

Chapter 7

Inference for numerical data¹

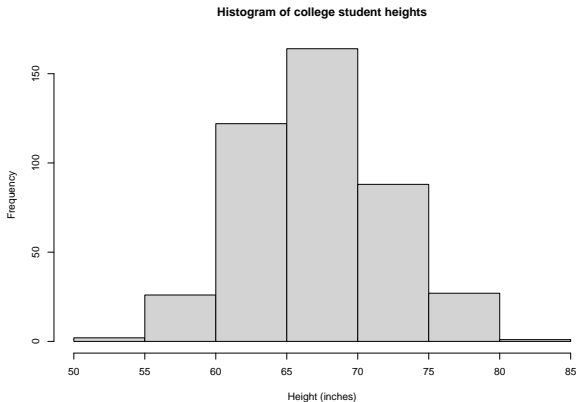
Department of Mathematics & Statistics
North Carolina A&T State University

¹These notes use content from OpenIntro Statistics Slides by Mine Cetinkaya-Rundel.

One-sample means with the t distribution

Heights

- ▶ According to the CDC, the mean height of U.S. adults ages 20 and older is about 66.5 inches (69.3 inches for males, and 63.8 inches for females).
- ▶ In our sample data, we have a sample of 430 college students from a single college.



Summary statistics

n	\bar{x}	s	minimum	maximum
430	67.09	4.86	53.78	83.21

Objective: We would like to investigate if the mean height of students at this college is significantly different than 66.5 inches.

Conditions

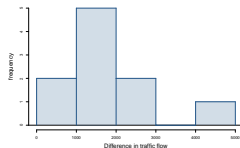
- ▶ **Independence:** We are told to assume that cases (rows) are independent.

Conditions

- ▶ **Independence:** We are told to assume that cases (rows) are independent.
- ▶ **Sample size / skew:**

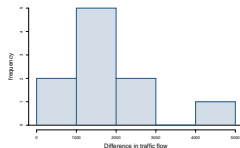
Conditions

- ▶ **Independence:** We are told to assume that cases (rows) are independent.
- ▶ **Sample size / skew:**
- ▶ The sample distribution does not appear to be extremely skewed, but it's very difficult to assess with such a small sample size. We might want to think about whether we would expect the population distribution to be skewed or not — probably not, it should be equally likely to have days with lower than average traffic and higher than average traffic.
- ▶ We do not know σ and n is too small to assume s is a reliable estimate for σ .



Conditions

- ▶ **Independence:** We are told to assume that cases (rows) are independent.
- ▶ **Sample size / skew:**
- ▶ The sample distribution does not appear to be extremely skewed, but it's very difficult to assess with such a small sample size. We might want to think about whether we would expect the population distribution to be skewed or not — probably not, it should be equally likely to have days with lower than average traffic and higher than average traffic.
- ▶ We do not know σ and n is too small to assume s is a reliable estimate for σ .



So what do we do when the sample size is small?

Review: what purpose does a large sample serve?

As long as observations are independent, and the population distribution is not extremely skewed, a large sample would ensure that...

- ▶ The sampling distribution of the mean is nearly normal.
- ▶ The estimate of the standard error, as $\frac{s}{\sqrt{n}}$, is reliable.

The normality condition

- ▶ The CLT, which states that sampling distributions will be nearly normal, holds true for **any** sample size as long as the population distribution is nearly normal.

The normality condition

- ▶ The CLT, which states that sampling distributions will be nearly normal, holds true for **any** sample size as long as the population distribution is nearly normal.
- ▶ While this is helpful special case, it's inherently difficult to verify normality in small data sets.

The normality condition

- ▶ The CLT, which states that sampling distributions will be nearly normal, holds true for **any** sample size as long as the population distribution is nearly normal.
- ▶ While this is helpful special case, it's inherently difficult to verify normality in small data sets.
- ▶ We should exercise caution when verifying the normality condition for small samples. It is important to not only examine the data but also think about where the data come from.
 - ▶ For example, ask: would I expect this distribution to be symmetric, and am I confident that outliers are rare?

The t distribution

- ▶ When the population standard deviation is unknown (almost always), the uncertainty of the standard error estimate is addressed by using a new distribution: the **t distribution**.

The t distribution

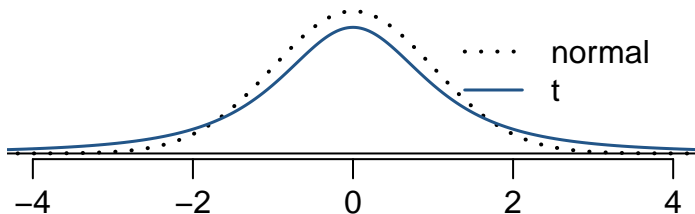
- ▶ When the population standard deviation is unknown (almost always), the uncertainty of the standard error estimate is addressed by using a new distribution: the **t distribution**.
- ▶ This distribution also has a bell shape, but its tails are **thicker** than the normal model's.

The t distribution

- ▶ When the population standard deviation is unknown (almost always), the uncertainty of the standard error estimate is addressed by using a new distribution: the **t distribution**.
- ▶ This distribution also has a bell shape, but its tails are **thicker** than the normal model's.
- ▶ Therefore observations are more likely to fall beyond two SDs from the mean than under the normal distribution.

