Warm Up

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

EXPERIMENTAL DESIGN 19

Two antidepressants are to be compared in the treatment of elderly patients in a nursing home. Each patient has his or her own room, some with spectacular views of the ocean. The experimental design is to create homogeneous blocks with respect to window view. How should randomization be used for a *randomized block design?*

(A) Within each block, randomly pick half the patients to receive each antidepressant.

(B) Randomly pick half of all patients to receive each antidepressant, but then analyze the results separately by blocks.

(C) Randomly choose which blocks will receive which antidepressant.

(D) Randomly choose half the blocks to receive each antidepressant for a given time period; then for the same time period switch the medication in each block and compare the results.

(E) For ethical reasons, allow patients to choose which medication they prefer taking, but then randomly assign patients to the blocks.

Answer: (A) In a randomized block design, we assign subjects to treatments at random *within* each of the homogeneous blocks. In effect, to reduce variability we run parallel experiments on the blocks.

On the AP Statistics exam there will be 6 free response questions, the last of which will be an investigative task.......

where you will need to think about a new concept that you did not cover in the AP Stats course.

It is always something that can be done using good statistical thinking and reasoning.

The activity we are about to do will take what you already know about significance tests and apply it in a new context.

You will be asked to do a full (four step) significance test for a difference of proportions..... before I show you the ins and outs.

Please do not refer to your textbook.

You can, however, use your AP formula sheet (first option) and each other. Use your previous notes (as a last option)

We'll refine later as a class.

Are some groups underrepresented? 10.1 Day 3

According to phys.org, Black and Hispanic females are underrepresented in STEM programs compared to non-STEM programs. A certain university would like to see if this is true for their student population. They took a random sample of 300 STEM students and found that 12 were Black or Hispanic females. A separate random sample of 500 non-STEM students had 75 Black or Hispanic females.

Do the data provide convincing evidence that Black and Hispanic females are underrepresented in STEM programs? Use a 5% significance level.

STATE: Parameter: Statistic:

Hypotheses: Significance level:

PLAN: Name of procedure:

Check conditions:

You can skip the conditions

We'll look at the work of individuals for $\boxed{\text{State}}$ and $\boxed{\text{DO}}$ → Stop before you calculate the test statistic

Are some groups underrepresented? 10.1 Day 3

According to phys.org, Black and Hispanic females are underrepresented in STEM programs compared to non-STEM programs. A certain university would like to see if this is true for their student population. They took a random sample of 300 STEM students and found that 12 were Black or Hispanic females. A separate random sample of 500 non-STEM students had 75 Black or Hispanic females.

Do the data provide convincing evidence that Black and Hispanic females are underrepresented in STEM programs? Use a 5% significance level.

STATE: Parameter: $p_1 - p_2$ **differ. in the true prop. of Black and HIspanic females in STEM and non-stem.** Statistic: $\hat{p}_1 - \hat{p}_2 = .04 - .15 = -.11$

Hypotheses: $p_1 - p_2 = 0$, $p_1 - p_2 < 0$ Significance level: $\alpha = .05$

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

Check conditions:

You can skip the conditions

DO:

Mean: $\mu_{\hat{p}_1 - \hat{p}_2} = 0$ Standard deviation: $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{.04(.96)}{300} + \frac{.15(.85)}{500}} = .02$

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

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

Work: Picture:

Test statistic: P-value:

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

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.

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

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.

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.

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

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.

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

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

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

in some places (on-line sites / older textbooks) $\hat{p}_c$ might be seen

don't use it because we'll be saving the "c" for something else.

