The solutions for 10.1 #4 and #36 are available on the pink sheet.

Warm Up

A significance test begins by assuming that H_0: $p_1 - p_2 = 0$ is true. In that case, $p_1 = p_2$. We call the common value of these two parameters p.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}}$$

Unfortunately, we don't know the common value of p. To estimate p, we combine (or "pool") the data from the two samples as if they came from one larger sample.

$$\hat{p} = \frac{\text{number of successes in both samples combined}}{\text{number of individuals in both samples combined}} = \frac{X_1 + X_2}{n_1 + n_2}$$

$$\boxed{z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}}$$

$\hat{p}_c$ is the combined prop.

Think about the conditions that need to be met in order to create confidence intervals for proportions. Match the condition on the left with its purpose on the right.

10% Condition	So we can generalize to both populations or, in an experiment, we can show causation.
Large Counts	So sampling without replacement is OK. If the condition is met, we can use the formulas for the standard devtation.
Random Condition	So that the sampling distribution of $\hat{p}_1 - \hat{p}_2$ will be approximately Normal and we can uthen use z^* to do calculations

Lesson 10.1 – Day 3
Which grade is more likely to go to prom?

Lesson 10.1 – Day 3

Which grade is more likely to go to prom?

CLASS OF 2020 CLASS OF 2021

The student council at a large high school is wondering if Juniors or Seniors are more likely to attend Prom. They take a random sample of 50 Juniors and find that 28 are planning on attending Prom. They select a random sample of 45 Seniors and 29 are planning on attending.

Do the data provide convincing evidence that a higher proportion of Seniors are going to prom than Juniors? Use a 5% significance level.

STATE: Parameter: Statistics:

Hypotheses: Significance level:

Lesson 10.1 – Day 3

Which grade is more likely to go to prom?

CLASS OF 2020 CLASS OF 2021

The student council at a large high school is wondering if Juniors or Seniors are more likely to attend Prom. They take a random sample of 50 Juniors and find that 28 are planning on attending Prom. They select a random sample of 45 Seniors and 29 are planning on attending.

Do the data provide convincing evidence that a higher proportion of Seniors are going to prom than Juniors? Use a 5% significance level.

STATE: Parameter: $p_1 - p_2$ true difference in proportions of grade going to prom (Junior – Senior) Statistics:

Hypotheses: Significance level:

Lesson 10.1 – Day 3

Which grade is more likely to go to prom?

CLASS OF 2020 CLASS OF 2021

The student council at a large high school is wondering if Juniors or Seniors are more likely to attend Prom. They take a random sample of 50 Juniors and find that 28 are planning on attending Prom. They select a random sample of 45 Seniors and 29 are planning on attending.

Do the data provide convincing evidence that a higher proportion of Seniors are going to prom than Juniors? Use a 5% significance level.

STATE: Parameter: $p_1 - p_2$ true difference in proportions of grade going to prom (Junior – Senior)

Statistics:

Hypotheses:

Significance level:

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 < 0$

Lesson 10.1 – Day 3

Which grade is more likely to go to prom?

CLASS OF 2020 CLASS OF 2021

The student council at a large high school is wondering if Juniors or Seniors are more likely to attend Prom. They take a random sample of 50 Juniors and find that 28 are planning on attending Prom. They select a random sample of 45 Seniors and 29 are planning on attending.

Do the data provide convincing evidence that a higher proportion of Seniors are going to prom than Juniors? Use a 5% significance level.

STATE: Parameter: $p_1 - p_2$ true difference in proportions of grade going to prom (Junior – Senior)

Statistics: $\hat{p}_1 = \frac{28}{50} = 0.56$

$\hat{p}_2 = 29/45 = 0.64$

Hypotheses:

Significance level: $p_1 - p_2 = -.08$

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 < 0$

$\alpha = 0.05$

PLAN: Name of procedure:

Check conditions:

Random

10%

Large Counts
(wait... we'll need to use the pooled proportion)

PLAN: Name of procedure: Two-Sample z test for $p_1 - p_2$

Check conditions:

Random indep. rand. samples of JRS and SRS

10% assume
- $50 < \frac{1}{10}$(all JRS)
- $45 < \frac{1}{10}$(all SRS)

Large Counts
(wait... we'll need to use the pooled proportion)

NOTE

In a two-sample z test for a diff. in proportions, we assume the null hypothesis is true ($p_1 = p_2$).