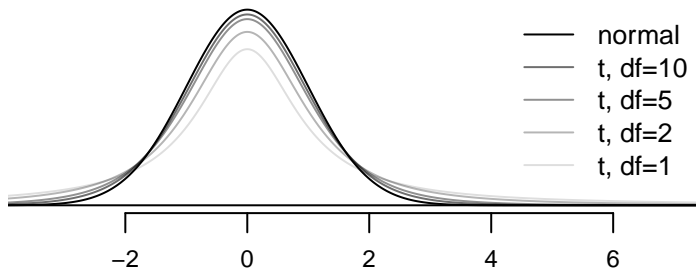
The t distribution

- ▶ When the population standard deviation is unknown (almost always), the uncertainty of the standard error estimate is addressed by using a new distribution: the **t distribution**.
- ▶ This distribution also has a bell shape, but its tails are **thicker** than the normal model's.
- ▶ Therefore observations are more likely to fall beyond two SDs from the mean than under the normal distribution.
- ▶ Extra thick tails are helpful for resolving our problem with a less reliable estimate the standard error (since n is small).



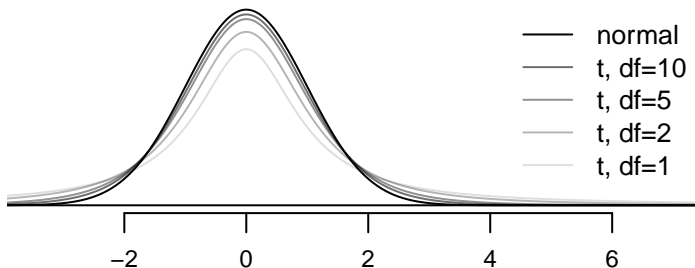
The t distribution

- ▶ Always centered at zero, like the standard normal (z) distribution.
- ▶ Has a single parameter: **degrees of freedom (df)**.



The t distribution

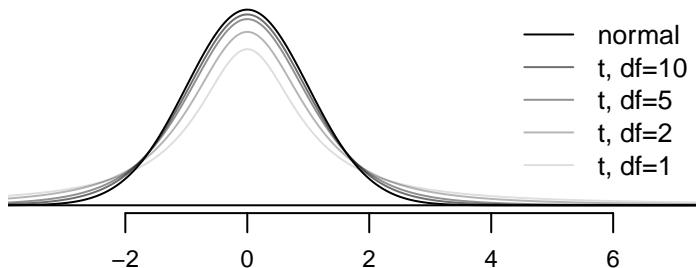
- ▶ Always centered at zero, like the standard normal (z) distribution.
- ▶ Has a single parameter: **degrees of freedom** (df).



What happens to shape of the t distribution as df increases?

The t distribution

- ▶ Always centered at zero, like the standard normal (z) distribution.
- ▶ Has a single parameter: **degrees of freedom** (df).



What happens to shape of the t distribution as df increases?

Approaches normal.

Back to the student heights survey

n	\bar{x}	s	minimum	maximum
430	67.09	4.86	53.78	83.21

Objective: We would like to investigate if the mean height of students at this college is significantly different than 66.5 inches.

Hypotheses

What are the hypotheses for testing for the mean of college student heights being different from 67 inches?

A) $H_0 : \mu = 66.5$

$H_A : \mu \neq 66.5$

B) $H_0 : \mu = 66.5$

$H_A : \mu > 66.5$

C) $H_0 : \mu = 66.5$

$H_A : \mu < 66.5$

D) $H_0 : \mu \neq 66.5$

$H_A : \mu > 66.5$

Hypotheses

What are the hypotheses for testing for the mean of college student heights being different from 66.5 inches?

A) $H_0 : \mu = 66.5$

$H_A : \mu \neq 66.5$

B) $H_0 : \mu = 66.5$

$H_A : \mu > 66.5$

C) $H_0 : \mu = 66.5$

$H_A : \mu < 66.5$

D) $H_0 : \mu \neq 66.5$

$H_A : \mu > 66.5$

Finding the test statistic

The test statistic for inference on sample mean is the T statistic with $df = n - 1$.

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

Finding the test statistic

The test statistic for inference on sample mean is the T statistic with $df = n - 1$.

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

in context...

$$\text{point estimate} = \bar{x} = 67.09$$

Finding the test statistic

The test statistic for inference on sample mean is the T statistic with $df = n - 1$.

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

in context...

$$\text{point estimate} = \bar{x} = 67.09$$

$$SE = \frac{s}{\sqrt{n}} = \frac{4.86}{\sqrt{430}} = 0.234$$

Finding the test statistic

The test statistic for inference on sample mean is the T statistic with $df = n - 1$.

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

in context...

$$\text{point estimate} = \bar{x} = 67.09$$

$$SE = \frac{s}{\sqrt{n}} = \frac{4.86}{\sqrt{430}} = 0.234$$

$$T = \frac{67.09 - 66.5}{0.234} = 2.52$$

Finding the test statistic

The test statistic for inference on sample mean is the T statistic with $df = n - 1$.

