Comparing 2 Populations

$\boxed{10.1}$ - 2 proportions

Notation:

	Population 1	Population 2
true proportions (parameters)	p_1	p_2
sample sizes	n_1	n_2
# of successes	X_1	X_2
sample proportions	$\hat{p}_1 = \frac{X_1}{n_1}$	$\hat{p}_2 = \frac{X_2}{n_2}$

Compare 2 populations by doing inference on the difference.

$\hat{p}_1 - \hat{p}_2$ estimates $p_1 - p_2$

Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

- shape: approx. Normal if
  $n_1 p_1 \geq 10$
  $n_1(1-p_1) \geq 10$
  $n_2 p_2 \geq 10$
  $n_2(1-p_2) \geq 10$
- Center: Mean = $p_1 - p_2$
- Spread: $\sigma = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

p. 607

Confidence Intervals for comparing 2 proportions

Statistic ± crit. value · St. dev. of statistic

$$\hat{p}_1 - \hat{p}_2 \pm \underbrace{z^* \underbrace{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}_{\text{standard error}}}_{\text{margin of error}}$$

*called a 2 sample z Interval for a difference between 2 proportions

on calc: $\boxed{\text{Stat}}$ → tests → B: 2-prop z Int follow same 4-step process

Conditions:
1. Random — both samples must be randomly selected.
2. Normal — $n_1\hat{p}_1 \geq 10$
$n_1(1-\hat{p}_1) \geq 10$
$n_2\hat{p}_2 \geq 10$
$n_2(1-\hat{p}_2) \geq 10$
3. Independent — samples taken independently and 10% rule should be met for both samples.

---

ex:

1. State: p_1 = the true proportion of U.S. teens who use social network sites
p_2 = the " " " " adults " " " ".

C-level: 95%

2. Plan: Method is a 2-Sample Z Interval for a difference between 2 proportions.

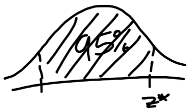
Conditions:
1. Random — random sample of 800 teens and another random sample of 2253 adults were taken separately
2. Normal — $n_1\hat{p}_1 = 800(.73) = 584 \geq 10$
$n_1(1-\hat{p}_1) = 800(.27) = 216 \geq 10$
$n_2(\hat{p}_2) = 2253(.47) = 1059 \geq 10$
$n_2(1-\hat{p}_2) = 2253(.53) = 1194 \geq 10$
3. Independent — 2 samples were taken independently.
there are more than 8000 U.S. teens and 22530 U.S. adults.

---

3. Do

$$.73 - .47 \pm 1.96\sqrt{\frac{.73(.27)}{800} + \frac{(.47)(.53)}{2253}}$$

$n_1 = 800$
$n_2 = 2253$
$\hat{p}_1 = .73$
$\hat{p}_2 = .47$



0.26 ± 0.037

$(.223, .297)$

4. Conclude: We are 95% confident that the interval (.223, .297) captures the true difference in the proportions of U.S. teens & adults who use social networking sites.