1. **Problem**
An experiment is run with a treatment group of size 221 and a control group of size 270. The results are summarized in the table below.

	treatment	control
cold	112	111
not cold	109	159

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

(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.

$$\hat{p}_1 = \frac{112}{221} = 0.507$$
$$\hat{p}_2 = \frac{111}{270} = 0.411$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.411 - 0.507 = -0.096$$

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

$$\hat{p} = \frac{112 + 111}{221 + 270} = 0.454$$

Determine the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}$$
$$= \sqrt{\frac{(0.454)(0.546)}{221} + \frac{(0.454)(0.546)}{270}}$$
$$= 0.0452$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.0452)$$

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.

$$\begin{aligned} z &= \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE} \\ &= \frac{(0.411 - 0.507) - 0}{0.0452} \\ &= -2.12 \end{aligned}$$

Determine the p-value.

$$\begin{aligned} p\text{-value} &= 2 \cdot \Phi(-|z|) \\ &= 2 \cdot \Phi(-2.12) \\ &= 0.034 \end{aligned}$$

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 = 0.034

(b) We reject the null, so yes

2. **Problem**
In one sample of 100 cases, 40.4% are fluorescent ($\hat{p}_1 = 0.404$). In a second sample of 400 cases, 18.5% are fluorescent ($\hat{p}_2 = 0.185$). 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).

$$\begin{aligned}\hat{p}_2 - \hat{p}_1 &= 0.185 - 0.404 \\ &= -0.219\end{aligned}$$

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

$$z^\star = 2.33$$

Determine the standard error.

$$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$= \sqrt{\frac{(0.404)(0.596)}{100} + \frac{(0.185)(0.815)}{400}}$$

$$= 0.0528$$

Determine the lower bound.

$$\begin{aligned}LB &= \text{point estimate} - ME \\ &= (\hat{p}_2 - \hat{p}_1) - z^\star SE \\ &= -0.219 - (2.33)(0.0528) \\ &= -0.342\end{aligned}$$

Determine the upper bound.

$$\begin{aligned}UB &= \text{point estimate} + ME \\ &= (\hat{p}_2 - \hat{p}_1) + z^\star SE \\ &= -0.219 + (2.33)(0.0528) \\ &= -0.096\end{aligned}$$

We are 98% confident that $p_2 - p_1$ is between -0.342 and -0.096.

   (a) The lower bound = -0.342
   (b) The upper bound = -0.096

3. **Problem**
In one population, 39.2% are sorry ($p_1 = 0.392$). In a second population, 69.5% are sorry ($p_2 = 0.695$). When random samples of sizes 60 and 5000 are taken from the first and second populations respectively, what is the chance that $\hat{P}_2 - \hat{P}_1$ is between 0.29 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.

$$\begin{aligned} p_2 - p_1 &= 0.695 - 0.392 \\ &= 0.303 \end{aligned}$$

Calculate the standard error.

$$\begin{aligned} SE &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\ &= \sqrt{\frac{0.392(1-0.392)}{60} + \frac{0.695(1-0.695)}{5000}} \\ &= 0.0634 \end{aligned}$$

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

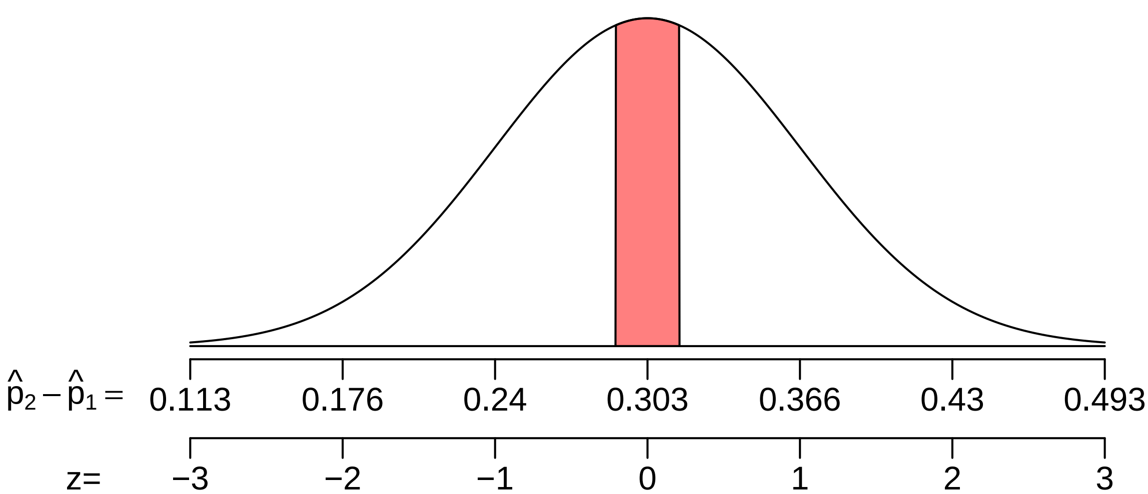
$$\hat{P}_2 - \hat{P}_1 \quad \sim \quad \mathcal{N}(0.303,\ 0.0634)$$