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

in context...

$$\text{point estimate} = \bar{x} = 67.09$$

$$SE = \frac{s}{\sqrt{n}} = \frac{4.86}{\sqrt{430}} = 0.234$$

$$T = \frac{67.09 - 66.5}{0.234} = 2.52$$

$$df = 430 - 1 = 429$$

Note: Null value is 66.5 because in the null hypothesis we set $\mu = 66.5$

Finding the p-value

- ▶ The p-value is, once again, calculated as the area tail area under the t distribution.

Finding the p-value

- ▶ The p-value is, once again, calculated as the area tail area under the t distribution.
- ▶ Using R:

```
2 * pt(2.52, df = 429, lower.tail = FALSE)
```

```
## [1] 0.0120975
```

Finding the p-value

- ▶ The p-value is, once again, calculated as the area tail area under the t distribution.
- ▶ Using R:

```
2 * pt(2.52, df = 429, lower.tail = FALSE)
```

```
## [1] 0.0120975
```

- ▶ Using a web app:
https://gallery.shinyapps.io/dist_calc/

Finding the p-value

- ▶ The p-value is, once again, calculated as the area tail area under the t distribution.
- ▶ Using R:

```
2 * pt(2.52, df = 429, lower.tail = FALSE)
```

```
## [1] 0.0120975
```

- ▶ Using a web app:
https://gallery.shinyapps.io/dist_calc/
- ▶ Or when these aren't available, we can use a t -table.

Conclusion of the test

What is the conclusion of this hypothesis test?

Conclusion of the test

What is the conclusion of this hypothesis test?

We saw that the p-value was extremely low. Thus, we reject the null hypothesis. Based on the p-value, we conclude that the survey provide strong evidence that the mean of the college students height is different from the mean height of U.S. adults over 20.

What is the difference?

- ▶ We concluded that there is a difference in the mean heights of the college students compared to the mean height of U.S. adults

What is the difference?

- ▶ We concluded that there is a difference in the mean heights of the college students compared to the mean height of U.S. adults
- ▶ But it would be more interesting to find out what exactly this difference is.

What is the difference?

- ▶ We concluded that there is a difference in the mean heights of the college students compared to the mean height of U.S. adults
- ▶ But it would be more interesting to find out what exactly this difference is.
- ▶ We can use a confidence interval to estimate this difference.

Confidence interval for a sample mean

- Confidence intervals are always of the form

$$\text{point estimate} \pm ME$$

Confidence interval for a sample mean

- ▶ Confidence intervals are always of the form

$$\text{point estimate} \pm ME$$

- ▶ ME is always calculated as the product of a critical value and SE.

Confidence interval for a sample mean

- ▶ Confidence intervals are always of the form

$$\text{point estimate} \pm ME$$

- ▶ ME is always calculated as the product of a critical value and SE.

- ▶ $ME = t^* \times SE$

$$\text{point estimate} \pm t^* \times SE$$

Finding the critical $t(t^*)$

- ▶ We want to find the 95% confidence interval.
- ▶ Using R:

```
qt(p = (1+0.95)/2, df = 429)
```

```
## [1] 1.965509
```


Constructing a CI for a small sample mean

Which of the following is the correct calculation of a 95% confidence interval for the heights of the college students?

$$\bar{x} = 67.09 \quad s = 4.86 \quad n = 430 \quad SE = 0.234$$

- A) $66.5 \pm 1.96 \times 0.234$
- B) $67.09 \pm 1.97 \times 0.234$
- C) $67.09 \pm -2.26 \times 0.234$
- D) $66.5 \pm 2.26 \times 4.86$

Constructing a CI for a small sample mean

Which of the following is the correct calculation of a 95% confidence interval for the heights of the college students?

$$\bar{x} = 67.09 \quad s = 4.86 \quad n = 430 \quad SE = 0.234$$

- A) $66.5 \pm 1.96 \times 0.234$
- B) $67.09 \pm 1.97 \times 0.234 \rightarrow (66.63, 67.55)$
- C) $67.09 \pm -2.26 \times 0.234$
- D) $66.5 \pm 2.26 \times 4.86$

Synthesis

Does the conclusion from the hypothesis test agree with the findings of the confidence interval?

Synthesis

Does the conclusion from the hypothesis test agree with the findings of the confidence interval?

Yes, the hypothesis test found a significant difference, and the CI does not contain the null value of 66.5.

Recap: Inference using the t -distribution

- ▶ If σ is unknown, use t -distribution with $SE = \frac{s}{\sqrt{n}}$.

Recap: Inference using the t -distribution

- ▶ If σ is unknown, use t -distribution with $SE = \frac{s}{\sqrt{n}}$.
- ▶ Conditions:
 - ▶ Independence of observations (often verified by random sample, and if sampling w/o replacement, $n < 10\%$ of population).
 - ▶ No extreme skew.

