AP Statistics
Unit 06 – Day 06 Notes
Significance Tests: Proportions (One Sample)

Name Key
Period______

Significance tests are particularly useful when we must make a decision about a situation.
EXAMPLES:

- Is the defendant guilty or not?
- Should we choose print or television?
- Should I hate Tim Tebow as much as I do?

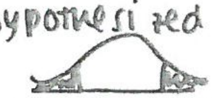
Significance Test: procedure for using observed data to decide between 2 competing claims.
→ Claims are often statements about a parameter

Null Hypothesis: starting hypothesis about a population parameter.
→ the claim we weigh evidence against
→ a statement of "no difference"

- Denoted by H_0
- Specifies a population model parameter of interest and proposes a value for that parameter
- Usually write the null hypothesis in the form: H_0: parameter (p or μ) = hypothesized value
- Purpose is to identify the parameter we hope to learn about and find a specific hypothesized value for that parameter

Alternative Hypothesis: the claim about the population that we are trying to find evidence <u>for</u>. (when the null isn't true.)

- Denoted by H_A
- Usually write the alternative hypotheses in the form:

H_A: $p <$ hypothesized value 
H_A: $p >$ hypothesized value 
H_A: $p \neq$ hypothesized value 

hypothesized value = p_0

- Contains values of the parameter that we consider plausible if we reject the null hypothesis

P-value: The probability, computed assuming H_0 is true, that the statistic ($\hat{p}$ or $\bar{x}$) would take a value as extreme as or more extreme than the one actually observed, in the direction specified by H_A.

- Probability of an event
- We want to find the probability of seeing data like these given that the null hypothesis is true

HIGH P-value

- we haven't seen anything unlikely or surprising at all
- mostly consistent with the model from the null hypothesis and therefore no reason to reject the null
- you say that you "fail to reject the null" because we can't prove that it is true we just say it doesn't appear to be false

LOW P-value

- we say that its very unlikely we'd observe data like these if our null hypothesis were true and therefore we "reject the null"

Example 1:
Write the null hypotheses and alternate hypotheses you would usé to test each of the following situations.

a. A governor is concerned about his "negatives"-the percentage of state residents who express disapproval of his job performance. His political committee pays for a series of TV ads, hoping that they can keep the negatives below 30%. They will use follow-up polling to assess the ads' effectiveness.

$H_0: p = 0.30$

$H_A: p < 0.30$

b. Is a coin fair?

$H_0: p = 0.50$ $H_A: p \neq 0.50$

c. Only about 20% of people who try to quit smoking succeed. Sellers of a motivational tape claim that listening to the recorded messages can help people quit.

$H_0: p = 0.20$ $H_A: p > 0.20$

Significance Level: the fixed value α that we use as a cutoff for deciding whether an observed result is too unlikely to happen by chance alone when the null hypothesis is true.

- Denoted by α
- Most commonly used is $\alpha = 0.05$
- Also used is $\alpha = 0.10$ and $\alpha = 0.01$

Options for Statistical Significance:

- **Reject H_0:** If the observed result is too unlikely to occur just by chance when the null hypothesis is true, we can reject H_0 and say there is convincing evidence for H_A.

- **Fail to Reject H_0:** If the observed result is not very unlikely to occur when the null hypothesis is true, we should fail to reject H_0 and say that we do not have convincing evidence for H_A.

Significance Tests: A Four-Step Process

1. **State:** What hypotheses do you want to test (H_o and H_A), and at what significance level? Define any parameters you use.

2. **Plan:** Choose the appropriate inference method. Check conditions.
   - Random:
   - 10 %
   - Large Counts ($np \geq 10$ and $nq \geq 10$)

   If the conditions are met, we will perform a **one-sample z-test for** the population proportion p.

3. **Do:** If the conditions are met, perform calculations:
   - Compute the **test statistic**
   - Find the **p-value**

   Draw a picture! use probability notation when appropriate

4. **Conclude:** Make a decision about the hypotheses in the context of the problem.