Determine z scores of the boundaries.

$$z_{\text{lower}} = \frac{(\hat{p}_2 - \hat{p}_1)_{\text{lower}} - (p_2 - p_1)}{SE}$$
$$= \frac{(0.29) - (0.303)}{0.0634}$$
$$= -0.21$$

$$z_{\text{upper}} = \frac{(\hat{p}_2 - \hat{p}_1)_{\text{upper}} - (p_2 - p_1)}{SE}$$
$$= \frac{(0.316) - (0.303)}{0.0634}$$
$$= 0.21$$

Draw a sketch. The phrase "between 0.29 and 0.316" suggests finding a central area.



Use a z table.

$$\begin{aligned}\Pr\left(0.29 < \hat{P}_2 - \hat{P}_1 < 0.316\right) &= \Pr(|Z| < 0.21)\\ &= 2 \cdot \Phi(0.21) - 1\\ &= 0.1664\end{aligned}$$

Thus, we conclude that there is a 16.64% chance that $\hat{P}_2 - \hat{P}_1$ is between 0.29 and 0.316.

4. **Problem**
It is generally accepted that a population's proportion is 0.393. However, you think that maybe the population proportion is not 0.393, so you decide to run a two-tail hypothesis test with a significance level of 0.02 with a sample size of 4000.

Then, when you collect the random sample, you find its proportion is 0.413. 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.

$$H_0: \quad p = 0.393$$
$$H_A: \quad p \neq 0.393$$

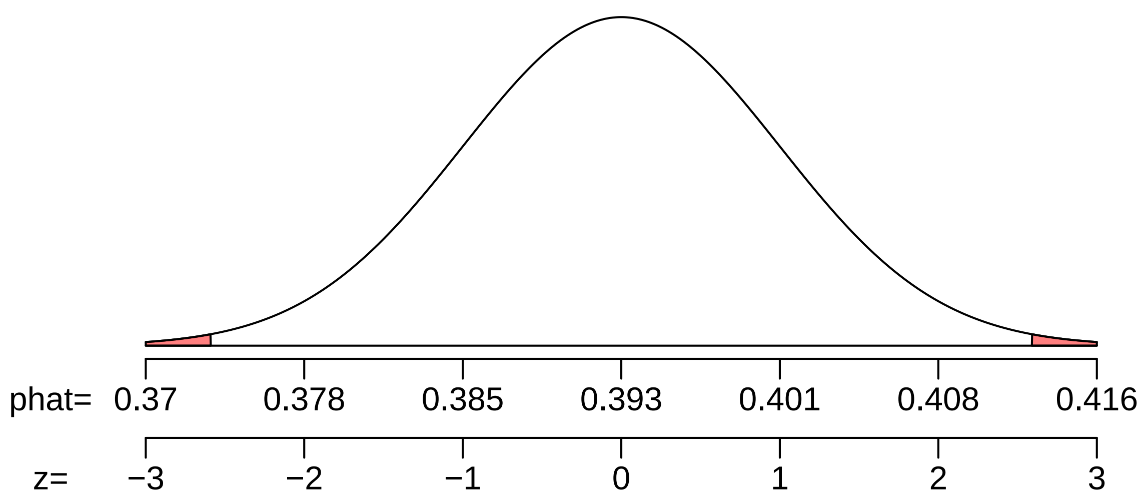
Determine the standard error.

$$SE = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.393(1-0.393)}{4000}} = 0.00772$$

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

$$z = \frac{\hat{p} - p_0}{SE} = \frac{0.413 - 0.393}{0.00772} = 2.59$$

The p-value is a two-tail area.



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

$$\Pr\left(\hat{P} > 0.413\right) = 2 \cdot \Phi(-2.59) = 0.0096$$

In other words:

$$p\text{-value} = 0.0096$$

Compare p-value to α (which is 0.02).

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

Make the conclusion: we reject the null hypothesis.

