AP Statistics
Unit 07 – Day 02 Notes
Confidence Intervals: 2-Sample Proportions

Name Key
Period

HE SAMPLING DISTRIBUTION OF $\hat{p}_1 - \hat{p}_2$

SHAPE: When $n_1p_1 \geq 10$ and $n_2p_2 \geq 10$ then the sampling distribution
$n_1q_1 \geq 10$ $n_2q_2 \geq 10$ of $\hat{p}_1 - \hat{p}_2$ is approximately Normal.

CENTER: the mean of the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is $p_1 - p_2$.

SPREAD: Standard deviation = $\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$ as long as $10n < N$

Follow the same 4 steps as with one-sample:

STATE:

p_1 = $\widehat{p_1}$ =
p_2 = $\widehat{p_2}$ =
We want to estimate the difference $p_1 - p_2$ with ___% confidence.

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 (no hat)
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-interval for a difference between two proportions $p_1 - p_2$.**

DO:

include ~~the~~ the formula with the appropriate numbers to the problem (then you can plug it into the calculator!).

$$(\widehat{p_1} - \widehat{p_2}) \pm z^* \sqrt{\frac{\widehat{p_1}(\widehat{q_1})}{n_1} + \frac{\widehat{p_2}(\widehat{q_2})}{n_2}} = \text{(lower value, upper value)}$$

CONCLUDE:

We are C% confident that the interval from [lower value] to [upper value] captures the true difference in the proportion of [what the situation is in context].

CALCULATOR USE:

Option B

Stats → Tests → 2-PropZInt
- X1: sample 1
- n1: total number in sample 1
- X2: sample 2
- n2: total number in sample 2
- C-level: C%
- Calculate

Example 1: Thirty-five (x_1) people from a random sample of 125 (n_1) workers from Company A admitted to using sick leave when they weren't really ill. Seventeen (x_2) employees from a random sample of 68 (n_2) workers from Company B admitted that they had used sick leave when they weren't ill. Construct and interpret a 95% confidence interval for the difference in the proportions of workers at the two companies who would admit to using sick leave when they weren't really sick.

State: p_1 = the true proportion of company A workers who use sick leave when not ill.

p_2 = the true proportion of company B workers who use sick leave when not ill.

we want to estimate the difference in proportion $p_1 - p_2$ with 95% confidence.

$\hat{p}_1 = \frac{35}{125} = 0.28$ $\quad \hat{p}_2 = \frac{17}{68} = 0.25$

Plan: random: the data come from 2 independent random samples ✓

10% condition: $n_1 = 125$ $\quad 1250 <$ all workers at company A

$n_2 = 68$ $\quad 680 <$ all workers at company B ✓

Large counts: $n_1\hat{p}_1 = 35 \geq 10$ ✓ $\quad n_2\hat{p}_2 = 17 \geq 10$ ✓

$n_1\hat{q}_1 = 90 \geq 10$ ✓ $\quad n_2\hat{q}_2 = 51 \geq 10$ ✓

Because our conditions are met, we will construct a 2-sample z-interval for difference in 2 proportions $p_1 - p_2$.

Do: $$(0.28 - 0.25) \pm 1.96\sqrt{\frac{(0.28)(1-0.28)}{125} + \frac{(0.25)(1-0.25)}{68}} = \boxed{(-0.0996, 0.15957)}$$

$z^* = 1.96$

Conclude: we are 95% confident that the interval from -0.0996 to 0.15957 captures the true difference in proportions of workers at Companies A and B who admit to using sick leave when not really ill.

Example 2: Mrs. Mapstone wanted to determine if there were a difference in proportion of passing scores between her classes and Mrs. DeMarre's classes since teaching at BHS. Mrs. Mapstone took a random sample of 125 students and Mrs. DeMarre took a random sample of 53 students. Mrs. Mapstone found that 102 of her students' scores were passing and Mrs. DeMarre found that 43 of her students' scores were passing.

a) Construct and interpret a 90% confidence interval to determine the difference in proportions between the scores of Mrs. Mapstone's classes and Mrs. DeMarre's classes.

State: p_1 = the true proportion of passing students in Mrs. D's classes
p_2 = the true proportion of passing students in Mrs. Mapstone's classes
We want to estimate the difference in proportion $p_1 - p_2$ with 90% confidence. $\hat{p}_1 = \frac{43}{53} = 0.811$ $\hat{p}_2 = \frac{102}{125} = 0.816$

Plan: random: the data come from 2 independent random samples ✓
10% condition: $n_1 = 53$ 530 < all of Mrs. D's students ever ✓
$n_2 = 125$ 1250 < all of Mrs. Mapstone's students ever

Large counts: $n_1\hat{p}_1 = 43 \geq 10$ ✓ $n_2\hat{p}_2 = 102 \geq 10$ ✓
$n_1\hat{q}_1 = 10 \geq 10$ ✓ $n_2\hat{q}_2 = 23 \geq 10$ ✓

Because our conditions are met, we will construct a 2-sample z-interval for the difference between 2 proportions $p_1 - p_2$.

Do: $$(0.811 - 0.816) \pm 1.645\sqrt{\frac{(0.811)(1-0.811)}{53} + \frac{(0.816)(1-0.816)}{125}}$$

$z^* = 1.645$

$$= \boxed{(-0.1099, 0.10051)}$$

Conclude: We are 90% confident that the interval from -0.1099 to 0.10051 captures the true difference in proportions of students passing Mrs. D's and Mrs. Mapstone's classes ($p_1 - p_2$).

b) Does your interval from part (a) give convincing evidence of a difference between the population proportions? Explain. Nope. Because 0 is in our interval of plausible values, there is not convincing evidence of a difference between the true proportion of passing students between Mrs. D's and Mrs. Mapstone's classes.

0 = no difference