1. **Problem**
   In one population, 39.6% are omnivorous ($p_1 = 0.396$). In a second population, 24.3% are omnivorous ($p_2 = 0.243$). When random samples of sizes 1000 and 1000 are taken from the first and second populations respectively, what is the chance that $\hat{p}_2 - \hat{p}_1$ is under $-0.17$?

**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:

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

Calculate the expected difference.

$$\begin{aligned} p_2 - p_1 &= 0.243 - 0.396 \\ &= -0.153 \end{aligned}$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}} \\ &= \sqrt{\frac{0.396(1 - 0.396)}{1000} + \frac{0.243(1 - 0.243)}{1000}} \\ &= 0.0206 \end{aligned}$$

Determine a $z$ score.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)}{SE} \\ &= \frac{(-0.17) - (-0.153)}{0.0206} \\ &= -0.83 \end{aligned}$$

Draw a sketch. The phrase “under -0.17” suggests finding a left area.

Use a z table.

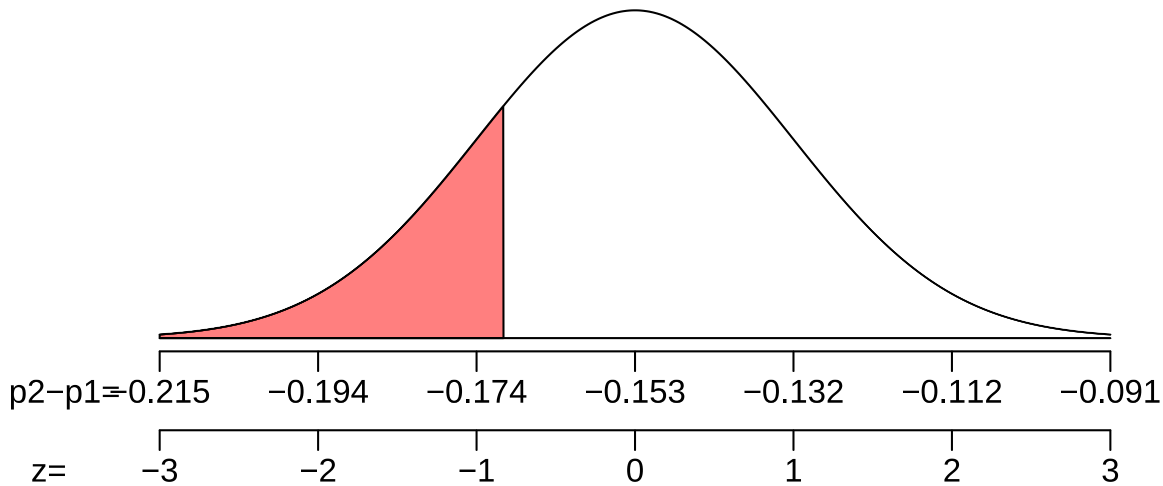


Figure 1:

$$
\begin{aligned}
\Pr\left(\hat{P}_2 - \hat{P}_1 < -0.17\right) &= \Pr(Z < -0.83) \\
&= \Phi(-0.83) \\
&= 0.2033
\end{aligned}
$$

Thus, we conclude that there is a 20.33% chance that $\hat{P}_2 - \hat{P}_1$ is under -0.17.

2. **Problem**
   In one population, 37.6% are happy ($p_1 = 0.376$). In a second population, 23% are happy ($p_2 = 0.23$). When random samples of sizes 6000 and 4000 are taken from the first and second populations respectively, what is the chance that $\hat{p}_2 - \hat{p}_1$ is more than $-0.141$?

**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:

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

Calculate the expected difference.

$$\begin{aligned} p_2 - p_1 &= 0.23 - 0.376 \\ &= -0.146 \end{aligned}$$

Calculate the standard error.

$$SE = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$= \sqrt{\frac{0.376(1-0.376)}{6000} + \frac{0.23(1-0.23)}{4000}}$$

$$= 0.00913$$

Determine a $z$ score.

$$z = \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)}{SE}$$

$$= \frac{(-0.141) - (-0.146)}{0.00913}$$

$$= 0.55$$

Draw a sketch. The phrase “more than -0.141” suggests finding a right area.

Use a z table.

