1. **Problem**
   In one population, 29.2% are omnivorous ($p_1 = 0.292$). In a second population, 78.3% are omnivorous ($p_2 = 0.783$). When random samples of sizes 5000 and 100 are taken from the first and second populations respectively, what is the chance that $\hat{P}_2 - \hat{P}_1$ is at most 0.497?

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.

$$\begin{aligned} p_2 - p_1 &= 0.783 - 0.292 \\ &= 0.491 \end{aligned}$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\ &= \sqrt{\frac{0.292(1-0.292)}{5000} + \frac{0.783(1-0.783)}{100}} \\ &= 0.0417 \end{aligned}$$

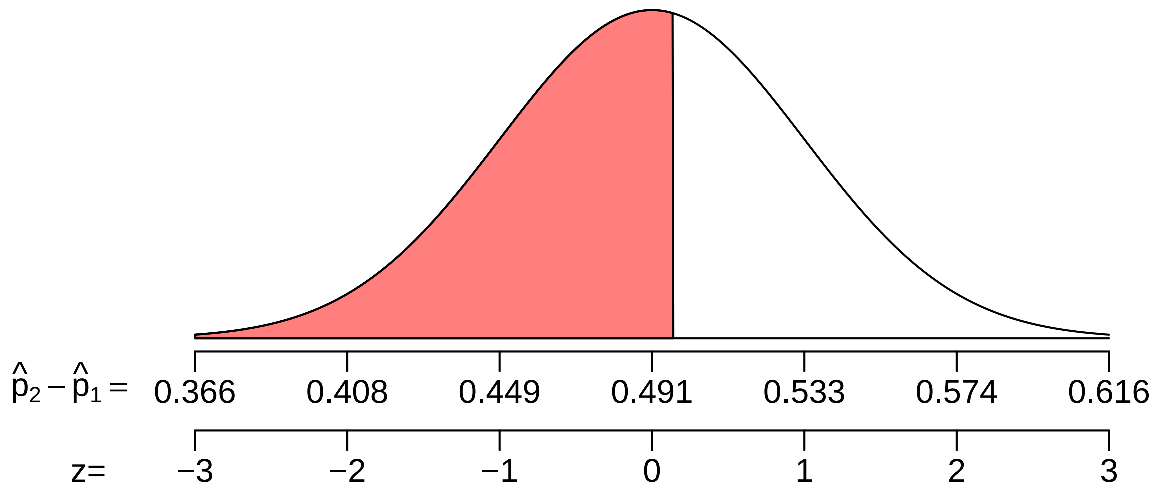
We have the parameters for $\hat{P}_2 - \hat{P}_1$ sampling.

$$\hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0.491,\, 0.0417)$$

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

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)}{SE} \\ &= \frac{(0.497) - (0.491)}{0.0417} \\ &= 0.14 \end{aligned}$$

Draw a sketch. The phrase "at most 0.497" suggests finding a left area.



Use a z table.

$$\begin{aligned}\Pr\left(\hat{P}_2 - \hat{P}_1 < 0.497\right) &= \Pr(Z < 0.14)\\ &= \Phi(0.14)\\ &= 0.5557\end{aligned}$$

Thus, we conclude that there is a 55.57% chance that $\hat{P}_2 - \hat{P}_1$ is at most 0.497.

2. **Problem**
   In one population, 68.9% are abysmal ($p_1 = 0.689$). In a second population, 63.8% are abysmal ($p_2 = 0.638$). When random samples of sizes 1000 and 800 are taken from the first and second populations respectively, what is the chance that $\hat{P}_2 - \hat{P}_1$ is more than -0.04?