   SENTENCE FRAMES:

   Because our P-value = ____ is greater than the significance level α = ___, we fail to reject H_0. There is not convincing evidence that (alternative hypothesis).

   Because our P-value = ____ is less than the significance level α = ___, we reject H_0. There is convincing evidence that (alternative hypothesis).

   **Note:* *We NEVER ACCEPT the null. Instead, we reject or fail to reject.*

Test Statistic: a calculation that measures how far a sample statistic diverges from what we would expect if the null hypothesis H_0 were true, in standardized units.

$$\text{test statistic} = \frac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$$

One-Sample Z-Test for a proportion

$$\boxed{z = \frac{(\hat{p} - p_o)}{\sqrt{\frac{p_o(1-p_o)}{n}}}}$$

In your calculator:

1. STAT
2. TESTS
3. Option 5: 1-PropZTest
   - p_0 =
   - x =
   - n =
   - prop: $\neq p_0$ $< p_0$ $> p_0$
4. CALCULATE

Example 2:

A large city's Department of Motor Vehicles claimed that 80% of candidates pass driving tests but a newspaper reporter's survey of 90 randomly selected local teens that had taken the test found only 61 who passed.

a. Does this finding suggest that the passing rate for teenagers is lower than the DMV reported? Write appropriate hypothesis.

State:

H_0: $p = 0.80$ where p = the passing rate of teens who take DMV driving tests

H_A: $p < 0.80$

$\hat{p} = \frac{61}{90} = 0.678$ $\alpha =$

b. Are the conditions for inference satisfied?

Random: randomly selected ✓

plan:

10% condition: 900 < all teens who take the driving test

Large counts: $np \geq 10$ $nq \geq 10$

$90 \cdot 0.8 \geq 10$ $90 \cdot 0.2 \geq 10$

$72 \geq 10$ ✓ $18 \geq 10$ ✓

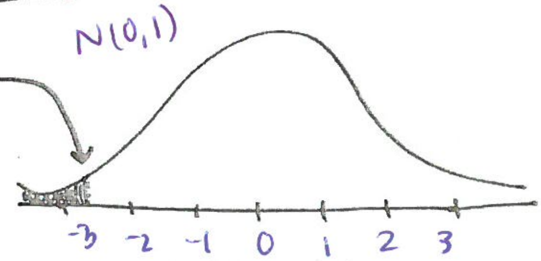
because our conditions are met, we will perform a 1-sample z-test for the population proportion p.

c. What's the P-value for the one-sample z-test for a proportion?

$P(\hat{p} < 0.678) = P(z < -2.89) = \boxed{0.00191}$

Do:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.678 - 0.80}{\sqrt{\frac{(0.8)(0.2)}{90}}} = -2.89$$



d. What can the reporter conclude? And how might the reporter explain what the P-value means for the newspaper story knowing that the significance level is $\alpha = 0.05$?

conclude:

Because the p-value 0.00191 is less than our significance level $\alpha = 0.05$, we reject the null hypothesis. There is convincing evidence that the passing rate of teens who take the DMV driving test is lower than 80%.

Example 3:

Anyone who plays or watches sports has her of the "home field advantage." Teams tend to win more often when they play at home. Or do they? If there were no home field advantage, the home teams would win about half of all games played. In the 2007 MLB season, there were 2431 regular-season games. It turns out that the home team won 1319 of the 2431 games or 54.25% of the time. Could this deviation from 50% be explained just from natural sampling variability, or is it evidence to suggest that there are really is a home field advantage, or least in professional baseball? Given that the significance level is $\alpha = 0.05$, determine if home field advantage really exists.

State: $H_0: p = 0.50$ where p = the true proportion of games won by a home team.

