Section 6.4: cont'd:

Wednesday, March 7, 2018 4:24 PM

example: Camosun wants to know the difference (if any) between rates of smoking for students on the two different campuses. Assume that equal numbers of students will be surveyed on each campus. The sampling error in the difference between the two proportions is required to be no larger than $\pm 3\%$. How large a sample size is required if

a) the rate of smoking for young adults in Canada is known to be 28%

b) if the rate of smoking is initially unknown

a) want MOE $\leq B$ where $B = 0.03$

$z_{\alpha/2} SE \leq B$

$$z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \leq B$$

$n_1 = n_2 = n$ and assume $\hat{p}_1 \approx \hat{p}_2 \approx \hat{p}$

$$z_{\alpha/2} \sqrt{\frac{2\hat{p}\hat{q}}{n}} \leq B$$

$$n \geq 2\hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{B}\right)^2$$

$$\geq 2(0.28)(0.72)\left(\frac{1.96}{0.03}\right)^2$$

$$\geq 1721$$

b) will still have

$$n \geq 2\hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{B}\right)^2$$

but we have no info on $\hat{p}$ and $\hat{q}$

brute force technique: try $\hat{p} = 0.5$
$\hat{p} = 0.95$

and look for worst case scenario

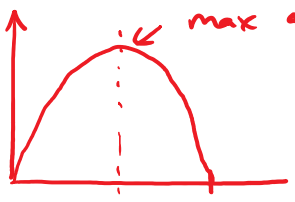
for $\hat{p} = 0.5$ (50%) get $n \geq 2134$ — worst case - go with this!
$\hat{p} = 0.95$ (95%) get $n \geq 406$

---

more sophisticated approach:

want to maximize $\hat{p}\hat{q}$

which equals $\underbrace{\hat{p}(1-\hat{p})}_{\text{parabola}}$



so stick with $\hat{p} = 0.5$ and plan on a sample size of ~~2134~~ 2150