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.

$$\begin{aligned} p_2 - p_1 &= 0.638 - 0.689 \\ &= -0.051 \end{aligned}$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\ &= \sqrt{\frac{0.689(1-0.689)}{1000} + \frac{0.638(1-0.638)}{800}} \\ &= 0.0224 \end{aligned}$$

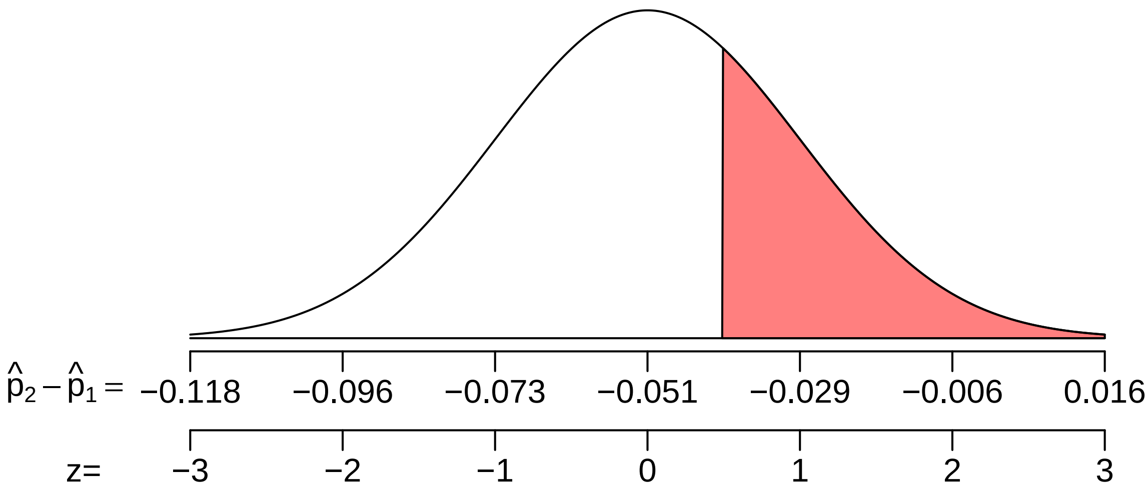
We have the parameters for $\hat{P}_2 - \hat{P}_1$ sampling.

$$\hat{P}_2 - \hat{P}_1 \ \sim \ \mathcal{N}(-0.051,\, 0.0224)$$

Determine a z score for the boundary $\hat{p}_2 - \hat{p}_1 = -0.04$.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)}{SE} \\ &= \frac{(-0.04) - (-0.051)}{0.0224} \\ &= 0.49 \end{aligned}$$

Draw a sketch. The phrase “more than -0.04” suggests finding a right area.



Use a z table.

$$\begin{aligned}\Pr\left(\hat{P}_2 - \hat{P}_1 > -0.04\right) &= \Pr(Z > 0.49) \\ &= 1 - \Phi(0.49) \\ &= 0.3121\end{aligned}$$

Thus, we conclude that there is a 31.21% chance that $\hat{P}_2 - \hat{P}_1$ is more than -0.04.

3. **Problem**
   In one population, 56% are angry ($p_1 = 0.56$). In a second population, 21.2% are angry ($p_2 = 0.212$). When random samples of sizes 2000 and 900 are taken from the first and second populations respectively, what is the chance that $\hat{P}_2 - \hat{P}_1$ is between -0.38 and -0.316?

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.

$$\begin{aligned} p_2 - p_1 &= 0.212 - 0.56 \\ &= -0.348 \end{aligned}$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\ &= \sqrt{\frac{0.56(1-0.56)}{2000} + \frac{0.212(1-0.212)}{900}} \\ &= 0.0176 \end{aligned}$$

