1. **Problem**
   An experiment is run with a control group of size 157 and a treatment group of size 180. The results are summarized in the table below.

	treatment	control
special	52	77
not special	105	103

   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.

$$\hat{p}_1 = \frac{52}{157} = 0.331$$
$$\hat{p}_2 = \frac{77}{180} = 0.428$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.428 - 0.331 = 0.097$$

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

$$\hat{p} = \frac{52 + 77}{157 + 180} = 0.383$$

Determine the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}$$
$$= \sqrt{\frac{(0.383)(0.617)}{157} + \frac{(0.383)(0.617)}{180}}$$
$$= 0.0531$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0531)$$

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.

$$z = \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE}$$
$$= \frac{(0.428 - 0.331) - 0}{0.0531}$$
$$= 1.83$$

Determine the p-value.

$$p\text{-value} = 2 \cdot \Phi(-|z|)$$
$$= 2 \cdot \Phi(-1.83)$$
$$= 0.0672$$

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 = 0.0672

(b) We retain the null, so no

2. **Problem**

   In one sample of 200 cases, 39% are folksy ($\hat{p}_1 = 0.39$). In a second sample of 50 cases, 40.4% are folksy ($\hat{p}_2 = 0.404$). Determine a 99.5% 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).

$$\begin{aligned}\hat{p}_2 - \hat{p}_1 &= 0.404 - 0.39 \\ &= 0.014\end{aligned}$$

Determine the critical $z^\star$ value such that $P(|Z| < z^\star) = 0.995$.

$$z^\star = 2.81$$

Determine the standard error.

$$\begin{aligned}SE &= \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\ &= \sqrt{\frac{(0.39)(0.61)}{200} + \frac{(0.404)(0.596)}{50}} \\ &= 0.0775\end{aligned}$$

Determine the lower bound.

$$\begin{aligned}LB &= \text{point estimate} - ME \\ &= (\hat{p}_2 - \hat{p}_1) - z^\star SE \\ &= 0.014 - (2.81)(0.0775) \\ &= -0.204\end{aligned}$$

Determine the upper bound.

$$\begin{aligned}UB &= \text{point estimate} + ME \\ &= (\hat{p}_2 - \hat{p}_1) + z^\star SE \\ &= 0.014 + (2.81)(0.0775) \\ &= 0.232\end{aligned}$$

We are 99.5% confident that $p_2 - p_1$ is between -0.204 and 0.232.

(a) The lower bound = -0.204

(b) The upper bound = 0.232

3. **Problem**
   In one population, 64.2% are cold ($p_1 = 0.642$). In a second population, 39.2% are cold ($p_2 = 0.392$). When random samples of sizes 700 and 300 are taken from the first and second populations respectively, what is the chance that $\hat{P}_2 - \hat{P}_1$ is between -0.32 and -0.18?

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.

$$\begin{aligned} p_2 - p_1 &= 0.392 - 0.642 \\ &= -0.25 \end{aligned}$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\ &= \sqrt{\frac{0.642(1-0.642)}{700} + \frac{0.392(1-0.392)}{300}} \\ &= 0.0335 \end{aligned}$$

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

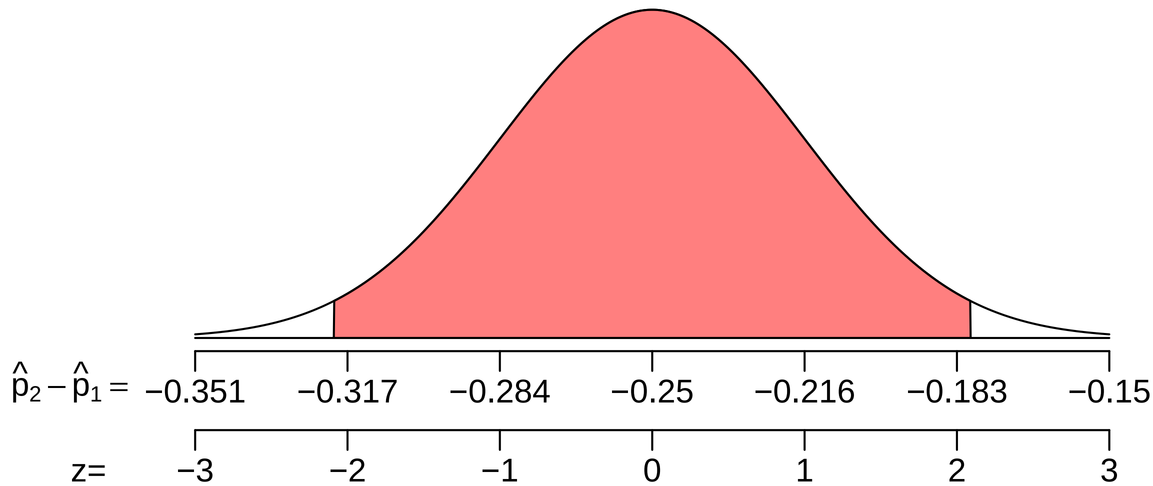
$$\hat{P}_2 - \hat{P}_1 \;\sim\; \mathcal{N}(-0.25,\, 0.0335)$$