So, when we get the formula for SD it does not make sense to use a different values (for p) because we are assuming the two proportions are equal.

So we must combine them

PLAN: Name of procedure: Two-Sample Z test for $p_1 - p_2$

Check conditions:

Random indep. rand. samples of JRS and SRS ✓

10% $50 < \frac{1}{10}$(all JRS) ✓
$45 < \frac{1}{10}$(all SRS)

Large Counts
(wait... we'll need to use the pooled proportion)

Use $\hat{p}_c = \frac{28+29}{50+45} = \frac{57}{95} = 0.60$

$n_1 \hat{p}_1 = 50(.60) = 30$
$n_1(1-\hat{p}_1) = 50(.40) = 20$
$n_2 \hat{p}_2 = 45(.60) = 27$
$n_2(1-\hat{p}_2) = 45(.40) = 18$

≥ 10

DO: Mean: $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2 = 0$ assumed to be equal

Standard deviation:

General Formula:

Specific Formula:

Work:

Picture for standardizing:

Test statistic:

P-value:

DO: Mean: $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2 = 0$ assumed to be equal

$\hat{p}_c$

Standard deviation:

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{0.6(.4)}{50} + \frac{0.6(.4)}{45}} \approx 0.101 \quad \text{not } 0.10$$

General Formula:

Specific Formula:

Work:

Picture for standardizing:

Test statistic:

P-value:

DO: Mean: $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2 = 0$ ← assumed to be equal

Picture for standardizing:

Standard deviation:

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{0.6(.4)}{50} + \frac{0.6(.4)}{45}} \approx 0.101 \text{ not } 0.10$$

General Formula: $\text{test stat} = \dfrac{\text{statistic} - \text{parameter}}{\text{std error}}$

Specific Formula:

Work:

$$z = \frac{(\hat{p}_1-\hat{p}_2) - (p_1-p_2)}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}}$$

$$z = \frac{-.08 - 0}{.101}$$

$$= -.792$$

Test statistic: $-.792$

P-value:

DO: Mean: $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2 = 0$ ← assumed to be equal

Picture for standardizing:

Standard deviation:

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{0.6(.4)}{50} + \frac{0.6(.4)}{45}} \approx 0.101 \text{ not } 0.10$$

General Formula: $\text{test stat} = \dfrac{\text{statistic} - \text{parameter}}{\text{std error}}$

Specific Formula:

Work:

$$z = \frac{(\hat{p}_1-\hat{p}_2) - (p_1-p_2)}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}}$$

Test statistic:

P-value:

$$z = \frac{-0.08 - 0}{.101}$$

$$z = -.79$$

DO: Mean: $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2 = 0$ assumed to be equal

Picture for standardizing:

Standard deviation:

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{0.6(.4)}{50} + \frac{0.6(.4)}{45}} \approx 0.101 \quad \text{not } 0.10$$

General Formula: $\text{test stat} = \dfrac{\text{statistic} - \text{parameter}}{\text{std error}}$

Specific Formula:

Work: $$z = \frac{(\hat{p}_1-\hat{p}_2) - (p_1-p_2)}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}}$$

Test statistic: $-.79$

P-value: 0.2148

$$z = \frac{-0.08 - 0}{.101}$$

$$z = -.79$$

CONCLUDE:

Because the p-value of $0.2148 > \alpha = .05$, *we fail to reject* H_0. *There is not convincing evidence that a higher proportion of seniors are going to prom.*