$H_A: p > 0.5$

$\alpha = 0.05$ $\hat{p} = 0.5425$

Plan: Random: all of the games in the season ✓

10% condition: 24310 < all MLB games ever played ✓

Large Counts: $np \geq 10$ $nq \geq 10$

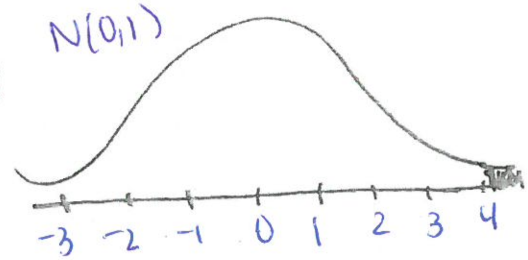
$(0.5)(2431) \geq 10$ $(0.5)(2431) \geq 10$

$1215 \geq 10$ ✓ $1215 \geq 10$ ✓

because our conditions are met, we will perform a 1-sample z-test for the population proportion p.

DO: $P(\hat{p} > 0.5) = P(z > 4.191) = \boxed{0.0000139}$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.5425 - 0.50}{\sqrt{\frac{(0.5)(0.5)}{2431}}} = 4.191$$



or 1-prop z Test

p_0: 0.5

X: 1319

n: 2431

prop: $> p_0$

$z = 4.198$

p-value = $\boxed{0.0000135}$

Conclude: Because our p-value 0.0000139 is less than our significance level $\alpha = 0.05$, we reject the null. There is convincing evidence that more than 50% of games are won by the home team.

Example 4:

According to the Centers for Disease Control and Prevention (CDC) Website, 50% of high school students have never smoked a cigarette. Bob wonders whether this national result holds true in his large, urban high school. For his AP statistics class project, Bob surveys an SRS of 150 students from his school. He gets responses from all 150 students, and 90 say that they have never smoked a cigarette.

a. What should Bob conclude? Give appropriate evidence to support your answer with a significance level of $\alpha = 0.05$.

State: $H_0: p = 0.50$ where p = the true proportion of students at a large urban HS who have never smoked a cigarette.

$H_A: p \neq 0.50$

$\alpha = 0.05$

$\hat{p} = \frac{90}{150} = 0.60$

Plan: Random: SRS ✓

10% condition: 1500 < all students at the large urban HS ✓

Large counts: $np \geq 10$ $\quad nq \geq 10$

$(150)(0.5) \geq 10$ $\quad 150(0.5) \geq 10$

$75 \geq 10$ ✓ $\quad 75 \geq 10$ ✓

because our conditions are met, we will perform a 1-sample z-test for the population proportion p.

DO: 1-PropZTest

p_0: 0.50

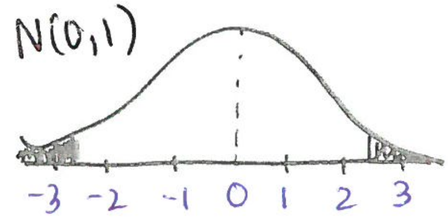
x: 90

n: 150

prop: $\neq p_0$

$z = 2.449$

p-value: 0.0143



Conclude: Because our p-value 0.0143 is less than our significance level $\alpha = 0.05$, we reject the null. There is convincing evidence that the

true prop. of students at the large urban HS who have tried a cigarette differs from the 50% stated by the CDC website.

b. Find a 95% confidence interval that will capture the true proportion of students in high school that have never smoked a cigarette before.

State: We want to find the true proportion of students in HS who have never smoked a cigarette with 95% confidence.

$\hat{p} = 0.60$

Plan: conditions met above ✓ we will do a 1-sample z-interval to estimate p.

DO: $0.6 \pm 1.96 \cdot \sqrt{\frac{(0.6)(0.4)}{150}} = (0.5216, 0.6784)$

Conclude: We are 95% confident that the interval from 0.5216 to 0.6784 captures the true proportion of students at large urban HS who have never smoked a cigarette.