11.1 Inference about Two Population Proportions MATH 241 THOMPSON

By definition, a sampling method is independent when an individual selected for one sample does not dictate which individual is to be in the second sample.

Conversely, a sampling method is dependent when an individual selected for one sample does dictate which individual is to be in the second sample.

1) Determine whether the following sampling is dependent or independent. Indicate whether the response variable is qualitative or quantitative.

A researcher wishes to compare annual salaries of married mathematicians and their spouses. She obtains a random sample of 967 such couples who work and determines each spouse's annual salary.

Determine whether the following sampling is dependent or independent.

✔ D. The sampling is dependent because an individual selected for one sample does dictate which individual is to be in the second sample.

✔ D. The variable is quantitative because it is a numerical measure.

A sampling method is independent when an individual selected for one sample does not dictate which individual is to be in the second sample. A sampling method is dependent when an individual selected for one sample does dictate which individual is to be in the second sample. Dependent samples are often referred to as matched-pairs samples.

Qualitative variables allow for classification of individuals based on some attribute or characteristic. Quantitative variables provide numerical measures of individuals, and can be added or subtracted.

2) An educator wants to determine whether a new curriculum significantly improves standardized test scores for third grade students. She randomly divides 80 third-graders into two groups. Group 1 is taught using the new curriculum, whi group 2 is taught using the traditional curriculum. At the end of the school year, both groups are given the standardized test and the mean scores are compared. Determine whether the sampling is dependent or independent. Indicate whether the response variable is qualitative or quantitative.

Determine whether the sampling is dependent or independent.

A. The sampling is independent because the individuals selected to be in one sample are used to determine the individuals to be in the second sample.

B. The sampling is dependent because the individuals selected for one sample do not dictate which individuals are to be in a second sample.

C. The sampling is dependent because the individuals selected to be in one sample are used to determine the individua to be in the second sample.

✔ D. This sampling is independent because the individuals selected for one sample do not dictate which individuals are to be in a second sample.

Indicate whether the response variable is qualitative or quantitative.

A. The variable is qualitative because it is a numerical measure.

B. The variable is quantitative because it is an attribute classification.

C. The variable is qualitative because it is an attribute classification.

✔ D. The variable is quantitative because it is a numerical measure.

3)

Researchers wondered if there was a difference between males and females in regard to some common annoyances. They asked a random sample of males and females, the following question: "Are you annoyed by people who repeatedly check their mobile phones while having an in-person conversation?" Among the 582 males surveyed, 189 responded "Yes"; among the 583 females surveyed, 202 responded "Yes." Does the evidence suggest a higher proportion of females are annoyed by this behavior? Complete parts (a) through (g) below.

(a) Determine the sample proportion for each sample.

The proportions of the females and males who took the survey who are annoyed by the behavior in question are .3465 and .3247, respectively.
(Round to four decimal places as needed.)

189/582 = 0.3247 202/583 = 0.3465

smaller one(larger decimal) on the right

(b) Explain why this study can be analyzed using the methods for conducting a hypothesis test regarding two independent proportions. Select all that apply.

- [x] A. $n_1\hat{p}_1\left(1-\hat{p}_1\right) \geq 10$ and $n_2\hat{p}_2\left(1-\hat{p}_2\right) \geq 10$
- [ ] B. The samples are dependent.
- [ ] C. The sample size is more than 5% of the population size for each sample.
- [ ] D. The data come from a population that is normally distributed.
- [x] E. The sample size is less than 5% of the population size for each sample.
- [x] F. The samples are independent.

(c) What are the null and alternative hypotheses? Let p_1 represent the population proportion of females who are annoyed by the behavior in question and p_2 represent the population proportion of males who are annoyed by the behavior in question.

H_0: p_1 = p_2
H_1: p_1 > p_2

STATS- PROPORTION STATS – TWO SAMPLE -WITH SUMMARY

.3465 - .3247 = 0.0218

Hypothesis test results:

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	Z-Stat	P-value
p_1 - p_2	202	583	189	582	0.021741437	0.027669425	-0.0021165264	0.5008

Two Sample Prop. Summary

Sample 1:
# of successes: 202
# of observations: 583

Sample 2:
# of successes: 189
# of observations: 582

Perform:
Hypothesis test for p_1 - p_2
H_0: p_1 - p_2 = .0218
H_A: p_1 - p_2 > .0218

(d) Describe the sampling distribution of $p_{female} - p_{male}$. Draw a normal model with the area representin hypothesis test.

