```
Stat 145, Mon 29-Mar-2021 -- Mon 29-Mar-2021
Biostatistics
Spring 2021


-------------------------------
Monday, March 29th 2021
-------------------------------
Due::   PS10 due at 11 pm

-------------------------------
Monday, March 29th 2021
-------------------------------
Wk 9, Mo
Topic:: Inference on two proportions
Read::  Lock5 6.7-6.9

Warmup
 3. paired data (a.k.a. matched-pairs t)
     Wetsuits data

Math symbols in R Markdown
 - two math modes
    what comes between single dollar signs
    what comes between double dollar signs
 - greek letters: a backslash followed by letter name spelled out
    \alpha
    \rho
    \sigma
 - "hat" and "bar" additions to a symbol
 - subscripts and superscripts
 - square roots
 - fractions
 - special symbols: less than, greater than, not equal to, plus or minus
 - hypotheses

 For more, see  http://scofield.site/courses/s145/tutorials/mathSymbols.pdf

Inference on two proportions
 - have done already using bootstrapping
 - parameter: both p_1, p_2 are relevant, but its their difference in focus
```

- statistics
  p-hat_1 - p-hat_2
- normality?

Have two populations

p_1 = proportion in 1st, p_2 = proportion in 2nd

True focus on $p_1 - p_2$

Have from independent samples

from Population 1	from Population 2
$\hat{p}_1$ from sample of size n_1	$\hat{p}_2$ from sample of size n_2

Data sets: Skip for now

1. 379 of 460 females support tougher gun-control laws, 318 of 520 males
2. 10 of 24 cocaine addicts treated with desipramine had relapses, compared with 20 of 24 who received placebo

Normal dists for $\hat{p}_1$, $\hat{p}_2$? Can check rules of thumb

$$\left.\begin{matrix} n_1\hat{p}_1 \geq 10 & \quad n_2\hat{p}_2 \geq 10 \\ n_1(1-\hat{p}_1) \geq 10 & \quad n_2(1-\hat{p}_2) \geq 10 \end{matrix}\right\}$$

If met, then normality is valid

$$\Rightarrow \left.\begin{matrix} \hat{p}_1 \sim \text{Norm}\left(p_1, \sqrt{\dfrac{p_1(1-p_1)}{n_1}}\right) \\ \hat{p}_2 \sim \text{Norm}\left(p_2, \sqrt{\dfrac{p_2(1-p_2)}{n_2}}\right) \end{matrix}\right\}$$

Comes from CLT and Sections 6.1–6.3

It's their difference

$$\hat{p}_1 - \hat{p}_2$$

that I'm interested in — as estimate for $p_1 - p_2$.

$$\hat{p}_1 - \hat{p}_2 \sim \text{Norm}\left(p_1 - p_2, \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$$

Ex.) Construct a 90% CI for the difference $p_F - p_M$

where

p_F = proportion of females in population favoring stronger gun control laws

p_M = same for men

estimate is

$\hat{p}_F - \hat{p}_M$

$= \frac{379}{460} - \frac{318}{520}$

Have data

$$\hat{p}_F = \frac{379}{460} \quad (n_F = 460)$$

$$\hat{p}_M = \frac{318}{520} \quad (n_M = 520)$$

The rules of thumb are met

$$460 \, \frac{379}{460} = 379, \quad 460\left(1 - \frac{379}{460}\right) = 81$$

$$520\left(\frac{318}{520}\right) = 318, \quad 520\left(1 - \frac{318}{520}\right) = 202$$

Expect

$$SE_{\hat{p}_F - \hat{p}_M} = \sqrt{\frac{\left(379/460\right)\left(1 - \frac{379}{460}\right)}{460} + \frac{\left(318/520\right)\left(1 - 318/520\right)}{520}}$$

$= 0.0278$

CI : Centred interval approach

$$\text{estimate } \hat{p}_F - \hat{p}_M \pm (z^*\text{-critical val})(SE_{\hat{p}_1 - \hat{p}_2})$$

$$\left(\frac{379}{460} - \frac{318}{520}\right) \pm (1.645)(0.0278)$$

$$[0.167, 0.258]$$