Determine z scores of the boundaries.

$$\begin{aligned} z_{\text{lower}} &= \frac{(\hat{p}_2 - \hat{p}_1)_{\text{lower}} - (p_2 - p_1)}{SE} \\ &= \frac{(-0.32) - (-0.25)}{0.0335} \\ &= -2.09 \end{aligned}$$

$$\begin{aligned} z_{\text{upper}} &= \frac{(\hat{p}_2 - \hat{p}_1)_{\text{upper}} - (p_2 - p_1)}{SE} \\ &= \frac{(-0.18) - (-0.25)}{0.0335} \\ &= 2.09 \end{aligned}$$

Draw a sketch. The phrase "between -0.32 and -0.18" suggests finding a central area.



Use a z table.

$$\begin{aligned}\Pr\left(-0.32 < \hat{P}_2 - \hat{P}_1 < -0.18\right) &= \Pr(|Z| < 2.09)\\ &= 2 \cdot \Phi(2.09) - 1\\ &= 0.9634\end{aligned}$$

Thus, we conclude that there is a 96.34% chance that $\hat{P}_2 - \hat{P}_1$ is between -0.32 and -0.18.

4. **Problem**

   It is generally accepted that a population's proportion is 0.802. However, you think that maybe the population proportion is not equal to 0.802, so you decide to run a two-tail hypothesis test with a significance level of 0.1 with a sample size of 1000.

   Then, when you collect the random sample, you find its proportion is 0.82. Do you reject or retain the null hypothesis?

   (a) Determine the p-value.

   (b) Decide whether we reject or retain the null hypothesis.

Solution
State the hypotheses.

$$H_0 : \quad p = 0.802$$
$$H_A : \quad p \neq 0.802$$

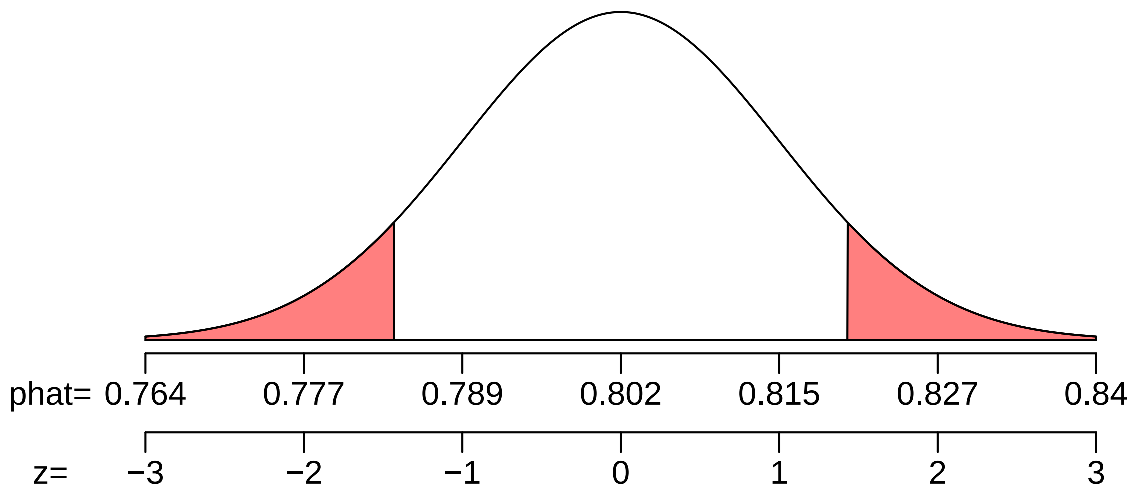
Determine the standard error.

$$SE = \sqrt{\frac{p_0(1 - p_0)}{n}} = \sqrt{\frac{0.802(1 - 0.802)}{1000}} = 0.0126$$

Determine a z score. For simplicity, we ignore the continuity correction.

$$z = \frac{\hat{p} - p_0}{SE} = \frac{0.82 - 0.802}{0.0126} = 1.43$$

The p-value is a two-tail area.



To determine that two-tail area, we use the z table.

$$\Pr\left(\hat{P} > 0.82\right) = 2 \cdot \Phi(-1.43) = 0.1528$$

In other words:

$$p\text{-value} = 0.1528$$

Compare p-value to α (which is 0.1).

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

Make the conclusion: we retain the null hypothesis.

(a) The p-value is 0.1528
(b) We retain the null hypothesis.

5. **Problem**
   If you suspect that $\hat{p}$ will be near 0.33, how large of a sample is needed to guarantee a margin of error less than 0.01 when building a 99.5% confidence interval?

