1. **Problem**
   In one population, 12.1% are abysmal ($p_1 = 0.121$). In a second population, 62% are abysmal ($p_2 = 0.62$). When random samples of sizes 700 and 900 are taken from the first and second populations respectively, what is the chance that $\hat{P}_2 - \hat{P}_1$ is at most 0.447?

Solution
Check if we expect the $\hat{P}_2 - \hat{P}_1$ sampling to follow a normal distribution. The random sampling from two (presumably very large) populations allows us to assume independence. The inequalities are also satisfied:

$$\begin{aligned} n_1 p_1 &> 10 \\ n_1(1-p_1) &> 10 \\ n_2 p_2 &> 10 \\ n_2(1-p_2) &> 10 \end{aligned}$$

So, we do expect $\hat{P}_2 - \hat{P}_1$ sampling to follow a normal distribution.

$$\hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(p_2 - p_1,\ SE)$$

Calculate the expected difference.

$$\begin{aligned} p_2 - p_1 &= 0.62 - 0.121 \\ &= 0.499 \end{aligned}$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\ &= \sqrt{\frac{0.121(1-0.121)}{700} + \frac{0.62(1-0.62)}{900}} \\ &= 0.0203 \end{aligned}$$

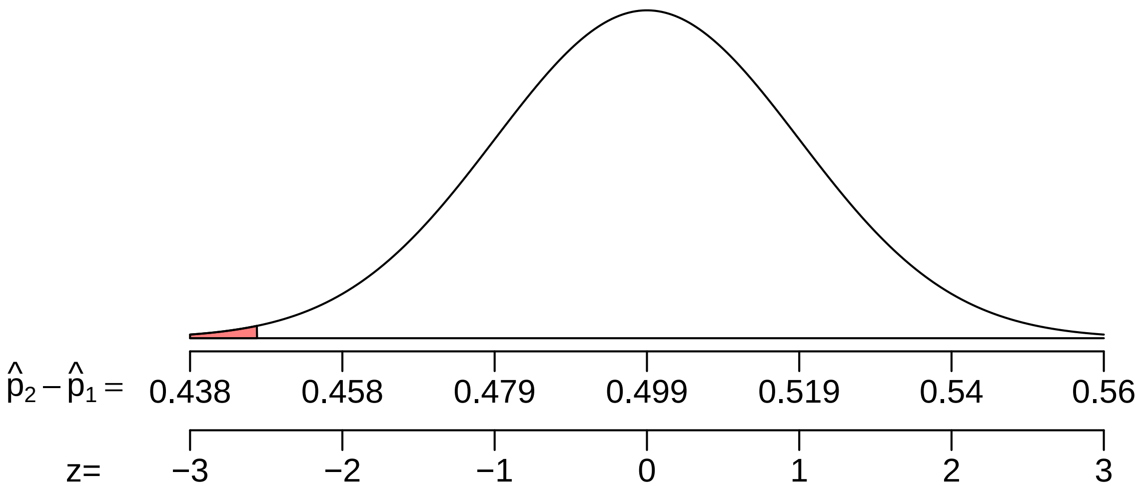
We have the parameters for $\hat{P}_2 - \hat{P}_1$ sampling.

$$\hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0.499,\ 0.0203)$$

Determine a z score for the boundary $\hat{p}_2 - \hat{p}_1 = -0.427$.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)}{SE} \\ &= \frac{(0.447) - (0.499)}{0.0203} \\ &= -2.56 \end{aligned}$$

Draw a sketch. The phrase “at most 0.447” suggests finding a left area.



Use a z table.

$$\begin{aligned}\Pr\left(\hat{P}_2 - \hat{P}_1 < 0.447\right) &= \Pr(Z < -2.56)\\ &= \Phi(-2.56)\\ &= 0.0052\end{aligned}$$

Thus, we conclude that there is a 0.52% chance that $\hat{P}_2 - \hat{P}_1$ is at most 0.447.