Recap: Inference using the t -distribution

- ▶ If σ is unknown, use t -distribution with $SE = \frac{s}{\sqrt{n}}$.
- ▶ Conditions:
 - ▶ Independence of observations (often verified by random sample, and if sampling w/o replacement, $n < 10\%$ of population).
 - ▶ No extreme skew.
- ▶ Hypothesis testing:

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}, \text{ where } df = n - 1$$

Recap: Inference using the t -distribution

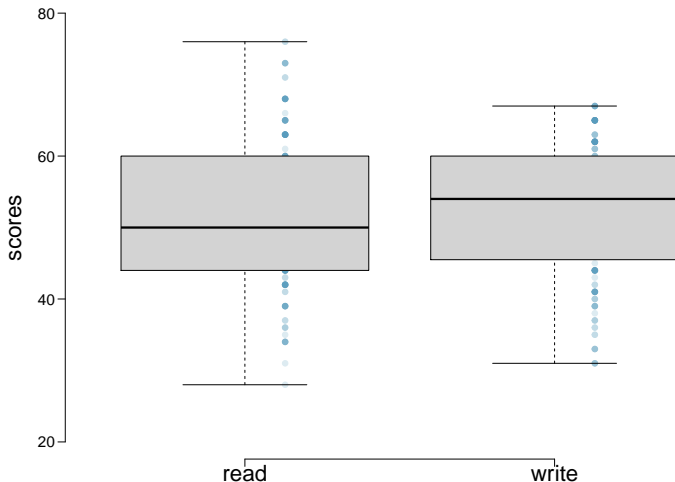
- ▶ If σ is unknown, use t -distribution with $SE = \frac{s}{\sqrt{n}}$.
- ▶ Conditions:
 - ▶ Independence of observations (often verified by random sample, and if sampling w/o replacement, $n < 10\%$ of population).
 - ▶ No extreme skew.
- ▶ Hypothesis testing:

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}, \text{ where } df = n - 1$$

- ▶ Confidence interval: $\text{point estimate} \pm t_{df}^* \times SE$

Paired data

200 observations were randomly sampled from the High School and Beyond survey. The same students took a reading and writing test and their scores are shown below. At a first glance, does there appear to be a difference between the average reading and writing test score?



Practice

The same students took a reading and writing test and their scores are shown below. Are the reading and writing scores of each student independent of each other?

	id	read	write
1	70	57	52
2	86	44	33
3	141	63	44
4	172	47	52
⋮	⋮	⋮	⋮
200	137	63	65

A) Yes

B) No

Practice

The same students took a reading and writing test and their scores are shown below. Are the reading and writing scores of each student independent of each other?

	id	read	write
1	70	57	52
2	86	44	33
3	141	63	44
4	172	47	52
⋮	⋮	⋮	⋮
200	137	63	65

A) Yes

B) No

Analyzed paired data

- ▶ When two sets of observations have this special correspondence (not independent), they are said to be **paired**.

Analyzed paired data

- ▶ When two sets of observations have this special correspondence (not independent), they are said to be **paired**.
- ▶ To analyze paired data, it is often useful to look at the difference in outcomes of each pair of observations.

$$\text{diff} = \text{read} - \text{write}$$

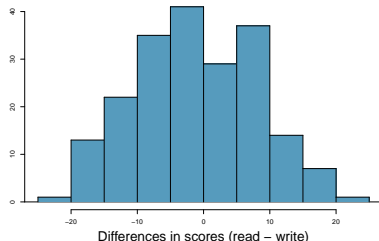
Analyzed paired data

- ▶ When two sets of observations have this special correspondence (not independent), they are said to be **paired**.
- ▶ To analyze paired data, it is often useful to look at the difference in outcomes of each pair of observations.

$$\text{diff} = \text{read} - \text{write}$$

- ▶ It is important that we always subtract using a consistent order.

	id	read	write	diff
	1	70	57	52
	2	86	44	33
	3	141	63	44
	4	172	47	52
	⋮	⋮	⋮	⋮
	200	137	65	-2



Parameter and point estimate

- ▶ **Parameter of interest:** Average difference between the reading and writing scores of **all** high school students.

$$\mu_{diff}$$

Parameter and point estimate

- ▶ **Parameter of interest:** Average difference between the reading and writing scores of **all** high school students.

$$\mu_{diff}$$

- ▶ **Point estimate:** Average difference between the reading and writing scores of **sampled** high school students.

$$\bar{x}_{diff}$$

Setting the hypotheses

If in fact there was no difference between the scores on the reading and writing exams, what would you expect the average difference to be?

Setting the hypotheses

If in fact there was no difference between the scores on the reading and writing exams, what would you expect the average difference to be?

0

Setting the hypotheses

If in fact there was no difference between the scores on the reading and writing exams, what would you expect the average difference to be?

0

What are the hypotheses for testing if there is a difference between the average reading and writing scores?

Setting the hypotheses

If in fact there was no difference between the scores on the reading and writing exams, what would you expect the average difference to be?

0

What are the hypotheses for testing if there is a difference between the average reading and writing scores?

H_0 : There is no difference between the average reading and writing score.

$$\mu_{diff} = 0$$

H_A : There is a difference between the average reading and writing score.

$$\mu_{diff} \neq 0$$

Nothing new here

- ▶ The analysis is no different than what we have done before.
- ▶ We have data from **one** sample: differences.
- ▶ We are testing to see if average difference is different than 0.

Checking assumptions & conditions

Which of the following is true?

- A) Since students are sampled randomly and are less than 10% of all high school students, we can assume that the difference between the reading and writing scores of one student in the sample is independent of another.
- B) The distribution of differences is bimodal, therefore we cannot continue with the hypothesis test.
- C) In order for differences to be random we should have sampled with replacement.
- D) Since students are sampled randomly and are less than 10% all students, we can assume that the sampling distribution of the average difference will be nearly normal.