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

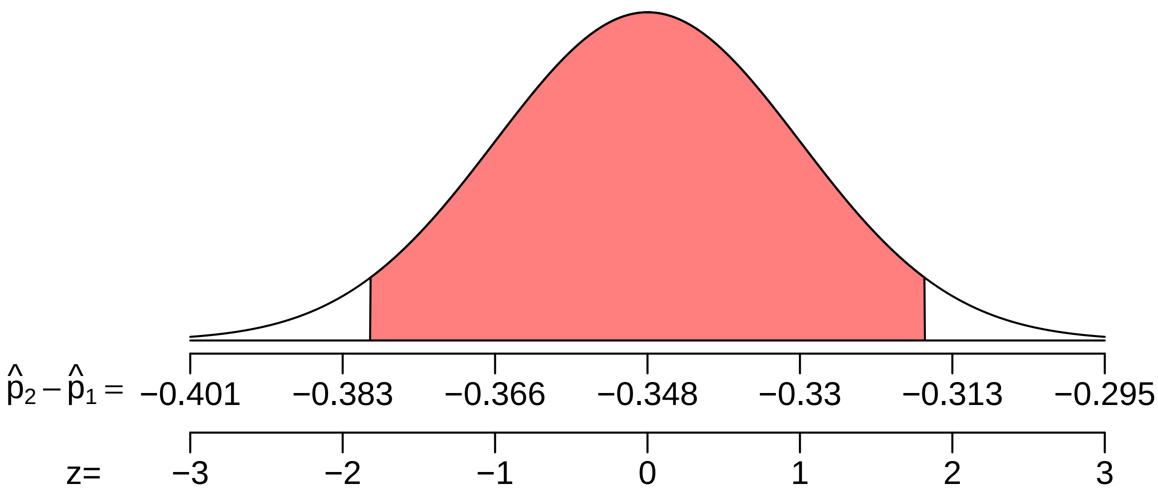
$$\hat{P}_2 - \hat{P}_1 \quad \sim \quad \mathcal{N}(-0.348,\ 0.0176)$$

Determine z scores of the boundaries.

$$\begin{aligned} z_{\text{lower}} &= \frac{(\hat{p}_2 - \hat{p}_1)_{\text{lower}} - (p_2 - p_1)}{SE} \\ &= \frac{(-0.38) - (-0.348)}{0.0176} \\ &= -1.82 \end{aligned}$$

$$\begin{aligned} z_{\text{upper}} &= \frac{(\hat{p}_2 - \hat{p}_1)_{\text{upper}} - (p_2 - p_1)}{SE} \\ &= \frac{(-0.316) - (-0.348)}{0.0176} \\ &= 1.82 \end{aligned}$$

Draw a sketch. The phrase "between -0.38 and -0.316" suggests finding a central area.



Use a z table.

$$\begin{aligned}\Pr\left(-0.38 < \hat{P}_2 - \hat{P}_1 < -0.316\right) &= \Pr(|Z| < 1.82)\\ &= 2 \cdot \Phi(1.82) - 1\\ &= 0.9312\end{aligned}$$

Thus, we conclude that there is a 93.12% chance that $\hat{P}_2 - \hat{P}_1$ is between -0.38 and -0.316.

4. **Problem**
   In one population, 38.6% are preoccupied ($p_1 = 0.386$). In a second population, 27.9% are preoccupied ($p_2 = 0.279$). When random samples of sizes 900 and 9000 are taken from the first and second populations respectively, what is the chance that $\hat{P}_2 - \hat{P}_1$ is outside the interval (-0.143, -0.071)?

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.

$$\begin{aligned} p_2 - p_1 &= 0.279 - 0.386 \\ &= -0.107 \end{aligned}$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\ &= \sqrt{\frac{0.386(1-0.386)}{900} + \frac{0.279(1-0.279)}{9000}} \\ &= 0.0169 \end{aligned}$$