Lesson 10.1: Day 3: **A Significance Test** for $p_1 - p_2$

Hypotheses

Important ideas:

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \;\begin{matrix} > \\ < \\ \neq \end{matrix}\; 0$

Lesson 10.1: Day 3: **A Significance Test** for $p_1 - p_2$

Hypotheses

Important ideas:

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \;\begin{matrix} > \\ < \\ = \end{matrix}\; 0$

$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2 = 0$

$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}$

Lesson 10.1: Day 3: **A Significance Test** for $p_1 - p_2$

Hypotheses

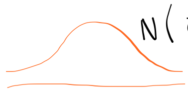
Important ideas:

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \;\overset{>}{\underset{\neq}{<}}\; 0$

$$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2 = 0$$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}$$



$N(0, \sigma_{\hat{p}_1 - \hat{p}_2})$

where $\hat{p}_c$ = Both proportions combined

Lesson 10.1: Day 3: **A Significance Test** for $p_1 - p_2$

Hypotheses

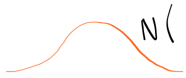
Important ideas:

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \;\overset{>}{\underset{\neq}{<}}\; 0$

$$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2 = 0$$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}$$



$N(0, \sigma_{\hat{p}_1 - \hat{p}_2})$

where $\hat{p}_c$ = Both proportions combined

$$= \frac{X_1 + X_2}{n_1 + n_2}$$

Lesson 10.1: Day 3: **A Significance Test** for $p_1 - p_2$

Hypotheses

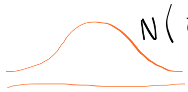
Important ideas:

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \; >, <, \neq \; 0$

$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2 = 0$

$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}$



$N(0, \sigma_{\hat{p}_1 - \hat{p}_2})$

where $\hat{p}_c$ = Both proportions combined

$= \frac{X_1 + X_2}{n_1 + n_2}$

Test statistic = $\frac{\text{Stat} - \text{Parameter}}{\text{std. error}}$

Lesson 10.1: Day 3: **A Significance Test** for $p_1 - p_2$

Hypotheses

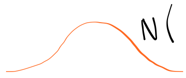
Important ideas:

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \; >, <, \neq \; 0$

$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2 = 0$

$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}$



$N(0, \sigma_{\hat{p}_1 - \hat{p}_2})$

where $\hat{p}_c$ = Both proportions combined

$= \frac{X_1 + X_2}{n_1 + n_2}$

Test statistic = $\frac{\text{Stat} - \text{Parameter}}{\text{std. error}}$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{}}$$

Lesson 10.1: Day 3: **A Significance Test** for $p_1 - p_2$

Hypotheses

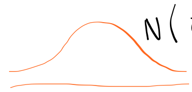
Important ideas:

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \; > / < / \neq \; 0$

$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2 = 0$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}$$



$N(0, \sigma_{\hat{p}_1 - \hat{p}_2})$

where $\hat{p}_c$ = Both proportions combined $= \frac{X_1 + X_2}{n_1 + n_2}$

Test statistic $= \frac{\text{Stat} - \text{Parameter}}{\text{std error}}$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\quad}}$$

Conclude

Because ___ $\lessgtr \alpha$, we reject / fail to reject the null and do/do not have convincing evidence of Ha context

Lesson 10.1: Day 3: A Significance Test for $p_1 - p_2$

Hypotheses

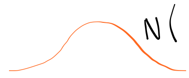
Important ideas:

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \; > / < / \neq \; 0$

$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2 = 0$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}$$



$N(0, \sigma_{\hat{p}_1 - \hat{p}_2})$

where $\hat{p}_c$ = Both proportions combined $= \frac{X_1 + X_2}{n_1 + n_2}$

Test statistic $= \frac{\text{Stat} - \text{Parameter}}{\text{std error}}$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\quad}}$$

Conclude

Because ___ $\lessgtr \alpha$, we reject / fail to reject the null and do/do not have convincing evidence of Ha context

Assuming the null is true, there is a ___ probability of getting $\hat{p}_1 - \hat{p}_2 =$ ___ or less/more/farther than purely by chance.