Solution

Determine $z^\star$ such that $P(|Z| < z^\star) = 0.995$.

$$z^\star = 2.81$$

Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.01}{2.81} = 0.00356$$

Calculate n. Because we have no idea what p is, we will use a conservative approach and use $p = 0.5$.

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.33)(0.67)}{(0.00356)^2} = 17445.713925$$

When determining a necessary sample size, always round up (ceiling).

$$n = 17446$$

Also, for simplicity, you should probably only use about 2 significant digits, while still rounding up.

$$n = 18000$$

6. **Problem**
   A random sample of size 1400 was found to have a sample proportion of 5.3%. Determine a 80% confidence interval of the population proportion.

   (a) Find the lower bound of the confidence interval.
   (b) Find the upper bound of the condifence interval.

Solution
Determine $z^{\star}$ such that $P(|Z| < z^{\star}) = 0.8$.

$$z^{\star} = 1.28$$

Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.053)(1-0.053)}{1400}} = 0.00599$$

Calculate the margin of error.

$$ME = z^{\star}SE = (1.28)(0.00599) = 0.00767$$

To find the confidence interval's bounds, find the sample proportion plus or minus the margin of error.

$$p \approx \hat{p} \pm ME$$

Determine the interval.

$$(0.0453\,,\ 0.0607)$$

We are 80% confident that the true population proportion is between 4.53% and 6.07%.

(a) The lower bound = 0.0453, which can also be expressed as 4.53%.

(b) The upper bound = 0.0607, which can also be expressed as 6.07%.

7. **Problem**
   In a very large population, 88.1% are special. When a random sample of size 4200 is taken, what is the chance that the sample proportion of special individuals is farther than $\pm$ 0.5 percentage points from 88.1%?

Solution
Determine the standard error.

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.881(1-0.881)}{4200}} = 0.005$$

Determine the upper and lower bounds on $\hat{p}$.

$$\hat{p}_{\text{lower}} = 0.881 - 0.005 = 0.876$$

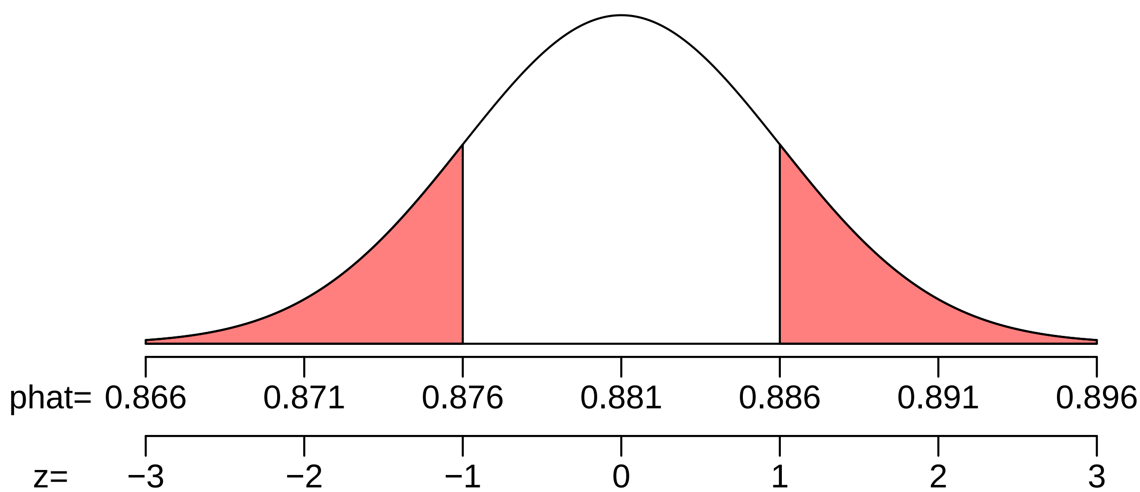
$$\hat{p}_{\text{upper}} = 0.881 + 0.005 = 0.886$$

Determine the z scores. For simplicity, we ignore the continuity correction.

$$z_{\text{lower}} = \frac{\hat{p}_{\text{lower}} - p}{SE} = \frac{0.876 - 0.881}{0.005} = \frac{-0.005}{0.005} = -1$$

$$z_{\text{upper}} = \frac{\hat{p}_{\text{upper}} - p}{SE} = \frac{0.886 - 0.881}{0.005} = \frac{0.005}{0.005} = 1$$

We are looking for a two-tail area ("farther than $\pm$ 0.5 percentage points from 88.1%").



To determine that central area, we use the z table.

$$\Pr\left(|\hat{P} - 0.881| > 0.005\right) = \Pr(|Z| > 1) = 2 \cdot \Phi(-1) = 0.3173$$

Thus, we conclude there is a 31.7% chance that the sample proportion is farther than $\pm$ 0.5 percentage points from 88.1%.