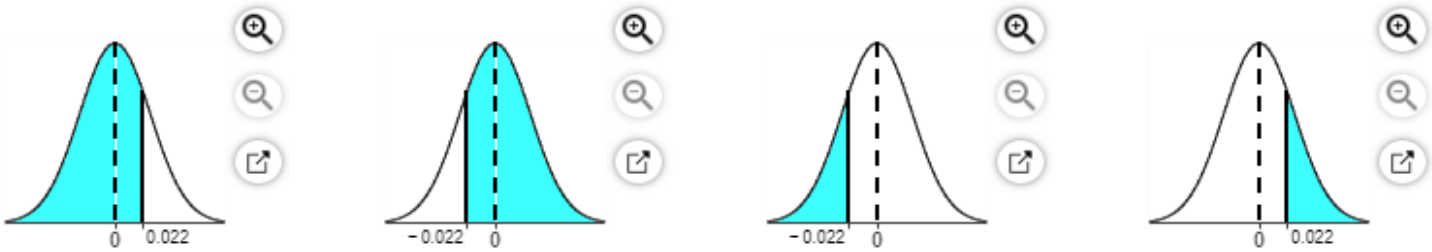
mean is always 0

The sampling distribution is approximately normal with mean 0 and standard deviation 0.0277.
(Type an integer or decimal rounded to four decimal places as needed.)

greater than from problem

Draw a normal model with the area representing the P-value shaded for this hypothesis test. Choose the correct graph below, where the horizontal axis represents $\hat{p}_{female} - \hat{p}_{male}$.

A. B. C. D.



STATS- PROPORTION STATS – TWO SAMPLE -WITH SUMMARY

Hypothesis test results:

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	Z-Stat	P-value
$p_1 - p_2$	202	583	189	582	0.021741437	0.027669425	0.78575674	0.216

(e) Determine the P-value based on the model from part (d).

First find the test statistic for this hypothesis test.

0.79 (Round to two decimal places as needed.)

Two Sample Prop. Summary

Sample 1:
# of successes: 202
# of observations: 583

Sample 2:
# of successes: 189
# of observations: 582

C

Perform:
Hypothesis test for $p_1 - p_2$
H_0: $p_1 - p_2$ = 0
H_A: $p_1 - p_2$ < 0

Determine the P-value for this hypothesis test.

0.216 (Round to three decimal places as needed.)

(f) Interpret the P-value.

If the population proportions are equal, one would expect a sample difference proportion greater than the one observed in about 216 out of 1000 repetitions of this experiment.
(Round to the nearest integer as needed.)

(g) Based on the P-value, what does the sample evidence suggest? That is, what is the conclusion of the hypothesis test? Assume an $\alpha = 0.01$ level of significance.

Do not reject H_0. There is not sufficient evidence at the $\alpha = 0.01$ level of significance to suggest a higher proportion of females are annoyed by the behavior in question.

4) In 1945, an organization surveyed 1100 adults and asked, "Are you a total abstainer from, or do you on occasion consume, alcoholic beverages?" Of the 1100 adults surveyed, 429 indicated that they were total abstainers. In a recent survey, the same question was asked of 1100 adults and 363 indicated that they were total abstainers. Complete parts (a) and (b) below.

(a) Determine the sample proportion for each sample.

The proportions of the adults who took the 1945 survey and the recent survey who were total abstainers are .39 and .33, respectively.
(Round to three decimal places as needed.)

429/1100 = 0.39 363/1100 = 0.33

(b) Has the proportion of adults who totally abstain from alcohol changed? Use the $\alpha = 0.10$ level of significance.

STATS- PROPORTION STATS – TWO SAMPLE -WITH SUMMARY

First verify the model requirements. Select all that apply.

- [ ] A. The samples are dependent.
- [ ] B. The data come from a population that is normally distributed.
- [x] C. The samples are independent.
- [x] D. The sample size is less than 5% of the population size for each sample.
- [ ] E. The sample size is more than 5% of the population size for each sample.
- [x] F. $n_1\hat{p}_1\left(1-\hat{p}_1\right) \geq 10$ and $n_2\hat{p}_2\left(1-\hat{p}_2\right) \geq 10$

Two Sample Prop. Summary

Sample 1:
# of successes: 429
# of observations: 1100

Sample 2:
# of successes: 363
# of observations: 1100

Perform:
Hypothesis test for $p_1 - p_2$
H_0: $p_1 - p_2$ = 0
H_A: $p_1 - p_2$ ≠ 0

Identify the null and alternative hypotheses for this test. Let p_1 represent the population proportion of 1945 adults who were total abstainers and p_2 represent the population proportion of recent

Determine the null and alternative hypotheses.