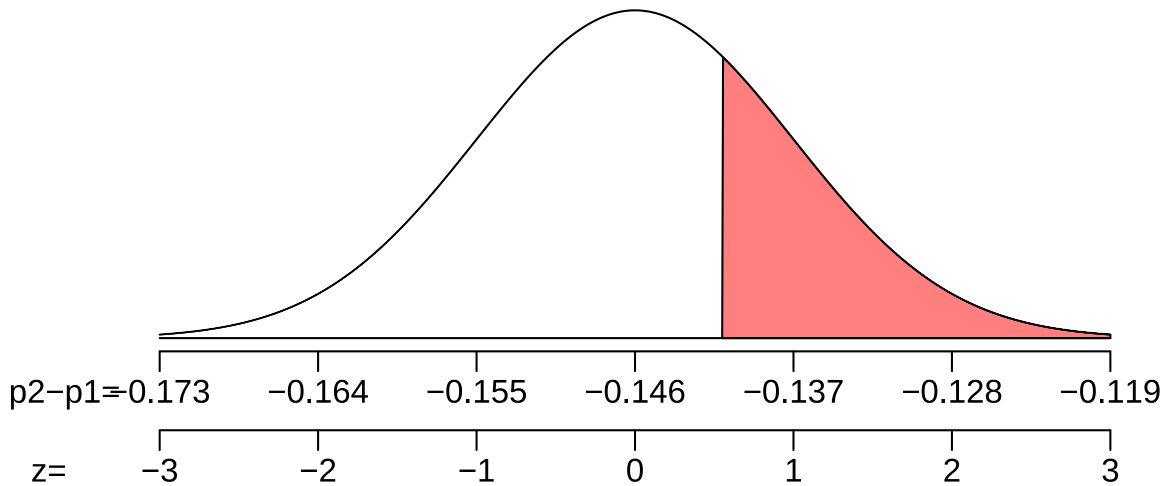


Figure 2:

$$\begin{aligned} \Pr\left(\hat{P}_2 - \hat{P}_1 > -0.141\right) &= \Pr(Z > 0.55) \\ &= 1 - \Phi(0.55) \\ &= 0.2912 \end{aligned}$$

Thus, we conclude that there is a 29.12% chance that $\hat{P}_2 - \hat{P}_1$ is more than -0.141.

3. **Problem**
In one population, 6.2% are sorry ($p_1 = 0.062$). In a second population, 51.3% are sorry ($p_2 = 0.513$). When random samples of sizes 800 and 200 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 0.522?

**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:

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

Calculate the expected difference.

$$\begin{aligned} p_2 - p_1 &= 0.513 - 0.062 \\ &= 0.451 \end{aligned}$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\ &= \sqrt{\frac{0.062(1-0.062)}{800} + \frac{0.513(1-0.513)}{200}} \\ &= 0.0364 \end{aligned}$$

Determine $z$ scores.

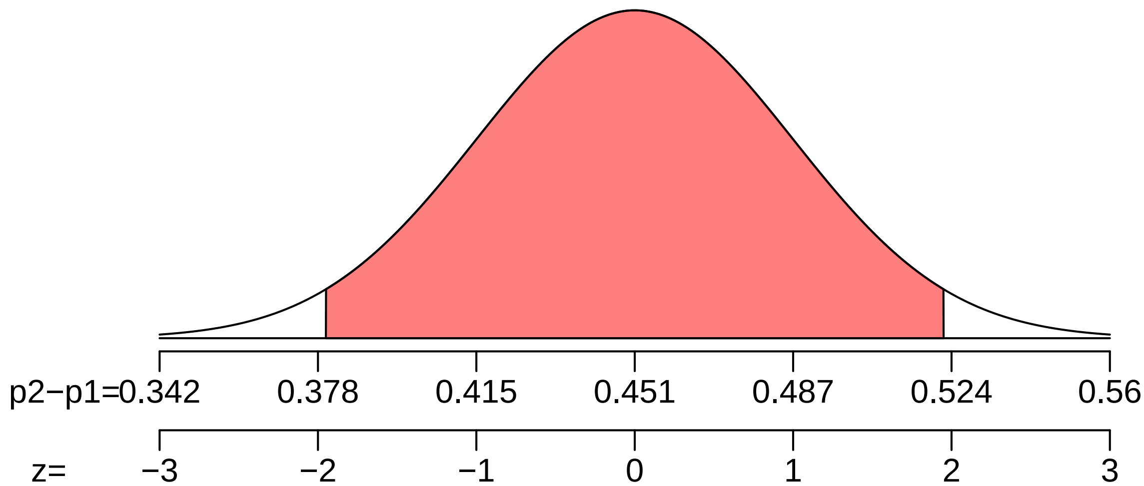


Figure 3:

$$z_{\text{lower}} = \frac{(\hat{p}_2 - \hat{p}_1)_{\text{lower}} - (p_2 - p_1)}{SE}$$

$$= \frac{(0.38) - (0.451)}{0.0364}$$

$$= -1.95$$

$$z_{\text{upper}} = \frac{(\hat{p}_2 - \hat{p}_1)_{\text{upper}} - (p_2 - p_1)}{SE}$$

$$= \frac{(0.522) - (0.451)}{0.0364}$$

$$= 1.95$$

Draw a sketch. The phrase “between 0.38 and 0.522” suggests finding a central area. Use a z table.

$$\begin{aligned} \text{Pr}\left(0.38 < \hat{P}_2 - \hat{P}_1 < 0.522\right) &= \text{Pr}(|Z| < 1.95) \\ &= 2 \cdot \Phi(1.95) - 1 \\ &= 0.9488 \end{aligned}$$

Thus, we conclude that there is a 94.88% chance that $\hat{P}_2 - \hat{P}_1$ is between 0.38 and 0.522.