(a) The p-value is 0.0096

(b) We reject the null hypothesis.

5. **Problem**
If you suspect that $\hat{p}$ will be near 0.76, how large of a sample is needed to guarantee a margin of error less than 0.04 when building a 99% confidence interval?

Solution
Determine $z^\star$ such that $P(|Z| < z^\star) = 0.99$.

$$z^\star = 2.58$$

Determine the maximal standard error.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.04}{2.58} = 0.0155$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.76)(0.24)}{(0.0155)^2} = 759.2091571$$

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

$$n = 760$$

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

$$n = 760$$

6. **Problem**
   A random sample of size 66000 was found to have a sample proportion of 6.4%. Determine a 90% 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 $P(|Z| < z^\star) = 0.9$.

$$z^\star = 1.64$$

   Calculate the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.064)(1-0.064)}{66000}} = 0.000953$$

   Calculate the margin of error.

$$ME = z^\star SE = (1.64)(0.000953) = 0.00156$$

   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.

$$(0.0624\,,\ 0.0656)$$

   We are 90% confident that the true population proportion is between 6.24% and 6.56%.

   (a) The lower bound = 0.0624, which can also be expressed as 6.24%.
   (b) The upper bound = 0.0656, which can also be expressed as 6.56%.

7. **Problem**
In a very large population, 76.8% are happy. When a random sample of size 1600 is taken, what is the chance that the sample proportion of happy individuals is beyond $\pm$ 0.8 percentage points from 76.8%?

Solution
Determine the standard error.

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.768(1-0.768)}{1600}} = 0.011$$

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

$$\hat{p}_{\text{lower}} = 0.768 - 0.008 = 0.76$$

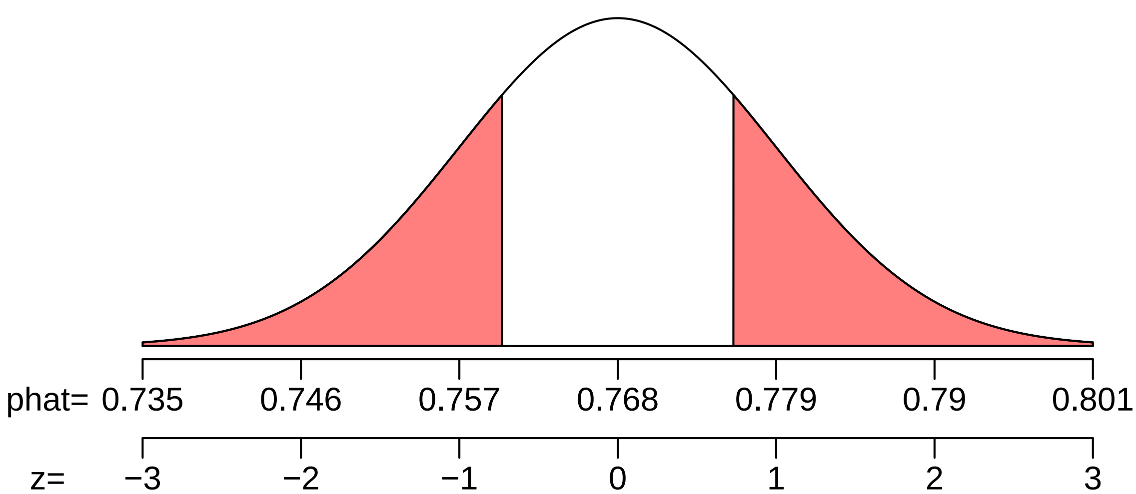
$$\hat{p}_{\text{upper}} = 0.768 + 0.008 = 0.776$$

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

$$z_{\text{lower}} = \frac{\hat{p}_{\text{lower}} - p}{SE} = \frac{0.76 - 0.768}{0.011} = \frac{-0.008}{0.011} = -0.73$$

$$z_{\text{upper}} = \frac{\hat{p}_{\text{upper}} - p}{SE} = \frac{0.776 - 0.768}{0.011} = \frac{0.008}{0.011} = 0.73$$

We are looking for a two-tail area (“beyond $\pm$ 0.8 percentage points from 76.8%”).



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

$$\Pr\left(|\hat{P} - 0.768| > 0.008\right) \quad = \quad \Pr(|Z| > 0.73) \quad = \quad 2 \cdot \Phi(-0.73) \quad = \quad 0.4654$$

Thus, we conclude there is a 46.5% chance that the sample proportion is beyond $\pm$ 0.8 percentage points from 76.8%.