they are overweight is 35% or not. A random sample of 120 students were chosen and 40 of them said they think that they are overweight. A student conducted the following hypothesis test with some errors.

Could you identify the mistakes they made?

1. They identified the null and alternative hypotheses as $\boldsymbol{H_0: \widehat{p} = 0.35}$ and $\boldsymbol{H_A: \widehat{p} > 0.35}$.

Solution: It should be p in the hypotheses, not $\hat{p}$. Also in the alternative hypothesis, it should be $\neq$ instead of $>$.

PROBLEM 4: EVALUATING A HYPOTHESIS TEST

We are interested in whether the proportion of college students who think they are overweight is 35% or not. A random sample of 120 students were chosen and 40 of them said they think that they are overweight. A student conducted the following hypothesis test with some errors.

Could you identify the mistakes they made?

2. They checked the conditions for using the CLT approximation: Random sample; $\boldsymbol{n\hat{p} = 40 \geq 10}$

There are two key differences when using confidence intervals vs the p-value method (hypothesis testing). When using hypothesis testing for proportions:

A) When checking the S-F condition for normality, use p_0 (not $\hat{p}$).

B) When calculating the SE, use p_0 (not $\hat{p}$).

This is because the sampling distribution is determined under the null proportion, so the null value p_0 is used for the proportion in these calculations rather than $\hat{p}$.

Solution: Use p_0, not $\hat{p}$ in the S-F condition calculation. That is, it should be $np_0 = 42 \geq 10$. Also, they missed the condition $n(1 - p_0) = 78 \geq 10$

PROBLEM 4: EVALUATING A HYPOTHESIS TEST

We are interested in whether the proportion of college students who think they are overweight is 35% or not. A random sample of 120 students were chosen and 40 of them said they think that they are overweight. A student conducted the following hypothesis test with some errors.

Could you identify the mistakes they made?

3. They calculated $\hat{p} = \frac{40}{120} = 0.33$; $SE = \sqrt{\frac{0.33*0.67}{120}} = 0.043$; $z = \frac{0.35-0.33}{0.043} = 0.465$; **p-value:** $P(Z > 0.465) = 0.32$

There are two key differences when using confidence intervals vs the p-value method (hypothesis testing). When using hypothesis testing for proportions:

A) When checking the S-F condition for normality, use p_0 (not $\hat{p}$).

B) When calculating the SE, use p_0 (not $\hat{p}$).

This is because the sampling distribution is determined under the null proportion, so the null value p_0 is used for the proportion in these calculations rather than $\hat{p}$.

Solution: Use p_0, not $\hat{p}$ in the SE condition calculation. That is, it should be $SE = \sqrt{\frac{0.35*0.65}{120}} = 0.044$. Consequently, the z-score should be $z = \frac{0.33-0.35}{0.044} = -0.45$, and the p-value should be calculated as $2 * P(Z < -0.45) = 0.654$.

PROBLEM 4: EVALUATING A HYPOTHESIS TEST

We are interested in whether the proportion of college students who think they are overweight is 35% or not. A random sample of 120 students were chosen and 40 of them said they think that they are overweight. A student conducted the following hypothesis test with some errors.

Could you identify the mistakes they made?

4. They concluded that we fail to reject the null hypothesis at significance level $\alpha = 0.05$, and we don't have strong evidence against the null hypothesis.

Solution: Their conclusion is valid based on their calculated p-value of 0.32.

PROBLEM 5: TYPE I AND II ERRORS

John is having mild symptoms recently and he wants to get tested for COVID-19. The null hypothesis is that he doesn't have COVID-19, while the alternative hypothesis is that he does have COVID. **What are the Type I and Type II errors that could occur based on his test result?**

Type I error: rejecting the null hypothesis (H_0) when H_0 is actually true (false positive)

Type II error: failing to reject the null hypothesis when the alternative (H_A) is actually true (false negative)

Solution: Type I error occurs when the test result says he does have COVID-19 when he doesn't (false positive); and Type II error occurs when the test says he doesn't have COVID-19 while he actually does have it (false negative).

PROBLEM 6: IDENTIFYING TYPE I AND II ERRORS

A hypothesis test is conducted to find out whether students in a university prefer Coke over Diet Coke. The study finds no evidence that the proportion of the people who prefer Coke is not 0.5. Later a company surveys all the students and finds that there is no difference in student preference. **Do you think a Type I error occurred, a Type II error occurred, or neither?**

Solution: No error has been made since the true proportion is 0.5, which wasn't rejected.