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

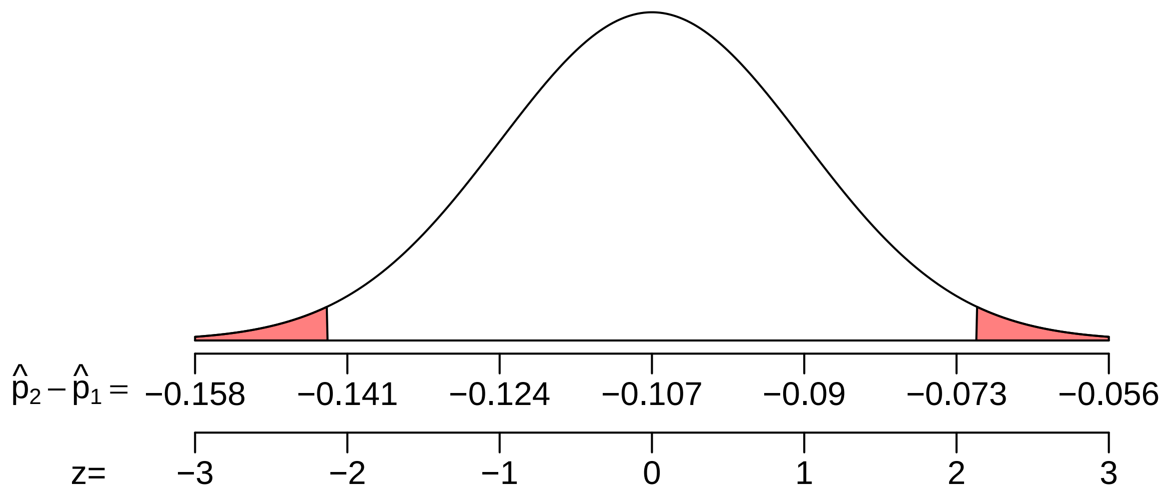
$$\hat{P}_2 - \hat{P}_1 \quad \sim \quad \mathcal{N}(-0.107,\ 0.0169)$$

Determine z scores of boundaries.

$$\begin{aligned} z_{\text{lower}} &= \frac{(\hat{p}_2 - \hat{p}_1)_{\text{lower}} - (p_2 - p_1)}{SE} \\ &= \frac{(-0.143) - (-0.107)}{0.0169} \\ &= -2.13 \end{aligned}$$

$$\begin{aligned} z_{\text{upper}} &= \frac{(\hat{p}_2 - \hat{p}_1)_{\text{upper}} - (p_2 - p_1)}{SE} \\ &= \frac{(-0.071) - (-0.107)}{0.0169} \\ &= 2.13 \end{aligned}$$

Draw a sketch. The phrase “outside the interval (-0.143, -0.071)” suggests finding a two-tail area.



Use a z table.

$$\begin{aligned}\Pr\left(\hat{P}_2 - \hat{P}_1 < -0.143 \quad \text{OR} \quad \hat{P}_2 - \hat{P}_1 > -0.071\right) &= \Pr(|Z| > 2.13) \\ &= 2 \cdot \Phi(-2.13) \\ &= 0.0332\end{aligned}$$

Thus, we conclude that there is a 3.32% chance that $\hat{P}_2 - \hat{P}_1$ is outside the interval (-0.143, -0.071).

5. **Problem**
   In one sample of 40 cases, 38.2% are reclusive ($\hat{p}_1 = 0.382$). In a second sample of 300 cases, 61.1% are reclusive ($\hat{p}_2 = 0.611$). Determine a 95% 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).

$$\begin{aligned}\hat{p}_2 - \hat{p}_1 &= 0.611 - 0.382 \\ &= 0.229\end{aligned}$$

Determine the critical $z^\star$ value such that $P(|Z| < z^\star) = 0.95$.

$$z^\star = 1.96$$

Determine the standard error.

$$\begin{aligned}SE &= \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\ &= \sqrt{\frac{(0.382)(0.618)}{40} + \frac{(0.611)(0.389)}{300}} \\ &= 0.0818\end{aligned}$$

Determine the lower bound.

$$\begin{aligned}LB &= \text{point estimate} - ME \\ &= (\hat{p}_2 - \hat{p}_1) - z^\star SE \\ &= 0.229 - (1.96)(0.0818) \\ &= 0.0687\end{aligned}$$

Determine the upper bound.

$$\begin{aligned}UB &= \text{point estimate} + ME \\ &= (\hat{p}_2 - \hat{p}_1) + z^\star SE \\ &= 0.229 + (1.96)(0.0818) \\ &= 0.389\end{aligned}$$

We are 95% confident that $p_2 - p_1$ is between 0.0687 and 0.389.