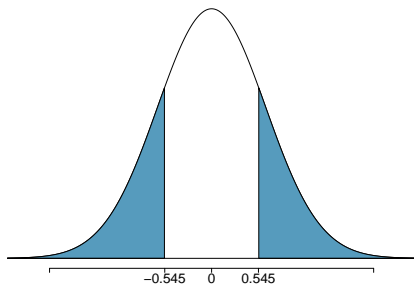
Checking assumptions & conditions

Which of the following is true?

- A) Since students are sampled randomly and are less than 10% of all high school students, we can assume that the difference between the reading and writing scores of one student in the sample is independent of another.
- B) The distribution of differences is bimodal, therefore we cannot continue with the hypothesis test.
- C) In order for differences to be random we should have sampled with replacement.
- D) Since students are sampled randomly and are less than 10% all students, we can assume that the sampling distribution of the average difference will be nearly normal.

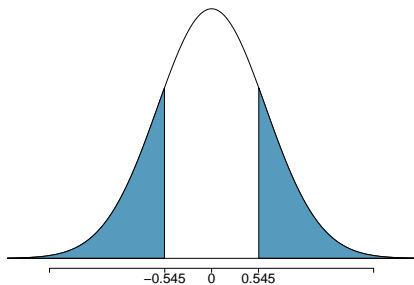
Calculating the test-statistic and the p-value

The observed average difference between the two scores is -0.545 points and the standard deviation of the difference is 8.887 points. Do these data provide convincing evidence of a difference between the average scores on the two exams? Use $\alpha = 0.05$.



Calculating the test-statistic and the p-value

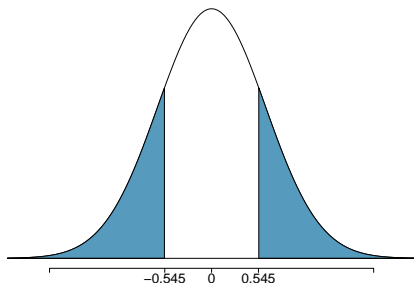
The observed average difference between the two scores is -0.545 points and the standard deviation of the difference is 8.887 points. Do these data provide convincing evidence of a difference between the average scores on the two exams? Use $\alpha = 0.05$.



$$\begin{aligned} T &= \frac{-0.545 - 0}{\frac{8.887}{\sqrt{200}}} \\ &= \frac{-0.545}{0.628} = -0.87 \\ df &= 200 - 1 = 199 \end{aligned}$$

Calculating the test-statistic and the p-value

The observed average difference between the two scores is -0.545 points and the standard deviation of the difference is 8.887 points. Do these data provide convincing evidence of a difference between the average scores on the two exams? Use $\alpha = 0.05$.



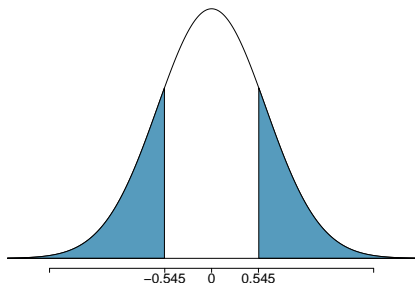
$$\begin{aligned} T &= \frac{-0.545 - 0}{\frac{8.887}{\sqrt{200}}} \\ &= \frac{-0.545}{0.628} = -0.87 \end{aligned}$$

$$df = 200 - 1 = 199$$

$$p\text{-value} = 0.1927 \times 2 = 0.3854$$

Calculating the test-statistic and the p-value

The observed average difference between the two scores is -0.545 points and the standard deviation of the difference is 8.887 points. Do these data provide convincing evidence of a difference between the average scores on the two exams? Use $\alpha = 0.05$.



$$\begin{aligned} T &= \frac{-0.545 - 0}{\frac{8.887}{\sqrt{200}}} \\ &= \frac{-0.545}{0.628} = -0.87 \end{aligned}$$

$$df = 200 - 1 = 199$$

$$p\text{-value} = 0.1927 \times 2 = 0.3854$$

Since $p\text{-value} > 0.05$, fail to reject, the data do not provide convincing evidence of a difference between the average reading and writing scores.

Interpretation of p-value

Which of the following is the correct interpretation of the p-value?

- A) Probability that the average scores on the reading and writing exams are equal.
- B) Probability that the average scores on the reading and writing exams are different.
- C) Probability of obtaining a random sample of 200 students where the average difference between the reading and writing scores is at least 0.545 (in either direction), if in fact the true average difference between the scores is 0.
- D) Probability of incorrectly rejecting the null hypothesis if in fact the null hypothesis is true.

Interpretation of p-value

Which of the following is the correct interpretation of the p-value?

- A) Probability that the average scores on the reading and writing exams are equal.
- B) Probability that the average scores on the reading and writing exams are different.
- C) Probability of obtaining a random sample of 200 students where the average difference between the reading and writing scores is at least 0.545 (in either direction), if in fact the true average difference between the scores is 0.
- D) Probability of incorrectly rejecting the null hypothesis if in fact the null hypothesis is true.

HT \leftrightarrow CI

Suppose we were to construct 95% confidence interval for the average difference between the reading and writing scores. Would you expect this interval to include 0?

- A) Yes
- B) No
- C) Cannot tell from the information given.

