AP Statistics
Unit 07 – Day 04 Notes
Significance Tests: 2-Sample Proportions

Name Key
Period______

STATE:

- p_1 =
- p_2 =
- $\widehat{p_1}$ =
- $\widehat{p_2}$ =

H_o: $p_1 - p_2 = 0$ OR H_o: $p_1 = p_2$

H_A: $p_1 - p_2 \neq 0$ OR H_A: $p_1 \neq p_2$

H_A: $p_1 - p_2 > 0$ OR H_A: $p_1 > p_2$

H_A: $p_1 - p_2 < 0$ OR H_A: $p_1 < p_2$

one of these

State the significance level: α =

PLAN:

Random: The data must come from two independent random samples or from two groups in a randomized experiment.

10% Condition: $n_1 < 10\%$ of the population
$n_2 < 10\%$ of the population

Large Counts: The distribution of $\widehat{p_1} - \widehat{p_2}$ is approximately Normal if…

$n_1\widehat{p_1} \geq 10$
$n_1\widehat{q_1} \geq 10$
$n_2\widehat{p_2} \geq 10$
$n_2\widehat{q_2} \geq 10$

State the test you are using: We will be using a **2-sample z-test for the difference between two proportions $p_1 - p_2$**

DO:

$$\boxed{z = \frac{(\widehat{p_1} - \widehat{p_2}) - 0}{\sqrt{\frac{\widehat{p_C}(\widehat{q_C})}{n_1} + \frac{\widehat{p_C}(\widehat{q_C})}{n_2}}}}$$

* we use the calculator :)

CONCLUDE:

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

A Test for the Differences Between Two Proportions Pooled Samples:

- Combining the counts to get an overall proportion.
- Whenever we have data from different sources or different groups but we believe that they really came from the same underlying population, we pool them to get better estimates.
- *We always use pooled values when doing significance tests, but not for confidence intervals.*

$\hat{p}_c =$ → $$\boxed{\hat{p}_{pooled} = \frac{success_1 + success_2}{n_1 + n_2}}$$

"combined"

$$\boxed{z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_1} + \frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_2}}}}$$

know this for mc Questions

CALCULATOR:

STAT → TESTS → OPTION 6: 2-PropZTest

$x_1 =$ # of successes $(n_1 \cdot \hat{p}_1)$
$n_1 =$
$x_2 =$ # of successes $(n_2 \cdot \hat{p}_2)$
$n_2 =$
$p1 \neq p2$ $< p2$ $> p2$
Calculate Draw

Example 1:

One concern of a study on teen's online profiles was safety and privacy. In a random sample, only 19% (62 girls) of 325 teen girls with profiles say that they are easy to find, while 28% (75 boys) of the 268 boys with profiles say the same. Are these results significant enough to show that girls were less likely than boys to say that they are easy to find online from their profiles.

State: p_1 = the true proportion of teen girls who say they are easy to find online.

p_2 = the true proportion of teen boys who say they are easy to find online.

$\hat{p}_1 = \frac{62}{325} = 0.19$ $\quad \hat{p}_2 = \frac{75}{268} = 0.28$ $\quad \alpha = 0.05$

$H_0: p_1 = p_2$

$H_A: p_1 < p_2$

Plan: random: data comes from 2 independent random samples ✓

10% condition: $n_1 = 325$ $\quad$ 3250 < all teen girls with online profiles. ✓

$n_2 = 268$ $\quad$ 2680 < all teen boys with online profiles. ✓

Large counts: $n_1\hat{p}_1 = 62 \geq 10$ ✓ $\quad n_2\hat{p}_2 = 75 \geq 10$ ✓

$n_1\hat{q}_1 = 263 \geq 10$ ✓ $\quad n_2\hat{q}_2 = 193 \geq 10$ ✓

Because our conditions are met, we will perform a 2-sample z-test for difference of two proportions $p_1 - p_2$.

