1. **Problem**
   A roughly symmetric population has a mean $\mu = 200$ and standard deviation $\sigma = 59$. What is the probability that a sample of size $n = 167$ has a mean above 209.55?

Solution

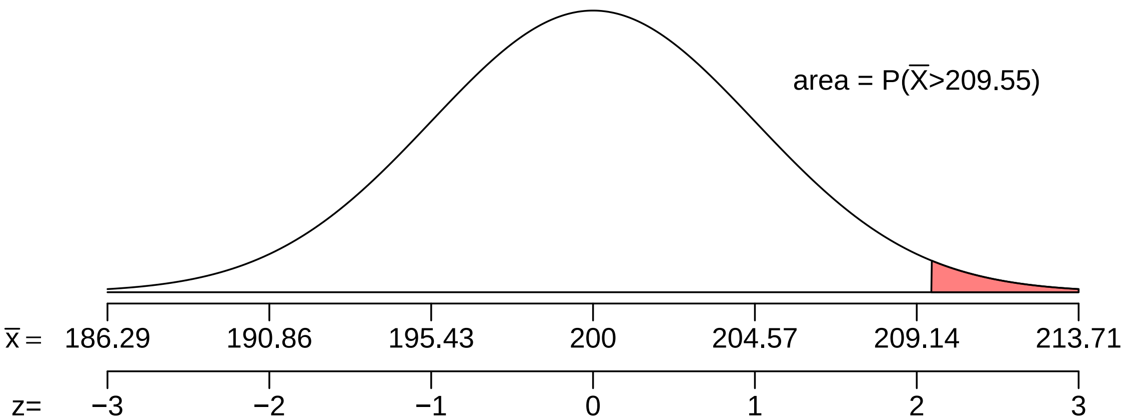
Find the standard error.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{59}{\sqrt{167}} = 4.57$$

Describe the sampling distribution.

$$\bar{X} \sim \mathcal{N}(200, 4.57)$$

Draw a sketch.



Calculate a z score.

$$z = \frac{209.55 - 200}{4.57} = 2.09$$

Determine the probability.

$$P(\bar{X} > 209.55) = 0.018$$

2. **Problem**
   A random sample has mean 39.3 and standard deviation 10.6. Assuming the population is large and roughly symmetric (so we can use our methods of inference), determine a 90% confidence level.

   (a) Determine the lower bound.

   (b) Determine the upper bound.

Solution
We are given the sample size, sample mean, sample standard deviation, and confidence level.

$$\begin{aligned} n &= 39 \\ \bar{x} &= 39.3 \\ s &= 10.6 \\ CL &= 0.9 \end{aligned}$$

Determine the degrees of freedom (because we don't know σ and we are doing inference so we need to use the t distribution).

$$df = n - 1 = 38$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.9$.

$$t^\star = 1.69$$

Calculate the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{10.6}{\sqrt{39}} = 1.7$$

We want to make an inference about the population mean.

$$\mu \approx \bar{x} \pm t^\star SE$$

Determine the bounds.

$$\begin{aligned} CI &= (\bar{x} - t^\star SE,\ \bar{x} + t^\star SE) \\ &= (39.3 - 1.69 \times 1.7,\ 39.3 - 1.69 \times 1.7) \\ &= (36.4,\ 42.2) \end{aligned}$$

We are 90% confident that the population mean is between 36.4 and 42.2.

(a) Lower bound = 36.4

(b) Upper bound = 42.2

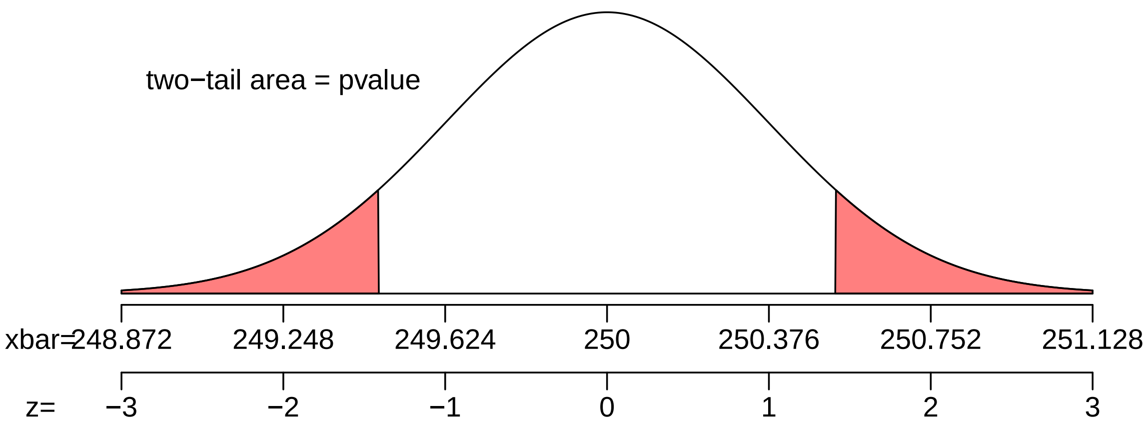
3. **Problem**
   A null hypothesis claims a roughly symmetric population has a mean $\mu = 250$ and a standard deviation $\sigma = 4$. Determine the *p*-value of a two-tail test if your sample of size $n = 113$ has mean $\bar{x} = 250.53$.