DO:

Mean: $\mu_{\hat p_1-\hat p_2}=0$ Standard deviation: $\sigma_{\hat p_1-\hat p_2}=\cancel{\sqrt{\frac{.04(.96)}{300}+\frac{.15(.85)}{500}}}$

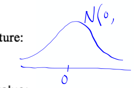
General Formula: Test Stat = $\frac{\text{Stat}-\text{Null}}{SD}$

Specific Formula: $Z=\cancel{\frac{(\hat p_1-\hat p_2)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}}$

$p=\frac{12+75}{300+500}=\frac{87}{800}=.109$

Work:

Picture: $N(0,\ )$



Test statistic: $Z=\frac{-.11-0}{\sqrt{\frac{.109(1-.109)}{300}+\frac{.109(1-.109)}{500}}}=-4.838$

P-value:

DO:

Mean: $\mu_{\hat p_1-\hat p_2}=0$ Standard deviation: $\sigma_{\hat p_1-\hat p_2}=\cancel{\sqrt{\frac{.04(.96)}{300}+\frac{.15(.85)}{500}}}$

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

Specific Formula: $Z=\cancel{\frac{(\hat p_1-\hat p_2)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}}$

$p=\frac{12+75}{300+500}=\frac{87}{800}=.109$

Work:

Picture:

Test statistic: $Z=\frac{-.11-0}{\sqrt{\frac{.109(1-.109)}{300}+\frac{.109(1-.109)}{500}}}=-4.838$

P-value:

$P(Z<-4.838)=.0000006$

CONCLUDE:

Since the P-Value of $.0000006<\alpha=.05$ We reject Ho

$p_1=$

$p_2=$

∴ There is convincing evidence to support the claim that the true difference in proportions of Black & ~~Hispanic~~ Latino ~~females in~~ STEM and NON-STEM is ~~greater~~ less than than 0.

Reminder

If you ever fail to reject Ho, be sure you DON'T "accept".

That would be like saying the true prop. is 0 but we would not necessarily know that.

Experience first

↓

Formalize

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

Important ideas:

Now Let's summarize

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

Important ideas:

Hypotheses:

Both Proportions Combined

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \gtrless \neq 0$

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

Important ideas:

Hypotheses:

Both Proportions Combined

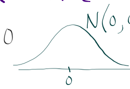
$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \gtrless \neq 0$

Same as $H_0: p_1 = p_2$ but we won't do this

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

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



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

Important ideas:

Hypotheses:

Both Proportions Combined

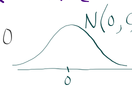
$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \gtrless \neq 0$

Same as $H_0: p_1 = p_2$ but we won't do this

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

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



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

where $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$

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.

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 < 0$

$\alpha = 0.05$

$p_1 =$

$p_2 =$

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

STATE

PLAN

DO

CONCLUDE

STATE

H_0: $p_1 - p_2 = 0$

H_a: $p_1 - p_2 < 0$

$p_1 - p_2 = -.19$

$p_1 \rightarrow$ 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 $\alpha = .05$

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

PLAN

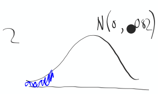
2-Sample z test for $p_1 - p_2$

DO

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

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

$\approx .082$



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

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

$z = -2.32$

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

normalcdf

don't forget to multiply by 2 in some problems

Conclude

Since the P-Value of .0102 < α = .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)

AP® Exam Tip

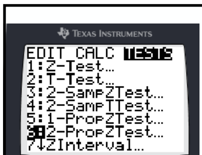
The formula for the two-sample z statistic for a test about $p_1 - p_2$ often leads to calculation errors by students.

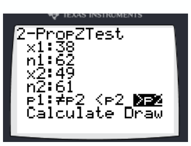
As a result, your teacher may recommend using the calculator's 2-PropZTest feature to perform calculations on the AP® Statistics exam.
(I Do !)

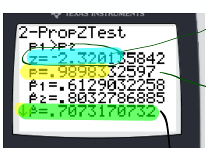
Be sure to name the procedure (two-sample z test for $p_1 - p_2$) in the "Plan" step and report the standardized test statistic (z = -2.32) and P-value (0.98) in the "Do" step.

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.







Statistic

P-Value

Combined proportion

$\hat{p}$

See your Ch. 9 Test

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.