2. **Problem**
   In one population, 51.4% are abysmal ($p_1 = 0.514$). In a second population, 69% are abysmal ($p_2 = 0.69$). When random samples of sizes 60 and 100 are taken from the first and second populations respectively, what is the chance that $\hat{P}_2 - \hat{P}_1$ is over 0.315?

Solution

Check if we expect the $\hat{P}_2 - \hat{P}_1$ sampling to follow a normal distribution. The random sampling from two (presumably very large) populations allows us to assume independence. The inequalities are also satisfied:

$$\begin{aligned} n_1 p_1 &> 10 \\ n_1(1-p_1) &> 10 \\ n_2 p_2 &> 10 \\ n_2(1-p_2) &> 10 \end{aligned}$$

So, we do expect $\hat{P}_2 - \hat{P}_1$ sampling to follow a normal distribution.

$$\hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(p_2 - p_1,\ SE)$$

Calculate the expected difference.

$$\begin{aligned} p_2 - p_1 &= 0.69 - 0.514 \\ &= 0.176 \end{aligned}$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\ &= \sqrt{\frac{0.514(1-0.514)}{60} + \frac{0.69(1-0.69)}{100}} \\ &= 0.0794 \end{aligned}$$

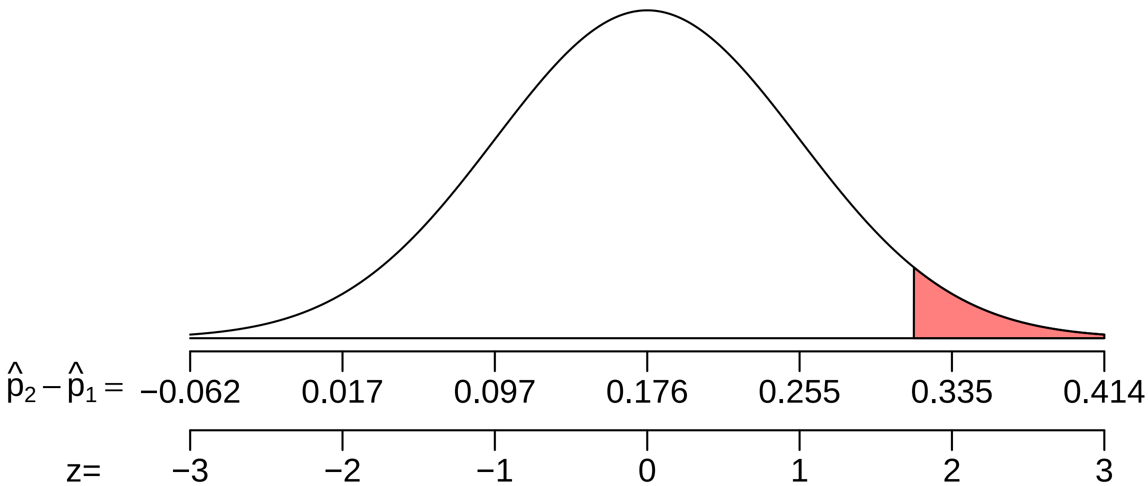
We have the parameters for $\hat{P}_2 - \hat{P}_1$ sampling.

$$\hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0.176,\ 0.0794)$$

Determine a z score for the boundary $\hat{p}_2 - \hat{p}_1 = 0.315$.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)}{SE} \\ &= \frac{(0.315) - (0.176)}{0.0794} \\ &= 1.75 \end{aligned}$$

Draw a sketch. The phrase “over 0.315” suggests finding a right area.



Use a z table.

$$\begin{aligned}\Pr\left(\hat{P}_2 - \hat{P}_1 > 0.315\right) &= \Pr(Z > 1.75)\\ &= 1 - \Phi(1.75)\\ &= 0.0401\end{aligned}$$

Thus, we conclude that there is a 4.01% chance that $\hat{P}_2 - \hat{P}_1$ is over 0.315.