Whew!

AP® Exam Tip

The formula for the two-sample z statistic for a test about $p_1 - p_2$ often leads to calculation errors by students.

As a result, your teacher may recommend using the calculator's 2-PropZTest feature to perform calculations on the AP® Statistics exam. **(I Do !)**

Be sure to name the procedure (two-sample z test for $p_1 - p_2$) in the "Plan" step and **report the standardized test statistic (z = -2.32) and P-value (0.98) in the "Do" step.**

Calculator Shortcut for the **DO** section:

2 Prop Z Test gives $z = -1.32$

and P-value $= 0.043$

This disadvantage of doing this on every problem on every Significance Test for a difference of proportions (including tonight's assignment) is that you might not develop/practice some of the details for multiple choice questions.

The same formulas will apply in Ch. 12.

On to Pre-school

4-Step Process
formally

(Check conditions but
you don't have
to write down)

Example of A Significance Test for p_1-p_2

Preschool - To study the long-term effects of preschool programs for poor children, researchers designed an experiment. They recruited 123 children who had never attended preschool from low-income families in Michigan. Researchers randomly assigned 62 of the children to attend preschool (paid for by the study budget) and the other 61 to serve as a control group who would not go to preschool. One response variable of interest was the need for social services as adults. Over a 10-year period, 38 children in the preschool group and 49 in the control group have needed social services.

1. Do these data provide convincing evidence that preschool reduces the later need for social services for children like the ones in this study? Justify your answer.

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 < 0$

$\alpha = 0.05$

It is acceptable to
define p_1 and p_2 separately

STATE

PLAN

DO

CONCLUDE

STATE

H_0: $p_1 - p_2 = 0$

H_a: $p_1 - p_2 < 0$

$p_1 \rightarrow$ true prop. of children (like the ones in the study) who attend pre-school and use services

$p_2 \rightarrow$ true prop. of children (like the ones in the study) who do not attend pre-school and use services.

Use $\alpha = .05$

$\hat{p}_c = \frac{38+49}{62+61} = \frac{87}{123} = .71$

$\hat{p}_1 = .61$

$\hat{p}_2 = .8$

$\hat{p}_1 - \hat{p}_2 = -.19$

PLAN

2-Sample z test for $p_1 - p_2$

DO

$$\text{Test Stat} = \frac{\text{Stat} - \text{Null}}{\text{SD}}$$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{.71(.29)}{62} + \frac{.71(.29)}{61}}$$

$$\approx .082$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sigma_{\hat{p}_1 - \hat{p}_2}}$$

$$z = \frac{-.19 - 0}{.082}$$

$$z = -2.32$$

$$P(z < -2.32) = .0102$$

normalcdf

don't forget to multiply by 2 in some problems

Conclude

Since the P-Value of $.0102 < \alpha = .05$, we reject H_0

- There is convincing evidence that the true proportion of children who attend preschool and use services is less than the true proportion of children who attend pre-school and don't use services,

2. Based on your conclusion to Question 1, could you have made a Type I error or a Type II error? Explain your reasoning.

Because we rejected H_0 it is possible we made a Type I error.

(finding convidence that services made a difference when they did not)

Take Home LCQ and ...

10.1....15, 19, 21, 29, 31-33

study pp. 645-654

Exp. Design 20

What is *bias* in conducting surveys?

(A) An example of sampling error

(B) Lack of a control group

(C) Confounding variables

(D) Difficulty in concluding cause and effect

(E) A tendency to favor the selection of certain members of a population

Answer: (E) Poorly designed sampling techniques result in bias, that is, in a tendency to favor the selection of certain members of a population. For example, door-to-door surveys ignore the homeless, radio call-in programs give too much emphasis to persons with strong opinions, and interviews at shopping malls typically give the opinions of a very select sample of the population.