HT \leftrightarrow CI

Suppose we were to construct 95% confidence interval for the average difference between the reading and writing scores. Would you expect this interval to include 0?

- A) Yes
- B) No
- C) Cannot tell from the information given.

$$\begin{aligned} -0.545 \pm 1.97 \frac{8.887}{\sqrt{200}} &= -0.545 \pm 1.97 \times 0.628 \\ &= -0.545 \pm 1.24 \\ &= (-1.785, 0.695) \end{aligned}$$

Friday the 13th

Between 1990 - 1992 researchers in the UK collected data on traffic flow, accidents, hospital admissions on Friday 13th and the previous Friday, Friday 6th. Below is an excerpt from this data set on traffic flow. We can assume that traffic flow on given day at locations 1 and 2 are independent.

	type	date	6 th	13 th	diff	location
1	traffic	1990, July	139246	138548	698	loc 1
2	traffic	1990, July	134012	132908	1104	loc 2
3	traffic	1991, September	137055	136018	1037	loc 1
4	traffic	1991, September	133732	131843	1889	loc 2
5	traffic	1991, December	123552	121641	1911	loc 1
6	traffic	1991, December	121139	118723	2416	loc 2
7	traffic	1992, March	128293	125532	2761	loc 1
8	traffic	1992, March	124631	120249	4382	loc 2
9	traffic	1992, November	124609	122770	1839	loc 1
10	traffic	1992, November	117584	117263	321	loc 2

Friday the 13th

- ▶ We want to investigate if people's behavior is different on Friday 13th compared to Friday 6th.

Friday the 13th

- ▶ We want to investigate if people's behavior is different on Friday 13th compared to Friday 6th.
- ▶ One approach is to compare the traffic flow on these two days.

Friday the 13th

- ▶ We want to investigate if people's behavior is different on Friday 13th compared to Friday 6th.
- ▶ One approach is to compare the traffic flow on these two days.
- ▶ We want to investigate if people's behavior is different on Friday 13th compared to Friday 6th.

Friday the 13th

- ▶ We want to investigate if people's behavior is different on Friday 13th compared to Friday 6th.
- ▶ One approach is to compare the traffic flow on these two days.
- ▶ We want to investigate if people's behavior is different on Friday 13th compared to Friday 6th.
- ▶ H_0 : Average traffic flow on Friday 6th and 13th are equal.
 H_A : Average traffic flow on Friday 6th and 13th are different.

Friday the 13th

- ▶ We want to investigate if people's behavior is different on Friday 13th compared to Friday 6th.
- ▶ One approach is to compare the traffic flow on these two days.
- ▶ We want to investigate if people's behavior is different on Friday 13th compared to Friday 6th.
- ▶ H_0 : Average traffic flow on Friday 6th and 13th are equal.
 H_A : Average traffic flow on Friday 6th and 13th are different.

Each case in the data set represents traffic flow recorded at the same location in the same month of the same year: one count from Friday 6th and the other Friday 13th. Are these two counts independent?

Friday the 13th

- ▶ We want to investigate if people's behavior is different on Friday 13th compared to Friday 6th.
- ▶ One approach is to compare the traffic flow on these two days.
- ▶ We want to investigate if people's behavior is different on Friday 13th compared to Friday 6th.
- ▶ H_0 : Average traffic flow on Friday 6th and 13th are equal.
 H_A : Average traffic flow on Friday 6th and 13th are different.

Each case in the data set represents traffic flow recorded at the same location in the same month of the same year: one count from Friday 6th and the other Friday 13th. Are these two counts independent?

No.

Hypotheses

What are the hypotheses for testing for a difference between the average traffic flow between Friday 6th and 13th?

A) $H_0 : \mu_{6th} = \mu_{13th}$

$$H_A : \mu_{6th} \neq \mu_{13th}$$

B) $H_0 : p_{6th} = p_{13th}$

$$H_A : p_{6th} \neq p_{13th}$$

C) $H_0 : \mu_{diff} = 0$

$$H_A : \mu_{diff} \neq 0$$

D) $H_0 : \bar{x}_{diff} = 0$

$$H_A : \bar{x}_{diff} \neq 0$$

Hypotheses

What are the hypotheses for testing for a difference between the average traffic flow between Friday 6th and 13th?

A) $H_0 : \mu_{6th} = \mu_{13th}$

$$H_A : \mu_{6th} \neq \mu_{13th}$$

B) $H_0 : p_{6th} = p_{13th}$

$$H_A : p_{6th} \neq p_{13th}$$

C) $H_0 : \mu_{diff} = 0$

$$H_A : \mu_{diff} \neq 0$$

D) $H_0 : \bar{x}_{diff} = 0$

$$H_A : \bar{x}_{diff} \neq 0$$

Friday the 13th

	type	date	6 th	13 th	diff	location
1	traffic	1990, July	139246	138548	698	loc 1
2	traffic	1990, July	134012	132908	1104	loc 2
3	traffic	1991, September	137055	136018	1037	loc 1
4	traffic	1991, September	133732	131843	1889	loc 2
5	traffic	1991, December	123552	121641	1911	loc 1
6	traffic	1991, December	121139	118723	2416	loc 2
7	traffic	1992, March	128293	125532	2761	loc 1
8	traffic	1992, March	124631	120249	4382	loc 2
9	traffic	1992, November	124609	122770	1839	loc 1
10	traffic	1992, November	117584	117263	321	loc 2

$$\bar{x}_{diff} = 1836$$

$$s_{diff} = 1176$$

$$n = 10$$

Finding the test statistic

The test statistic for inference on a small sample ($n < 50$) mean is the T statistic with $df = n - 1$.