3. **Problem**
   In one population, 45.7% are abysmal ($p_1 = 0.457$). In a second population, 27% are abysmal ($p_2 = 0.27$). When random samples of sizes 400 and 2000 are taken from the first and second populations respectively, what is the chance that $\hat{P}_2 - \hat{P}_1$ is between -0.205 and -0.169?

Solution

Check if we expect the $\hat{P}_2 - \hat{P}_1$ sampling to follow a normal distribution. The random sampling from two (presumably very large) populations allows us to assume independence. The inequalities are also satisfied:

$$\begin{aligned} n_1 p_1 &> 10 \\ n_1(1-p_1) &> 10 \\ n_2 p_2 &> 10 \\ n_2(1-p_2) &> 10 \end{aligned}$$

So, we do expect $\hat{P}_2 - \hat{P}_1$ sampling to follow a normal distribution.

$$\hat{P}_2 - \hat{P}_1 \quad \sim \quad \mathcal{N}(p_2 - p_1,\, SE)$$

Calculate the expected difference.

$$\begin{aligned} p_2 - p_1 &= 0.27 - 0.457 \\ &= -0.187 \end{aligned}$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\ &= \sqrt{\frac{0.457(1-0.457)}{400} + \frac{0.27(1-0.27)}{2000}} \\ &= 0.0268 \end{aligned}$$

We have the parameters for $\hat{P}_2 - \hat{P}_1$ sampling.

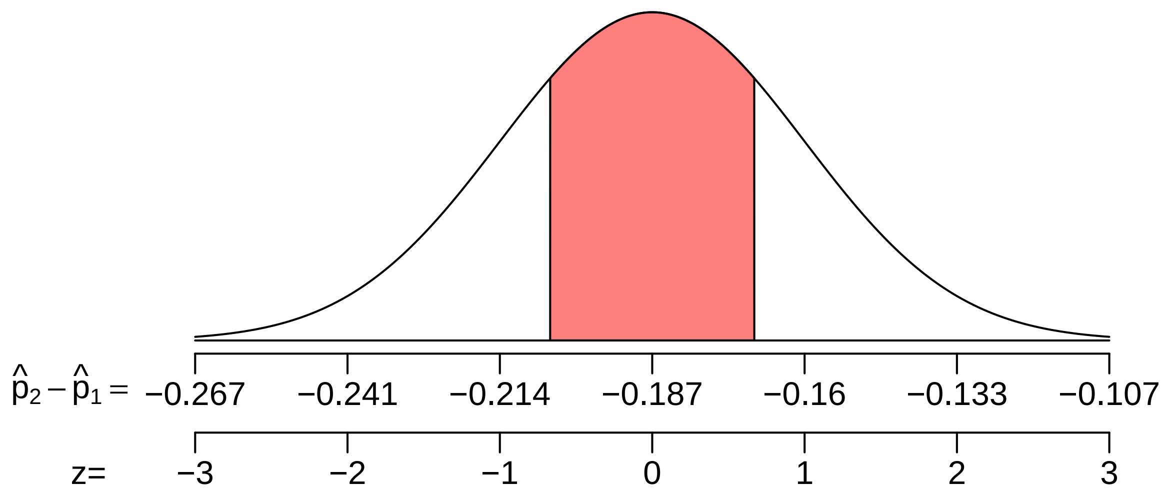
$$\hat{P}_2 - \hat{P}_1 \quad \sim \quad \mathcal{N}(-0.187,\, 0.0268)$$

Determine z scores of the boundaries.

$$\begin{aligned} z_{\text{lower}} &= \frac{(\hat{p}_2 - \hat{p}_1)_{\text{lower}} - (p_2 - p_1)}{SE} \\ &= \frac{(-0.205) - (-0.187)}{0.0268} \\ &= -0.67 \end{aligned}$$

$$\begin{aligned} z_{\text{upper}} &= \frac{(\hat{p}_2 - \hat{p}_1)_{\text{upper}} - (p_2 - p_1)}{SE} \\ &= \frac{(-0.169) - (-0.187)}{0.0268} \\ &= 0.67 \end{aligned}$$

Draw a sketch. The phrase “between -0.205 and -0.169” suggests finding a central area.