$H_0: p_1 = p_2$

$H_1: p_1 \neq p_2$

Find the test statistic for this hypothesis test.

2.93 (Round to two decimal places as needed.)

Determine the P-value for this hypothesis test.

Change inequality direction for the P-value

.003 (Round to three decimal places as needed.)

Interpret the P-value.

If the population proportions are equal, one would expect a sample difference proportion greater than the absolute value of the one observed in about 0 out of 100 repetitions of this experiment.

If P-value < α, reject the null hypothesis.

D. Reject H_0. There is sufficient evidence at the $\alpha = 0.10$ level of significance to suggest the proportion of adults who totally abstain from alcohol has changed.

5) Conduct a test at the $\alpha = 0.01$ level of significance by determining **(a)** the null and alternative hypotheses, **(b)** the test statistic, an **(c)** the P-value. Assume the samples were obtained independently from a large population using simple random sampling. Test whether $p_1 > p_2$. The sample data are $x_1 = 116$, $n_1 = 256$, $x_2 = 131$, and $n_2 = 301$.

(a) Choose the correct null and alternative hypotheses below.

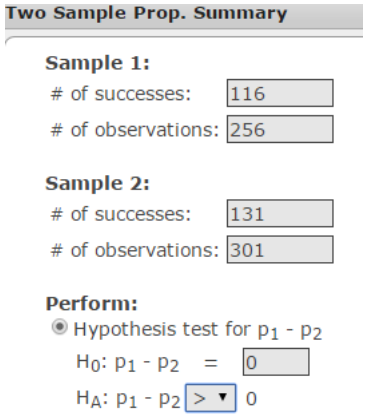
A. $H_0: p_1 = 0$ versus $H_1: p_1 \neq 0$

B. $H_0: p_1 = p_2$ versus $H_1: p_1 \neq p_2$

C. $H_0: p_1 = p_2$ versus $H_1: p_1 < p_2$

D. $H_0: p_1 = p_2$ versus $H_1: p_1 > p_2$

STATS- PROPORTION STATS – TWO SAMPLE - WITH SUMMARY



(b) Determine the test statistic.

$z_0 =$ .42 (Round to two decimal places as needed.)

(c) Determine the P-value.

Change inequality direction for the P-value

The P-value is .336.

(Round to three decimal places as needed.)

What is the result of this hypothesis test?

A. Do not reject the null hypothesis because there is not sufficient evidence to conclude that $p_1 > p_2$.

P-value > α, do not reject the null hypothesis

6) In randomized, double-blind clinical trials of a new vaccine, children were randomly divided into two groups. Subjects in group 1 received the new vaccine while subjects in group 2 received a control vaccine. After the second dose, 120 of 673 subjects in the experimental group (group 1) experienced drowsiness as a side effect. After the second dose, 77 of 554 of the subjects in the control group (group 2) experienced drowsiness as a side effect. Does the evidence suggest that a higher proportion of subjects in group 1 experienced drowsiness as a side effect than subjects in group 2 at the $\alpha = 0.10$ level of significance?

- [x] B. $n_1\hat{p}_1\left(1-\hat{p}_1\right) \geq 10$ and $n_2\hat{p}_2\left(1-\hat{p}_2\right) \geq 10$
- [ ] C. The samples are dependent.
- [ ] D. The data come from a population that is normally distributed.
- [x] E. The samples are independent.
- [x] F. The sample size is less than 5% of the population size for each sample.

Determine the null and alternative hypotheses.

H_0: p_1 = p_2

H_1: p_1 > p_2

Find the test statistic for this hypothesis test.

1.87 (Round to two decimal places as needed.)

Determine the P-value for this hypothesis test.

.031 (Round to three decimal places as needed.)

Interpret the P-value.

If the population proportions are equal, one would expect a sample difference proportion greater than the one observed in about 31 out of 1000 repetitions of this experiment.

- [x] C. Reject H_0. There is sufficient evidence to conclude that a higher proportion of subjects in group 1 experienced drowsiness as a side effect than subjects in group 2 at the $\alpha = 0.10$ level of significance.