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

Finding the test statistic

The test statistic for inference on a small sample ($n < 50$) mean is the T statistic with $df = n - 1$.

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

in context...

$$\text{point estimate} = \bar{x}_{diff} = 1836$$

Finding the test statistic

The test statistic for inference on a small sample ($n < 50$) mean is the T statistic with $df = n - 1$.

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

in context...

$$\text{point estimate} = \bar{x}_{diff} = 1836$$

$$SE = \frac{s_{diff}}{\sqrt{n}} = \frac{1176}{\sqrt{10}} = 372$$

Finding the test statistic

The test statistic for inference on a small sample ($n < 50$) mean is the T statistic with $df = n - 1$.

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

in context...

$$\text{point estimate} = \bar{x}_{diff} = 1836$$

$$SE = \frac{s_{diff}}{\sqrt{n}} = \frac{1176}{\sqrt{10}} = 372$$

$$T = \frac{1836 - 0}{372} = 4.94$$

Finding the test statistic

The test statistic for inference on a small sample ($n < 50$) mean is the T statistic with $df = n - 1$.

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

in context...

$$\text{point estimate} = \bar{x}_{diff} = 1836$$

$$SE = \frac{s_{diff}}{\sqrt{n}} = \frac{1176}{\sqrt{10}} = 372$$

$$T = \frac{1836 - 0}{372} = 4.94$$

$$df = 10 - 1 = 9$$

Note: Null value is 0 because in the null hypothesis we set

$\mu_{diff} = 0$

Finding the p-value

- ▶ The p-value is, once again, calculated as the area tail area under the t distribution.

Finding the p-value

- ▶ The p-value is, once again, calculated as the area tail area under the t distribution.
- ▶ Using R:

```
2 * pt(4.94, df = 9, lower.tail = FALSE)
```

```
## [1] 0.0008022394
```

Finding the p-value

- ▶ The p-value is, once again, calculated as the area tail area under the t distribution.
- ▶ Using R:

```
2 * pt(4.94, df = 9, lower.tail = FALSE)
```

```
## [1] 0.0008022394
```

- ▶ Using a web app:
https://gallery.shinyapps.io/dist_calc/

Finding the p-value

- ▶ The p-value is, once again, calculated as the area tail area under the t distribution.
- ▶ Using R:

```
2 * pt(4.94, df = 9, lower.tail = FALSE)
```

```
## [1] 0.0008022394
```

- ▶ Using a web app:
https://gallery.shinyapps.io/dist_calc/
- ▶ Or when these aren't available, we can use a t -table.

Conclusion of the test

What is the conclusion of this hypothesis test?

Conclusion of the test

What is the conclusion of this hypothesis test?

Since the p-value is quite low, we conclude that the data provide strong evidence of a difference between traffic flow on Friday 6th and 13th.

What is the difference?

- ▶ We concluded that there is a difference in the traffic flow between Friday 6th and 13th.

What is the difference?

- ▶ We concluded that there is a difference in the traffic flow between Friday 6th and 13th.
- ▶ But it would be more interesting to find out what exactly this difference is.

What is the difference?

- ▶ We concluded that there is a difference in the traffic flow between Friday 6th and 13th.
- ▶ But it would be more interesting to find out what exactly this difference is.
- ▶ We can use a confidence interval to estimate this difference.

Confidence interval for a small sample mean

- Confidence intervals are always of the form

$$\text{point estimate} \pm ME$$

Confidence interval for a small sample mean

- ▶ Confidence intervals are always of the form

$$\text{point estimate} \pm ME$$

- ▶ ME is always calculated as the product of a critical value and SE.

Confidence interval for a small sample mean

- ▶ Confidence intervals are always of the form

$$\text{point estimate} \pm ME$$

- ▶ ME is always calculated as the product of a critical value and SE.
- ▶ Since small sample means follow a t distribution (and not a z distribution), the critical value is a t^* (as opposed to a z^*).

$$\text{point estimate} \pm t^* \times SE$$

Finding the critical $t(t^*)$

Using R:

```
qt(p = 0.975, df = 9)
```

```
## [1] 2.262157
```

Constructing a CI for a small sample mean

Which of the following is the correct calculation of a 95% confidence interval for the difference between the traffic flow between Friday 6th and 13th?

$$\bar{x}_{diff} = 1836 \quad s_{diff} = 1176 \quad n = 10 \quad SE = 372$$

- A) $1836 \pm 1.96 \times 372$
- B) $1836 \pm 2.26 \times 372$
- C) $1836 \pm -2.26 \times 372$
- D) $1836 \pm 2.26 \times 1176$

Constructing a CI for a small sample mean

Which of the following is the correct calculation of a 95% confidence interval for the difference between the traffic flow between Friday 6th and 13th?

$$\bar{x}_{diff} = 1836 \quad s_{diff} = 1176 \quad n = 10 \quad SE = 372$$

- A) $1836 \pm 1.96 \times 372$
- B) $1836 \pm 2.26 \times 372 \rightarrow (995, 2677)$
- C) $1836 \pm -2.26 \times 372$
- D) $1836 \pm 2.26 \times 1176$

Interpreting the CI

Which of the following is the **best** interpretation for the confidence interval we just calculated?

$$\mu_{diff: 6th-13th} = (995, 2677)$$

We are 95% confident that...