Use a z table.

$$\begin{aligned}\Pr\left(-0.205 < \hat{P}_2 - \hat{P}_1 < -0.169\right) &= \Pr(|Z| < 0.67) \\ &= 2 \cdot \Phi(0.67) - 1 \\ &= 0.4972\end{aligned}$$

Thus, we conclude that there is a 49.72% chance that $\hat{P}_2 - \hat{P}_1$ is between -0.205 and -0.169.

4. **Problem**
In one population, 98.2% are sick ($p_1 = 0.982$). In a second population, 80.1% are sick ($p_2 = 0.801$). When random samples of sizes 600 and 400 are taken from the first and second populations respectively, what is the chance that $\hat{P}_2 - \hat{P}_1$ is outside the interval (-0.201, -0.161)?

Solution

Check if we expect the $\hat{P}_2 - \hat{P}_1$ sampling to follow a normal distribution. The random sampling from two (presumably very large) populations allows us to assume independence. The inequalities are also satisfied:

$$\begin{aligned} n_1 p_1 &> 10 \\ n_1(1-p_1) &> 10 \\ n_2 p_2 &> 10 \\ n_2(1-p_2) &> 10 \end{aligned}$$

So, we do expect $\hat{P}_2 - \hat{P}_1$ sampling to follow a normal distribution.

$$\hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(p_2 - p_1,\, SE)$$

Calculate the expected difference.

$$\begin{aligned} p_2 - p_1 &= 0.801 - 0.982 \\ &= -0.181 \end{aligned}$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\ &= \sqrt{\frac{0.982(1-0.982)}{600} + \frac{0.801(1-0.801)}{400}} \\ &= 0.0207 \end{aligned}$$

We have the parameters for $\hat{P}_2 - \hat{P}_1$ sampling.

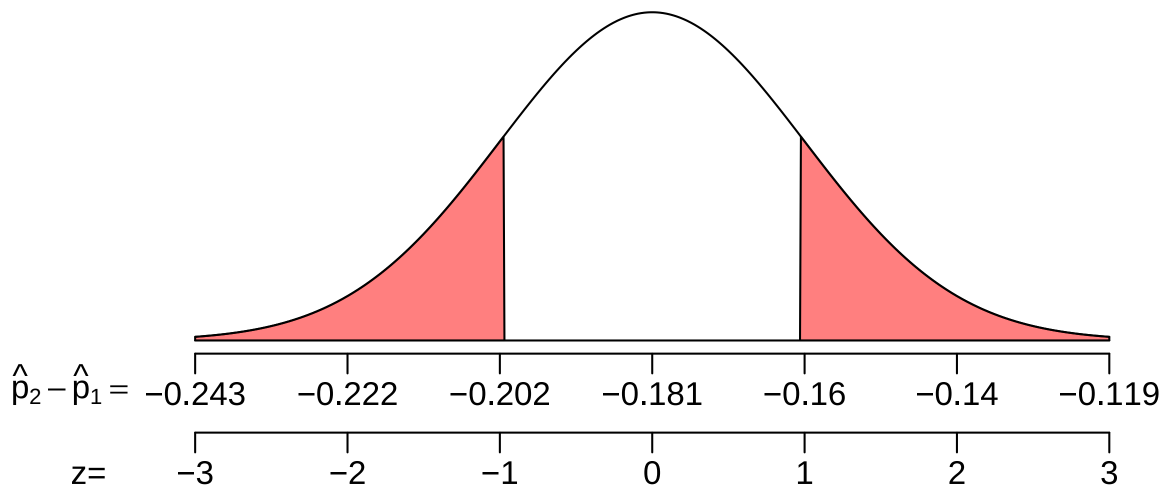
$$\hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(-0.181,\, 0.0207)$$