Do: 2-PropZTest

x_1: 62

n_1: 325

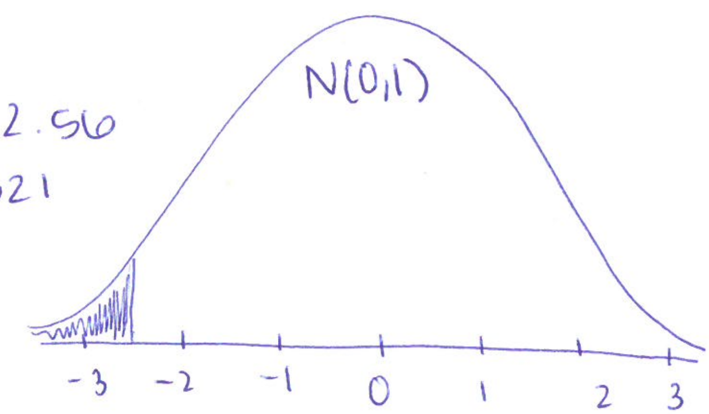
x_2: 75

n_2: 268

$p_1 < p_2$

test statistic: -2.56

p-value: 0.00521



Conclude: Because our p-value = 0.00521 is less than our significance level $\alpha = 0.05$, we reject the null hypothesis. There is convincing evidence that the true proportion of boys who say they are easy to find online is greater than the true proportion of girls who say they are easy to find online.

Example 2:

Of a random sample of 995 people, 37% of adults reported that they snored at least a few nights a week during the past year. Would you expect that percentage to be the same for all age groups? Split into two age categories, 26% of the 184 people under 30 snored, compared with 39% of the 811 in people older than 30. Determine the significance of snoring rates of the two age groups. ~~In the DO sections, please do it with pooling *and* without pooling so we can compare the values.~~

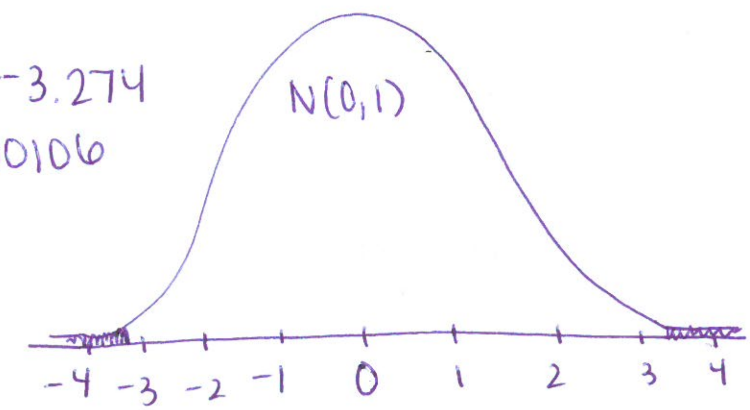
State: p_1 = the true proportion of adults under 30 who snore
p_2 = the true proportion of adults over 30 who snore

$H_0: p_1 = p_2$ $\qquad \hat{p}_1 = 0.26$ $\qquad \alpha = 0.05$
$H_A: p_1 \neq p_2$ $\qquad \hat{p}_2 = 0.39$

Plan: random: the data come from 2 independent groups in a random sample ✓
10% condition: $n_1 = 184$ $\quad 1840 <$ all adults under 30 ✓
$n_2 = 811$ $\quad 8110 <$ all adults over 30 ✓
Large Counts: $n_1\hat{p}_1 = 48 \geq 10$ ✓ $\quad n_2\hat{p}_2 = 316 \geq 10$ ✓
$n_1\hat{q}_1 = 136 \geq 10$ ✓ $\quad n_2\hat{q}_2 = 495 \geq 10$ ✓
Because our conditions are met, we will perform a 2-sample z-test for the difference of 2 proportions $p_1 - p_2$.

DO: 2-prop z Test
X_1: 48
n_1: 184
X_2: 316
n_2: 811
$p_1 \neq p_2$

test statistic: -3.274
p-value: 0.00106



Conclude: Because our p-value = 0.00106 is less than our significance level $\alpha = 0.05$, we reject the null hypothesis. There is convincing evidence that the true proportion of people under 30 who snore differs from the true proportion of people over 30 who snore.