7) Conduct the following test at the $\alpha = 0.05$ level of significance by determining **(a)** the null and alternative hypothese **(b)** the test statistic, and **(c)** the P-value. Assume that the samples were obtained independently using simple random sampling.

Test whether $p_1 \neq p_2$. Sample data are $x_1 = 28$, $n_1 = 255$, $x_2 = 38$, and $n_2 = 302$.

(a) Determine the null and alternative hypotheses. Choose the correct answer below.

- [ ] A. H_0: $p_1 = 0$ versus H_1: $p_1 = 0$
- [x] B. H_0: $p_1 = p_2$ versus H_1: $p_1 \neq p_2$

(b) The test statistic z_0 is −.58. (Round to two decimal places as needed.) /ersus H_1: $p_1 > p_2$

(c) The P-value is .560. (Round to three decimal places as needed.)

Test the null hypothesis. Choose the correct conclusion below.

- [x] C. Do not reject the null hypothesis because there is not sufficient evidence to conclude that $p_1 \neq p_2$.

8) A survey asked, "How many tattoos do you currently have on your body?" Of the 1232 males surveyed, 181 responded that they had at least one tattoo. Of the 1060 females surveyed, 146 responded that they had at least one tattoo. Construct a 90% confidence interval to judge whether the proportion of males that have at least one tattoo differs significantly from the proportion of females that have at least one tattoo. Interpret the interval.

Let p_1 represent the proportion of males with tattoos and p_2 represent the proportion of females with tattoos. Find the 90% confidence interval for $p_1 - p_2$.

The lower bound is $-.015$.
The upper bound is $.033$.
(Round to three decimal places as needed.)

B. There is 90% confidence that the difference of the proportions is in the interval. Conclude that there is insufficient evidence of a significant difference in the proportion of males and females that have at least one tattoo.

lower and upper bound – confidence interval

Sample 1:
# of successes: 181
# of observations: 1232

Sample 2:
# of successes: 146
# of observations: 1060

Perform:
○ Hypothesis test for $p_1 - p_2$
H_0: $p_1 - p_2$ = 0
H_A: $p_1 - p_2$ ≠ 0
◉ Confidence interval for $p_1 - p_2$
Level: 0.9

sufficient?? – Hypothesis test

Sample 1:
# of successes: 181
# of observations: 1232

Sample 2:
# of successes: 146
# of observations: 1060

Perform:
◉ Hypothesis test for $p_1 - p_2$
H_0: $p_1 - p_2$ = 0
H_A: $p_1 - p_2$ ≠ 0

9) Construct a confidence interval for $p_1 - p_2$ at the given level of confidence.

$x_1 = 357$, $n_1 = 549$, $x_2 = 425$, $n_2 = 557$, 90% confidence

The researchers are 90 % confident the difference between the two population proportions, $p_1 - p_2$, is between $-.157$ and $-.068$.
(Use ascending order. Type an integer or decimal rounded to three decimal places as needed.)

10) A physical therapist wants to determine the difference in the proportion of men and women who participate in regular sustained physical activity. What sample size should be obtained if he wishes the estimate to be within three percentage points with 95% confidence, assuming that
(a) he uses the estimates of 21.9% male and 18.1% female from a previous year?
(b) he does not use any prior estimates?

(a) n = 1363 (Round up to the nearest whole number.)

(b) n = 2135 (Round up to the nearest whole number.)

STAT-PROPORTION STATS-TWO SAMPLE
POWER/SAMPLE SIZE –confidence interval

from three % points
Width=0.03x2=0.06

Hypothesis Test Power | Confidence Interval Width

Press Compute to update.

No estimates – target = 0.5, make sure to reset the width to 0.06 again

Required parameters:		Enter one:	
Confidence level:	0.95	Width:	0.06
First proportion:	0.5	Sample size per group:	2135
Second proportion	0.5		

Compute

Required parameters:		Enter one:	
Confidence level:	0.95	Width:	.06
First proportion:	0.219	Sample size per group:	1363
Second proportion	0.181		

Compute

The experimental design is based on how the subjects are grouped by similarity. Different types of experimental designs and studies are explained below.

A completely randomized design is one in which each experimental unit is randomly assigned to a treatment.

A matched-pairs design is an experimental design in which the experimental units are paired up. The pairs are selected so that they are related in some way (that is, the same person before and after a treatment, twins, husband and wife, same geographical location, and so on).