4. **Problem**
   In one population, 98.7% are happy ($p_1$ = 0.987). In a second population, 90.8% are happy ($p_2$ = 0.908). When random samples of sizes 7000 and 700 are taken from the first and second populations respectively, what is the chance that $\hat{p}_2 - \hat{p}_1$ is outside the interval (-0.095, -0.063)?

**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:

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

Calculate the expected difference.

$$\begin{aligned} p_2 - p_1 &= 0.908 - 0.987 \\ &= -0.079 \end{aligned}$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\ &= \sqrt{\frac{0.987(1-0.987)}{7000} + \frac{0.908(1-0.908)}{700}} \\ &= 0.011 \end{aligned}$$

Determine $z$ scores.

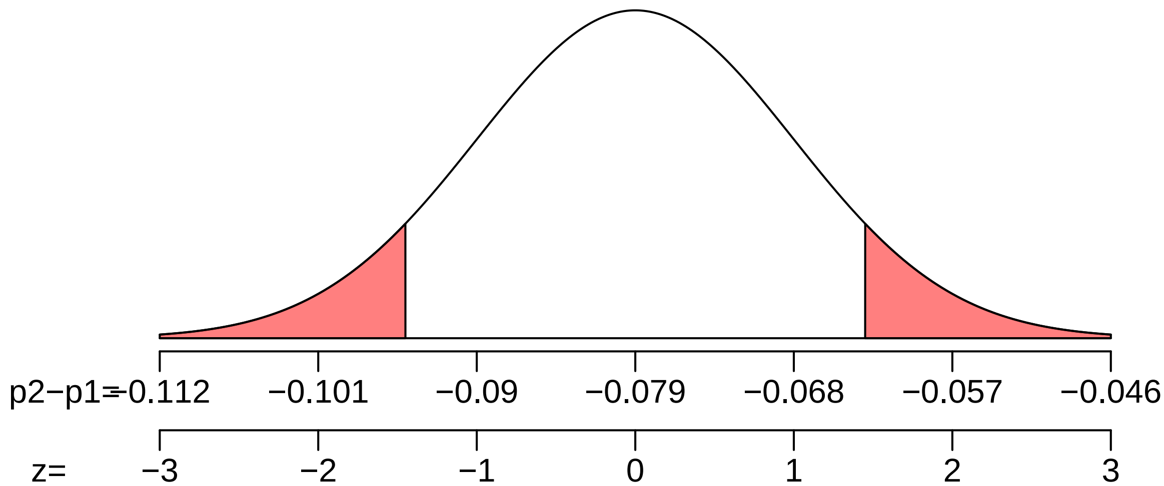


Figure 4:

$$\begin{aligned}
z_{\text{lower}} &= \frac{(\hat{p}_2 - \hat{p}_1)_{\text{lower}} - (p_2 - p_1)}{SE} \\
&= \frac{(-0.095) - (-0.079)}{0.011} \\
&= -1.45 \\
z_{\text{upper}} &= \frac{(\hat{p}_2 - \hat{p}_1)_{\text{upper}} - (p_2 - p_1)}{SE} \\
&= \frac{(-0.063) - (-0.079)}{0.011} \\
&= 1.45
\end{aligned}$$

Draw a sketch. The phrase “outside the interval (-0.095, -0.063)” suggests finding a two-tail area. Use a z table.

$$\begin{aligned}
\Pr\left(\hat{P}_2 - \hat{P}_1 < -0.095 \text{ OR } \hat{P}_2 - \hat{P}_1 > -0.063\right) &= \Pr(|Z| > 1.45) \\
&= 2 \cdot \Phi(-1.45) \\
&= 0.147
\end{aligned}$$

Thus, we conclude that there is a 14.7% chance that $\hat{P}_2 - \hat{P}_1$ is outside the interval (-0.095, -0.063).