Determine z scores of boundaries.

$$\begin{aligned} z_{\text{lower}} &= \frac{(\hat{p}_2 - \hat{p}_1)_{\text{lower}} - (p_2 - p_1)}{SE} \\ &= \frac{(-0.201) - (-0.181)}{0.0207} \\ &= -0.97 \end{aligned}$$

$$\begin{aligned} z_{\text{upper}} &= \frac{(\hat{p}_2 - \hat{p}_1)_{\text{upper}} - (p_2 - p_1)}{SE} \\ &= \frac{(-0.161) - (-0.181)}{0.0207} \\ &= 0.97 \end{aligned}$$

Draw a sketch. The phrase "outside the interval (-0.201, -0.161)" suggests finding a two-tail area.



Use a z table.

$$\begin{aligned}
\Pr\left(\hat{P}_2 - \hat{P}_1 < -0.201 \quad \text{OR} \quad \hat{P}_2 - \hat{P}_1 > -0.161\right) &= \Pr(|Z| > 0.97) \\
&= 2 \cdot \Phi(-0.97) \\
&= 0.332
\end{aligned}$$

Thus, we conclude that there is a 33.2% chance that $\hat{P}_2 - \hat{P}_1$ is outside the interval (-0.201, -0.161).

5. **Problem**
In one sample of 200 cases, 71.3% are omnivorous ($\hat{p}_1 = 0.713$). In a second sample of 500 cases, 50.4% are omnivorous ($\hat{p}_2 = 0.504$). Determine a 90% confidence interval of $p_2 - p_1$.
   (a) Determine the lower bound.
   (b) Determine the upper bound.

Solution

Determine the point estimate of $p_2 - p_1$ (our best guess for this population parameter is the corresponding sample statistic).

$$\begin{aligned}\hat{p}_2 - \hat{p}_1 &= 0.504 - 0.713 \\ &= -0.209\end{aligned}$$

Determine the critical $z^\star$ value such that $P(|Z| < z^\star) = 0.9$.

$$z^\star = 1.64$$

Determine the standard error.

$$\begin{aligned}SE &= \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\ &= \sqrt{\frac{(0.713)(0.287)}{200} + \frac{(0.504)(0.496)}{500}} \\ &= 0.039\end{aligned}$$

Determine the lower bound.

$$\begin{aligned}LB &= \text{point estimate} - ME \\ &= (\hat{p}_2 - \hat{p}_1) - z^\star SE \\ &= -0.209 - (1.64)(0.039) \\ &= -0.273\end{aligned}$$

Determine the upper bound.

$$\begin{aligned}UB &= \text{point estimate} + ME \\ &= (\hat{p}_2 - \hat{p}_1) + z^\star SE \\ &= -0.209 + (1.64)(0.039) \\ &= -0.145\end{aligned}$$

We are 90% confident that $p_2 - p_1$ is between -0.273 and -0.145.

(a) The lower bound = -0.273

(b) The upper bound = -0.145

6. **Problem**
   In one sample of 100 cases, 31.2% are reclusive ($\hat{p}_1 = 0.312$). In a second sample of 200 cases, 50.8% are reclusive ($\hat{p}_2 = 0.508$). Determine a 80% confidence interval of $p_2 - p_1$.
   (a) Determine the lower bound.
   (b) Determine the upper bound.