A randomized block design is used when the experimental units are divided into homogeneous groups called blocks. Within each block, the experimental units are randomly assigned to treatments.

Case-control studies are retrospective studies, meaning that they require individuals to look back in time or require the researcher to look at existing records. In case-control studies, individuals who have a certain characteristic may be matched with those who do not.

11) A doctor released the results of clinical trials for a vaccine to prevent a particular disease. In these clinical trials, 400,000 children were randomly divided in two groups. The subjects in group 1 (the experimental group) were give the vaccine, while the subjects in group 2 (the control group) were given a placebo. Of the 200,000 children in the experimental group, 43 developed the disease. Of the 200,000 children in the control group, 70 developed the disease. Complete parts (a) through (f) below.

(a) What type of experimental design is this?

- ○ Case-control
- ○ Matched-pairs design
- ✔ Completely randomized design
- ○ Randomized block design

(b) What is the response variable?

- ✔ A. Whether the subject contracted the disease or not
- ○ B. Whether the subject was in group 1 or group 2
- ○ C. The amount of the vaccine received
- ○ D. Whether the vaccine prevented the disease or not

(c) What are the treatments?

- ○ A. The amount of the placebo received
- ○ B. The amount of the vaccine received
- ○ C. Group 1 or group 2
- ✔ D. The vaccine or placebo

The number of subjects is so large because there is a low incidence rate of the disease.

(f) Does it appear to be the case that the vaccine was effective? Use the $\alpha = 0.05$ level of significance.

First verify the model requirements. Select all that apply.

- ☐ A. The sample size is more than 5% of the population size for each sample.
- ☐ B. The samples are dependent.
- ✔ C. $n_1\hat{p}_1\left(1-\hat{p}_1\right) \geq 10$ and $n_2\hat{p}_2\left(1-\hat{p}_2\right) \geq 10$
- ☐ D. The data come from a population that is normally distributed.
- ✔ E. The samples are independent.
- ✔ F. The sample size is less than 5% of the population size for each sample.

Identify the null and alternative hypotheses for this test. Let p_1 represent the population proportion of children given the vaccine who developed the disease and p_2 represent the population proportion of children given a placebo who developed the disease.

$H_0: p_1 = p_2$

$H_1: p_1 < p_2$

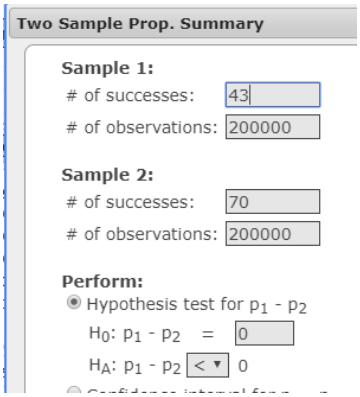
Find the test statistic for this hypothesis test.

-2.54 (Round to two decimal places as needed.)

Determine the P-value for this hypothesis test.

.006 (Round to three decimal places as needed.)

Interpret the P-value.



If the population proportions are equal, one would expect a sample difference proportion smaller than the one observed in about 1 out of 100 repetitions of this experiment.
(Round to the nearest integer as needed.)

If P-value < α, reject the null hypothesis.

12) State the conclusion for this hypothesis test.

Reject H_0. There is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude that the vaccine was effective.

Explain the difference between an independent and dependent sample.

Choose the correct answer below.

A sample is independent when an individual selected for one sample does not dictate which individual is to be in the second sample. A sample is dependent when an individual selected for one sample dictates which individual is to be in the second sample.

13)

In clinical trials of a medication, 2115 subjects were divided into two groups. The 1535 subjects in group 1 received the medication. The 580 in group 2 received a placebo. Of the 1535 subjects in group 1, 59 experienced dizziness as a side effect. In group 2, 18 experienced dizziness as a side effect. To test whether the proportion experiencing dizziness in group 1 is greater than that in group 2, the researchers entered the data into statistical software and obtained the following results. Test at $\alpha = 0.05$.

Sample	X	N	Sample p
1	59	1535	0.038436
2	18	580	0.031034

Estimate for p(1) − p(2): 0.007402
95% CI for p(1) − p(2): (− 0.009676, 0.02448)
Test for p(1) − p(2) = 0 (vs > 0): z = 0.81 P-value = 0.209

(This is a reading assessment question. Be certain of your answer because you only get one attempt on this question.)

What conclusion can be drawn at the $\alpha = 0.05$ level of significance?