5. **Problem**
   In one sample of 400 cases, 16.9% are reclusive ($\hat{p}_1 = 0.169$). In a second sample of 10000 cases, 93.3% are reclusive ($\hat{p}_2 = 0.933$). Determine a 98% 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).

$$\begin{aligned} \hat{p}_2 - \hat{p}_1 &= 0.933 - 0.169 \\ &= 0.764 \end{aligned}$$

Determine the critical $z^\star$ value such that $P(|Z| < z^\star) = 0.98$.

$$z^\star = 2.33$$

Determine the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\ &= \sqrt{\frac{(0.169)(0.831)}{400} + \frac{(0.933)(0.067)}{10000}} \\ &= 0.0189 \end{aligned}$$

Determine the lower bound.

$$\begin{aligned} LB &= \text{point estimate} - ME \\ &= (\hat{p}_2 - \hat{p}_1) - z^\star SE \\ &= 0.764 - (2.33)(0.0189) \\ &= 0.72 \end{aligned}$$

Determine the upper bound.

$$\begin{aligned} UB &= \text{point estimate} + ME \\ &= (\hat{p}_2 - \hat{p}_1) + z^\star SE \\ &= 0.764 + (2.33)(0.0189) \\ &= 0.808 \end{aligned}$$

(a) The lower bound = 0.72

(b) The upper bound = 0.808

6. **Problem**
   In one sample of 8000 cases, 79.8% are fluorescent ($\hat{p}_1 = 0.798$). In a second sample of 200 cases, 54.3% are fluorescent ($\hat{p}_2 = 0.543$). 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).

$$\begin{aligned} \hat{p}_2 - \hat{p}_1 &= 0.543 - 0.798 \\ &= -0.255 \end{aligned}$$

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

$$z^\star = 1.28$$

Determine the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\ &= \sqrt{\frac{(0.798)(0.202)}{8000} + \frac{(0.543)(0.457)}{200}} \\ &= 0.0355 \end{aligned}$$

Determine the lower bound.

$$\begin{aligned} LB &= \text{point estimate} - ME \\ &= (\hat{p}_2 - \hat{p}_1) - z^\star SE \\ &= -0.255 - (1.28)(0.0355) \\ &= -0.3 \end{aligned}$$

Determine the upper bound.

$$\begin{aligned} UB &= \text{point estimate} + ME \\ &= (\hat{p}_2 - \hat{p}_1) + z^\star SE \\ &= -0.255 + (1.28)(0.0355) \\ &= -0.21 \end{aligned}$$

(a) The lower bound = -0.3

(b) The upper bound = -0.21

7. **Problem**
An experiment is run with a control group of size 500 and a treatment group of size 40. The results are summarized in the table below.

| | treatment | control |
|---|---|---|
| angry | 262 | 29 |
| not angry | 238 | 11 |

Using a significance level of 0.01, 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.

$$\hat{p}_1 = \frac{262}{500} = 0.524$$
$$\hat{p}_2 = \frac{29}{40} = 0.725$$

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

$$\hat{p} = \frac{262 + 29}{500 + 40} = 0.539$$

Determine the standard error.

$$SE = \sqrt{\frac{\hat{p}(1 - \hat{p}}{n_1} + \frac{\hat{p}(1 - \hat{p})}{n_2}}$$

$$= \sqrt{\frac{(0.539)(0.461)}{500} + \frac{(0.539)(0.461)}{40}}$$

$$= 0.0819$$

Find the $z$ score.

$$z = \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE}$$

$$= \frac{(0.725 - 0.524) - 0}{0.0819}$$

$$= 2.45$$

Determine the $p$-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.45) \\ &= 0.0142 \end{aligned}$$

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 = 0.0142

(b) We retain the null, so no

8. **Problem**
An experiment is run with a control group of size 1000 and a treatment group of size 40. The results are summarized in the table below.

| | treatment | control |
|---|---|---|
| angry | 292 | 17 |
| not angry | 708 | 23 |

Using a significance level of 0.1, 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.

$$\hat{p}_1 = \frac{292}{1000} = 0.292$$
$$\hat{p}_2 = \frac{17}{40} = 0.425$$

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

$$\hat{p} = \frac{292 + 17}{1000 + 40} = 0.297$$

Determine the standard error.

$$SE = \sqrt{\frac{\hat{p}(1 - \hat{p}}{n_1} + \frac{\hat{p}(1 - \hat{p})}{n_2}}$$

$$= \sqrt{\frac{(0.297)(0.703)}{1000} + \frac{(0.297)(0.703)}{40}}$$

$$= 0.0737$$

Find the $z$ score.

$$z = \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE}$$

$$= \frac{(0.425 - 0.292) - 0}{0.0737}$$

$$= 1.8$$

Determine the $p$-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-1.8) \\ &= 0.0718 \end{aligned}$$

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 = 0.0718

(b) We reject the null, so yes