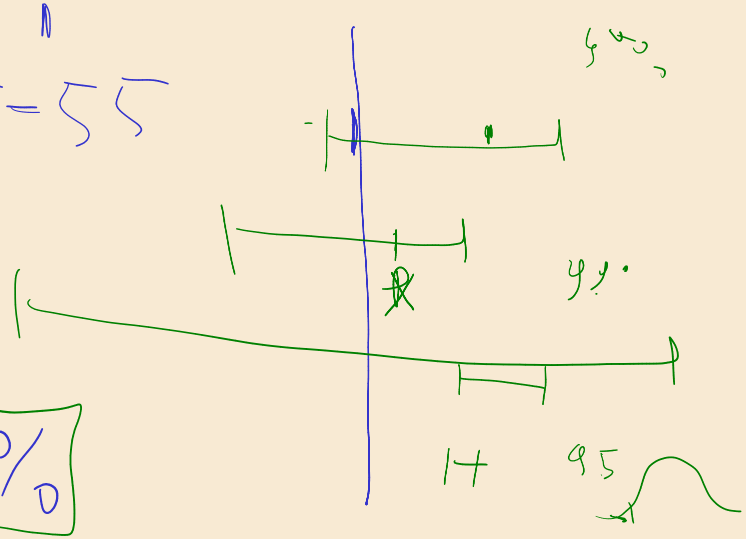
🧀 01 ✏️ Election Night Nail-Biter

A poll of $n = 900$ likely voters finds that 52% support candidate A.

- a. Build a 95% **Wald** CI for the true proportion p. Based on this interval, can you call the election for candidate A?
- b. Now compute the **Wilson** CI for the same data. How does it differ from the Wald CI?
- c. Imagine only $n = 20$ voters were polled and 11 (55%) said A. Recompute both Wald and Wilson CIs. Which one behaves better near the boundary, and why?

45-55



900

$\boxed{52\%}$

$\boxed{52\%}$

X%

0.33 - 0.76

0.34 - 0.76

1, 0, 0, ,

$\underbrace{52}$ $\pm\ z_{0.025} \frac{\sigma}{\sqrt{n}}$

1.96

$\boxed{p = 0.52}$ $\sigma^2 = p(1-p)$

$52 \pm 1.96 \frac{z}{1+z^2/n} \sqrt{\frac{p(1-p)}{900}} + \frac{z^2}{4n}$

Wald CI (0.48, 0.55) P = 0, 1

Wilson CI

48.55 $\frac{z^2}{n}$ $\frac{z}{1+z^2/n} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}$

🧀 🧀 02 ✏️ A/B Test: Ship It or Wait? 🔗

An e-commerce company runs an A/B test on their checkout page.

- Control ($n_1 = 200$): conversion rate $\hat{p}_1 = 0.08$ (8%).
- Treatment ($n_2 = 200$): conversion rate $\hat{p}_2 = 0.11$ (11%).

The product manager says “3% lift — ship it!”

P $\quad$ $P_1 - P_2$

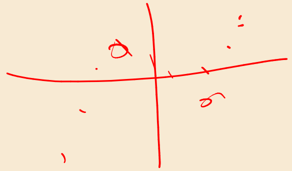
- a. Build a 95% CI for the difference $p_2 - p_1$.

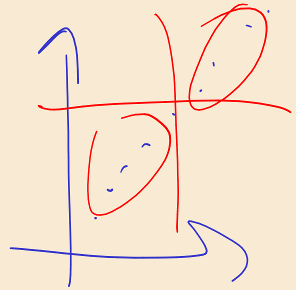
  Recall: $\mathrm{SE} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$.
- b. Does the CI contain 0? What does this tell the PM?
- c. How large would each group need to be (equal sizes) so that the expected CI width is narrow enough to exclude 0, assuming the true lift really is 3%? (Use the sample-size formula for two proportions.)

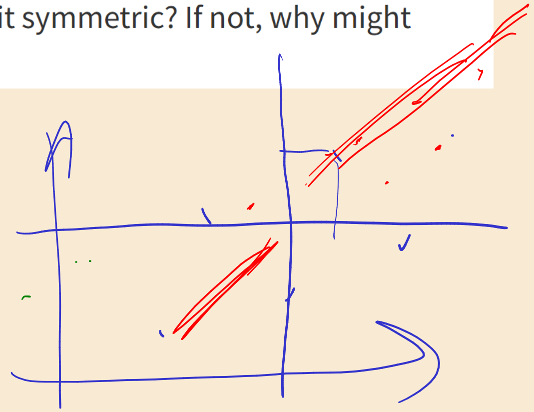
Given these (x, y) pairs:

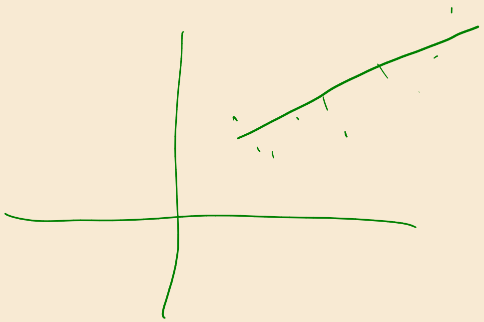
$$(1,2),\ (2,3),\ (3,5),\ (4,4),\ (5,7),\ (6,8),\ (7,6),\ (8,9),\ (9,10),\ (10,12)$$

- a. Compute the Pearson correlation r.
- b. Use the bootstrap ($B = 5{,}000$) to build a 95% percentile CI for the population correlation ρ.
- c. Plot the bootstrap distribution of r^*. Is it symmetric? If not, why might that be?









$(x - E\,x)$

$(y - E(y))$

Spe

X, y K

$\hat{\theta} = (X^T X)^{-1} X^T y$

$= cov(x,y)$

$$\frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i} (y_i - x)^2}}$$

rank(x)