Do not reject H_0, there is not enough evidence to conclude that the proportion experiencing dizziness in group 1 is greater than the proportion experiencing dizziness in group 2.

14) A sampling method is independent when an individual selected for one sample does not dictate which individual is to be in the second sample.

15) A sampling method is dependent when the individuals selected for one sample are used to determine the individuals in the second sample.

EXTRA EXAMPLES

In a survey of 3039 adults, a poll asked people whether they smoked cigarettes and whether they always wear a seat belt in a car. The table shows the results of the survey. For each activity, define a success as finding an individual that participates in the hazardous activity. Complete parts (a) and (b).

	No Seat Belt (success)	Seat Belt (failure)
Smoke (success)	57	443
Do not smoke (failure)	346	2193

(a) Why is this a dependent sample?

- A. More than 5% of the population was surveyed.
- B. The same person answered both questions. ✔
- C. Two questions where asked.

(b) Is there a significant difference in the proportion of individuals who smoke and the proportion of individuals that do not wear a seat belt? In other words, is there a significant difference between the proportion of individuals who engage in hazardous activities? Use the $\alpha = 0.10$ level of significance. Let p_1 represent the proportion of individuals who smoke and p_2 represent the proportion of individuals that do not wear a seat belt.

What are the hypotheses for this test?

- A. $H_0: p_1 < p_2$ $H_1: p_1 = p_2$
- B. $H_0: p_1 \neq p_2$ $H_1: p_1 = p_2$
- C. $H_0: p_1 = p_2$ $H_1: p_1 < p_2$
- D. $H_0: p_1 = p_2$ $H_1: p_1 \neq p_2$ ✔

Use proportion two sample summary to get p then divide the answer by 2

Most p-values are .000 if this formula does not work

Calculate the test statistic.

$z_0 =$ 3.42 (Round to two decimal places as needed.)

Calculate the P-value.

P-value = .001 (Round to three decimal places as needed.)

Which of the following is the correct conclusion for the hypothesis test?

Reject H_0. There is sufficient evidence at the $\alpha = 0.10$ level of significance to conclude that there is a difference in the proportion who do not use a seat belt and the proportion who smoke.

For two dependent samples, computer the test statistic using the McNemar's formula below, where f_{12} and f_{21} are values in the table.

$$z_0 = \frac{|f_{12} - f_{21}| - 1}{\sqrt{f_{12} + f_{21}}}$$

		Treatment A	
		Success	Failure
Treatment B	Success	f_{11}	f_{12}
	Failure	f_{21}	f_{22}

Find the values of f_{12} and f_{21}.

$f_{12} =$ 443

$f_{21} =$ 346

If P-value < α, reject the null hypothesis.

Researchers developed a new method of voice recognition (called a remapped network) that was thought to be an improvement over an existing neural network. The data shown in the accompanying table are based on results of their research. Does the evidence suggest that the remapped network has a different proportion of errors than the neural network? Use the $\alpha = 0.05$ level of significance.

Click the icon to view the data table.

Let p_1 represent the proportion of errors for the neural network and p_2 represent the proportion of errors for the remapped network. Determine the null and alternative hypotheses. Choose the correct answer below.

- A. $H_0: p_1 = p_2$ versus $H_1: p_1 < p_2$
- B. $H_0: p_1 = p_2$ versus $H_1: p_1 > p_2$
- C. $H_0: p_1 = 0$ versus $H_1: p_1 = 0$
- D. $H_0: p_1 = p_2$ versus $H_1: p_1 \neq p_2$

The test statistic z_0 is 2.41. (Round to two decimal places as needed.)

The P-value is .015. (Round to three decimal places as needed.)

Test the null hypothesis. Choose the correct conclusion below.

P-value ---two sample proportion

- A. Reject the null hypothesis because there is sufficient evidence to conclude that the remapped network has a different proportion of errors than the neural network.

Sample 1 errors of neutral network 445+27=472

Sample 2 errors 375+27=402

$$\frac{|375-445|-1}{\sqrt{375+445}} = 2.41$$

make sure |x| is positive answer

		Remapped network	
		Recognized word (success)	Did not recognize word (failure)
Neural network	Recognized word (success)	8,573	375
	Did not recognize word (failure)	445	27

Two Sample Prop. Summary

Sample 1:
# of successes: 472
# of observations: 9420

Sample 2:
# of successes: 402
# of observations: 9420

TOTAL = 9420