Solution

Determine the point estimate of $p_2 - p_1$ (our best guess for this population parameter is the corresponding sample statistic).

$$\begin{aligned}\hat{p}_2 - \hat{p}_1 &= 0.508 - 0.312\\ &= 0.196\end{aligned}$$

Determine the critical $z^\star$ value such that $P(|Z| < z^\star) = 0.8$.

$$z^\star = 1.28$$

Determine the standard error.

$$\begin{aligned}SE &= \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\\ &= \sqrt{\frac{(0.312)(0.688)}{100} + \frac{(0.508)(0.492)}{200}}\\ &= 0.0583\end{aligned}$$

Determine the lower bound.

$$\begin{aligned}LB &= \text{point estimate} - ME\\ &= (\hat{p}_2 - \hat{p}_1) - z^\star SE\\ &= 0.196 - (1.28)(0.0583)\\ &= 0.121\end{aligned}$$

Determine the upper bound.

$$\begin{aligned}UB &= \text{point estimate} + ME\\ &= (\hat{p}_2 - \hat{p}_1) + z^\star SE\\ &= 0.196 + (1.28)(0.0583)\\ &= 0.271\end{aligned}$$

We are 80% confident that $p_2 - p_1$ is between 0.121 and 0.271.

(a) The lower bound = 0.121

(b) The upper bound = 0.271

7. **Problem**
An experiment is run with a control group of size 28 and a treatment group of size 10. The results are summarized in the table below.

	treatment	control
fluorescent	11	6
not fluorescent	17	4

Using a significance level of 0.2, determine whether the treatment causes an effect on the proportion of cases that are fluorescent.

(a) Determine a *p*-value.

(b) Does the treatment have a significant effect? (yes or no)

Solution
State the hypotheses.

$$H_0: \ p_2 - p_1 = 0$$
$$H_A: \ p_2 - p_1 \neq 0$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{11}{28} = 0.393$$
$$\hat{p}_2 = \frac{6}{10} = 0.6$$

Determine the pooled proportion (because the null assumes the population proportions are equivalent).

$$\hat{p} = \frac{11+6}{28+10} = 0.447$$

Determine the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} \\ &= \sqrt{\frac{(0.447)(0.553)}{28} + \frac{(0.447)(0.553)}{10}} \\ &= 0.183 \end{aligned}$$

Find the z score.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.6 - 0.393) - 0}{0.183} \\ &= 1.13 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-1.13) \\ &= 0.2584 \end{aligned}$$

Compare the p-value to the signficance level.

$$p\text{-value} > \alpha$$

So, we retain the null hypothesis. Thus the difference in proportions is not significant.

(a) The p-value = 0.2584

(b) We retain the null, so no

8. **Problem**
An experiment is run with a control group of size 167 and a treatment group of size 136. The results are summarized in the table below.

	treatment	control
omnivorous	25	38
not omnivorous	142	98

Using a significance level of 0.005, determine whether the treatment causes an effect on the proportion of cases that are omnivorous.

(a) Determine a p-value.

(b) Does the treatment have a significant effect? (yes or no)

Solution
State the hypotheses.

$$H_0: \quad p_2 - p_1 = 0$$
$$H_A: \quad p_2 - p_1 \neq 0$$

Determine the sample proportions.

$$\hat{p}_1 = \frac{25}{167} = 0.15$$
$$\hat{p}_2 = \frac{38}{136} = 0.279$$

Determine the pooled proportion (because the null assumes the population proportions are equivalent).

$$\hat{p} = \frac{25 + 38}{167 + 136} = 0.208$$

Determine the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}$$
$$= \sqrt{\frac{(0.208)(0.792)}{167} + \frac{(0.208)(0.792)}{136}}$$
$$= 0.0469$$

Find the z score.

$$z = \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE}$$
$$= \frac{(0.279 - 0.15) - 0}{0.0469}$$
$$= 2.75$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.75) \\ &= 0.006 \end{aligned}$$

Compare the p-value to the signficance level.

$$p\text{-value} > \alpha$$

So, we retain the null hypothesis. Thus the difference in proportions is not significant.

(a) The p-value = 0.006
(b) We retain the null, so no