Solution

Find the standard error.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{113}} = 0.376$$

Make a sketch.



Find the z score.

$$z = \frac{250.53 - 250}{0.376} = 1.41$$

Find the p-value.

$$p\text{-value} = 2 \cdot \Phi(-1.41) = 0.159$$

4. **Problem**
   A null hypothesis claims a population has a mean 52. You decide to perform a two-tail hypothesis test with significance level 0.05. Your (roughly symmetric) sample of size 7 has mean 63.2 and standard deviation 10.7. Should we reject or retain the null hypothesis?

Solution

We are given sample size, sample mean, sample standard deviation, significance level, and null mean.

$$\begin{aligned} n &= 7 \\ \bar{x} &= 63.2 \\ s &= 10.7 \\ \alpha &= 0.05 \\ \mu_0 &= 52 \end{aligned}$$

State the hypotheses.

$$\begin{aligned} H_0 : &\quad \mu = 52 \\ H_A : &\quad \mu \neq 52 \end{aligned}$$

Find the degrees of freedom (because we don't know σ and we are doing inference so we need to use the t distribution).

$$df = n - 1 = 6$$

Determine a critical value $t^\star$ such that $P(|T| > t^\star) = 0.05$.

$$t^\star = 2.45$$

Find the standard error.

$$SE = \frac{s}{\sqrt{n}} = \frac{10.7}{\sqrt{7}} = 4.04$$

Find the test statistic from the observed mean.

$$t_{\text{obs}} = \frac{63.2 - 52}{4.04} = 2.77$$

Compare $|t_{\text{obs}}|$ and $t^\star$ to make a conclusion.

$$|t_{\text{obs}}| > t^\star$$

Thus, we reject the null hypothesis.

5. **Problem**
An experiment has $n_1 = 5$ plants in the treatment group and $n_2 = 6$ plants in the control group. After some time, the plants' heights (in cm) are measured, resulting in the following data:

	value1	value2	value3	value4	value5	value6
sample 1:	1.41	1.74	1.38	1.5	1.81	
sample 2:	1.09	1.23	1.42	0.81	0.55	0.54

(a) Determine degrees of freedom.
(b) Determine $t^{\star}$ for a 98% confidence interval.
(c) Determine SE.
(d) Determine a lower bound of the 98% confidence interval of $\mu_2 - \mu_1$.
(e) Determine an upper bound of the 98% confidence interval of $\mu_2 - \mu_1$.
(f) Determine $|t_{\text{obs}}|$ under the null hypothesis $\mu_2 - \mu_1 = 0$.
(g) Determine a lower bound of the two-tail p-value.
(h) Determine an upper bound of two-tail p-value.
(i) Do you reject the null hypothesis with a two-tail test using a significance level $\alpha = 0.02$? (yes or no)

Solution
These data are unpaired. We might as well find the sample means and sample standard deviations (use a calculator's built-in function for standard deviation).

$$\overline{x_1} = 1.57$$
$$\overline{x_2} = 0.94$$
$$s_1 = 0.196$$
$$s_2 = 0.365$$

We make a conservative estimate of the degrees of freedom using the appropriate formula.

$$df = \min(n_1,\ n_2) - 1 = \min(5, 6) - 1 = 4$$

We use the t table to find $t^{\star}$ such that $P(|T| < t^{\star}) = 0.98$

$$t^{\star} = 3.75$$

We use the SE formula for unpaired data.

$$SE = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} = \sqrt{\frac{(0.196)^2}{5} + \frac{(0.365)^2}{6}} = 0.173$$

We find the bounds of the confidence interval.

$$CI = (\overline{x_2} - \overline{x_1}) \pm t^{\star} SE$$
$$CI = (-1.279,\ 0.019)$$