(a) The lower bound = 0.0687

(b) The upper bound = 0.389

6. **Problem**
   In one sample of 1000 cases, 11.1% are fluorescent ($\hat{p}_1 = 0.111$). In a second sample of 90 cases, 43.6% are fluorescent ($\hat{p}_2 = 0.436$). 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).

$$\begin{aligned}\hat{p}_2 - \hat{p}_1 &= 0.436 - 0.111\\ &= 0.325\end{aligned}$$

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

$$z^\star = 1.28$$

Determine the standard error.

$$\begin{aligned}SE &= \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\\ &= \sqrt{\frac{(0.111)(0.889)}{1000} + \frac{(0.436)(0.564)}{90}}\\ &= 0.0532\end{aligned}$$

Determine the lower bound.

$$\begin{aligned}LB &= \text{point estimate} - ME\\ &= (\hat{p}_2 - \hat{p}_1) - z^\star SE\\ &= 0.325 - (1.28)(0.0532)\\ &= 0.257\end{aligned}$$

Determine the upper bound.

$$\begin{aligned}UB &= \text{point estimate} + ME\\ &= (\hat{p}_2 - \hat{p}_1) + z^\star SE\\ &= 0.325 + (1.28)(0.0532)\\ &= 0.393\end{aligned}$$

We are 80% confident that $p_2 - p_1$ is between 0.257 and 0.393.

(a) The lower bound = 0.257

(b) The upper bound = 0.393

7. **Problem**
An experiment is run with a control group of size 62 and a treatment group of size 91. The results are summarized in the table below.

	treatment	control
angry	20	45
not angry	42	46

Using a significance level of 0.05, determine whether the treatment causes an effect on the proportion of cases that are angry.

(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.

$$\hat{p}_1 = \frac{20}{62} = 0.323$$
$$\hat{p}_2 = \frac{45}{91} = 0.495$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.495 - 0.323 = 0.172$$

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

$$\hat{p} = \frac{20+45}{62+91} = 0.425$$

Determine the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} \\ &= \sqrt{\frac{(0.425)(0.575)}{62} + \frac{(0.425)(0.575)}{91}} \\ &= 0.0814 \end{aligned}$$

We can be more specific about what the null hypothesis claims.

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0814)$$

We want to describe how unusual our observation is under the null by finding the p-value. To do so, first find the z score.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.495 - 0.323) - 0}{0.0814} \\ &= 2.11 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.11) \\ &= 0.0348 \end{aligned}$$

Compare the p-value to the signficance level.

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

So, we reject the null hypothesis. Thus the difference in proportions is significant.

(a) The p-value = 0.0348

(b) We reject the null, so yes

8. **Problem**
An experiment is run with a control group of size 105 and a treatment group of size 104. The results are summarized in the table below.

	treatment	control
special	63	76
not special	42	28

Using a significance level of 0.04, determine whether the treatment causes an effect on the proportion of cases that are special.

(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.

$$\hat{p}_1 = \frac{63}{105} = 0.6$$
$$\hat{p}_2 = \frac{76}{104} = 0.731$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.731 - 0.6 = 0.131$$

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

$$\hat{p} = \frac{63 + 76}{105 + 104} = 0.665$$

Determine the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}$$
$$= \sqrt{\frac{(0.665)(0.335)}{105} + \frac{(0.665)(0.335)}{104}}$$
$$= 0.0653$$

We can be more specific about what the null hypothesis claims.

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0653)$$

We want to describe how unusual our observation is under the null by finding the p-value. To do so, first find the z score.

$$z = \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE}$$
$$= \frac{(0.731 - 0.6) - 0}{0.0653}$$
$$= 2.01$$

Determine the p-value.

$$p\text{-value} = 2 \cdot \Phi(-|z|)$$
$$= 2 \cdot \Phi(-2.01)$$
$$= 0.0444$$

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 = 0.0444
(b) We retain the null, so no