M358K : November 16^{th}, 2020.

$p_{<40K} - p_{\geq 40K}$

confidence interval : $(-0.16, 0.02)$

(a) No difference is the same as $p_{<40K} = p_{\geq 40K}$

$$\Longleftrightarrow \quad p_{<40K} - p_{\geq 40K} = 0$$

False!

(d) **At the 95% confidence level:**

$$-0.16 < p_{<40K} - p_{\geq 40K} < 0.02 \quad / \cdot (-1)$$

$$\Rightarrow \quad 0.16 > p_{\geq 40K} - p_{<40K} > -0.02$$

$\Rightarrow$ $(-0.02, 0.16)$ is the 95% conf. interval for $p_{\geq 40K} - p_{<40K}$

The standard error is the same regardless of how you order p_1 and p_2 in the difference:

$$\boxed{\text{pt. estimate} \pm \underbrace{z^*(\text{std error})}_{\text{m.e.}}}$$

For $p_1 - p_2$: $(\hat{p}_1 - \hat{p}_2) \pm \text{m.e.}$

For $p_2 - p_1$: $(\hat{p}_2 - \hat{p}_1) \pm \text{m.e.}$

#6.24.

$H_0: p_{OR} = p_{CA}$ vs. $H_a: p_{OR} \neq p_{CA}$

Conditions (Look @ grey boxes in the textbook)

1. Independence: sample size <10% population
2. Success-failure (under the null)

$\hat{p}$ ... pooled estimate of the total population proportion

$$\hat{p} = \frac{x_{CA} + x_{OR}}{n_{CA} + n_{OR}} = \frac{\hat{p}_{CA} \cdot n_{CA} + \hat{p}_{OR} \cdot n_{OR}}{n_{CA} + n_{OR}}$$

$n_{CA} \cdot \hat{p} \geq 10$ $\quad n_{CA}(1-\hat{p}) \geq 10$

$n_{OR} \cdot \hat{p} \geq 10$ $\quad n_{OR}(1-\hat{p}) \geq 10$