We find t_{obs}.

$$t_{\text{obs}} = \frac{(\overline{x_2} - \overline{x_1}) - (\mu_2 - \mu_1)_0}{SE} = \frac{(0.94 - 1.57) - 0}{0.173} = -3.64$$

We find $|t_{\text{obs}}|$.

$$|t_{\text{obs}}| = 3.64$$

We use the table to determine bounds on p-value. Remember, $df = 4$ and $p\text{-value} = P(|T| > |t_{\text{obs}}|)$.

$$0.02 < p\text{-value} < 0.04$$

We should consider both comparisons to make our decision.

$$|t_{\text{obs}}| < t^{\star}$$
$$p\text{-value} > \alpha$$

Thus, we retain the null hypothesis. Also notice the confidence interval does contain 0.

(a) 4
(b) 3.75
(c) 0.173
(d) -1.279
(e) 0.019
(f) 3.644
(g) 0.02
(h) 0.04
(i) no

6. **Problem**
   If you suspect that $\hat{p}$ will be near 0.23, how large of a sample is needed to guarantee a margin of error less than 0.02 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.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.02}{2.81} = 0.00712$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.23)(0.77)}{(0.00712)^2} = 3493.4825148$$

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

$$n = 3494$$

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

$$n = 3500$$

7. **Problem**
   How large of a sample is needed to guarantee a margin of error less than 0.03 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.

$$ME = z^\star SE$$

$$SE = \frac{ME}{z^\star} = \frac{0.03}{2.58} = 0.0116$$

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

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

$$n = \frac{p(1-p)}{SE^2} = \frac{(0.5)(0.5)}{(0.0116)^2} = 1857.9072533$$

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

$$n = 1858$$

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

$$n = 1900$$

8. **Problem**
   In one sample of 300 cases, 8.9% are special ($\hat{p}_1 = 0.089$). In a second sample of 300 cases, 43.7% are special ($\hat{p}_2 = 0.437$). Determine a 99% 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).

$$\begin{aligned}\hat{p}_2 - \hat{p}_1 &= 0.437 - 0.089\\ &= 0.348\end{aligned}$$

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

$$z^\star = 2.58$$

Determine the standard error.

$$\begin{aligned}SE &= \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\\ &= \sqrt{\frac{(0.089)(0.911)}{300} + \frac{(0.437)(0.563)}{300}}\\ &= 0.033\end{aligned}$$

Determine the lower bound.

$$\begin{aligned}LB &= \text{point estimate} - ME\\ &= (\hat{p}_2 - \hat{p}_1) - z^\star SE\\ &= 0.348 - (2.58)(0.033)\\ &= 0.263\end{aligned}$$

Determine the upper bound.

$$\begin{aligned}UB &= \text{point estimate} + ME\\ &= (\hat{p}_2 - \hat{p}_1) + z^\star SE\\ &= 0.348 + (2.58)(0.033)\\ &= 0.433\end{aligned}$$

We are 99% confident that $p_2 - p_1$ is between 0.263 and 0.433.

(a) The lower bound = 0.263
(b) The upper bound = 0.433

9. **Problem**

An experiment is run with a treatment group of size 285 and a control group of size 304. The results are summarized in the table below.

	treatment	control
sick	170	158
not sick	115	146

Using a significance level of 0.1, determine whether the treatment causes an effect on the proportion of cases that are sick.

(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.

$$\hat{p}_1 = \frac{170}{285} = 0.596$$
$$\hat{p}_2 = \frac{158}{304} = 0.52$$

Determine the difference of sample proportions.

$$\hat{p}_2 - \hat{p}_1 = 0.52 - 0.596 = -0.076$$

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

$$\hat{p} = \frac{170 + 158}{285 + 304} = 0.557$$

Determine the standard error.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}$$
$$= \sqrt{\frac{(0.557)(0.443)}{285} + \frac{(0.557)(0.443)}{304}}$$
$$= 0.041$$

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

$$H_0: \quad \hat{P}_2 - \hat{P}_1 \sim \mathcal{N}(0,\ 0.041)$$

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.

$$z = \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)_0}{SE}$$
$$= \frac{(0.52 - 0.596) - 0}{0.041}$$
$$= -1.85$$

Determine the p-value.

$$p\text{-value} = 2 \cdot \Phi(-|z|)$$
$$= 2 \cdot \Phi(-1.85)$$
$$= 0.0644$$

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 = 0.0644
(b) We reject the null, so yes