- A) The difference between the average number of cars on the road on Friday 6th and 13th is between 995 and 2677.
- B) On Friday 6th there are 995 to 2677 fewer cars on the road than on the Friday 13th, on average.
- C) On Friday 6th there are 995 to 2677 more cars on the road than on the Friday 13th, on average.
- D) On Friday 13th there are 995 to 2677 fewer cars on the road than on the Friday 6th, on average.

Interpreting the CI

Which of the following is the **best** interpretation for the confidence interval we just calculated?

$$\mu_{diff: 6th-13th} = (995, 2677)$$

We are 95% confident that...

- A) The difference between the average number of cars on the road on Friday 6th and 13th is between 995 and 2677.
- B) On Friday 6th there are 995 to 2677 fewer cars on the road than on the Friday 13th, on average.
- C) On Friday 6th there are 995 to 2677 more cars on the road than on the Friday 13th, on average.
- D) On Friday 13th there are 995 to 2677 fewer cars on the road than on the Friday 6th, on average.

Synthesis

Does the conclusion from the hypothesis test agree with the findings of the confidence interval?

Synthesis

Does the conclusion from the hypothesis test agree with the findings of the confidence interval?

Yes, the hypothesis test found a significant difference, and the CI does not contain the null value of 0.

Synthesis

Does the conclusion from the hypothesis test agree with the findings of the confidence interval?

Yes, the hypothesis test found a significant difference, and the CI does not contain the null value of 0.

Do you think the findings of this study suggests that people believe Friday 13th is a day of bad luck?

Synthesis

Does the conclusion from the hypothesis test agree with the findings of the confidence interval?

Yes, the hypothesis test found a significant difference, and the CI does not contain the null value of 0.

Do you think the findings of this study suggests that people believe Friday 13th is a day of bad luck?

No, this is an observational study. We have just observed a significant difference between the number of cars on the road on these two days. We have not tested for people's beliefs.

Difference of two means

Experiment

- ▶ A 2003 American Journal of Health Education Study investigated the effects of cell phone use on reaction time while driving.
- ▶ In the study, 60 participants were randomly selected and placed into one of two groups:
 - ▶ Treatment Group - Access to text documents on a cell phone.
 - ▶ Control Group - No distractions
- ▶ Participants in each group were then asked to take a computerized reaction time test.
- ▶ Researchers then recorded each subject's reaction time in seconds.

Data Summary

	Treatment	Control
	Phone	No Phone
\bar{x}	0.546	0.356
s	0.213	0.245
n	30	30

Parameter and point estimate

- ▶ **Parameter of interest:** Average difference between the reaction time of **all** drivers using a phone or not.

$$\mu_C - \mu_T$$

Parameter and point estimate

- ▶ **Parameter of interest:** Average difference between the reaction time of **all** drivers using a phone or not.

$$\mu_C - \mu_T$$

- ▶ **Point estimate:** Average difference between the reaction time of **participants** in the treatment and control group.

$$\bar{x}_C - \bar{x}_T$$

Hypotheses

Which of the following is the correct set of hypotheses for testing if the average reaction time of the drivers using a phone (μ_T) is higher than the average reaction time of the drivers not using a phone (μ_C)?

A) $H_0 : \mu_C = \mu_T$
 $H_A : \mu_C \neq \mu_T$

B) $H_0 : \mu_C = \mu_T$
 $H_A : \mu_C > \mu_T$

C) $H_0 : \mu_C = \mu_T$
 $H_A : \mu_C < \mu_T$

D) $H_0 : \bar{x}_C = \bar{x}_T$
 $H_A : \bar{x}_C < \bar{x}_T$

Hypotheses

Which of the following is the correct set of hypotheses for testing if the average reaction time of the drivers using a phone (μ_T) is higher than the average reaction time of the drivers not using a phone (μ_C)?

A) $H_0 : \mu_C = \mu_T$
 $H_A : \mu_C \neq \mu_T$

B) $H_0 : \mu_C = \mu_T$
 $H_A : \mu_C > \mu_T$

C) $H_0 : \mu_C = \mu_T$
 $H_A : \mu_C < \mu_T$

D) $H_0 : \bar{x}_C = \bar{x}_T$
 $H_A : \bar{x}_C < \bar{x}_T$

Conditions

Which of the following does not need to be satisfied in order to conduct this hypothesis test using theoretical methods?

- A) The reaction time of drivers not using a phone in the sample should be independent of another, and the reaction time of drivers using a phone should independent of another as well.
- B) The reaction times of drivers using and not using a phone in the sample should be independent.
- C) Distributions of reaction times of drivers in both groups should not be extremely skewed.
- D) Both sample sizes should be the same.

Conditions

Which of the following does not need to be satisfied in order to conduct this hypothesis test using theoretical methods?

- A) The reaction time of drivers not using a phone in the sample should be independent of another, and the reaction time of drivers using a phone should independent of another as well.
- B) The reaction times of drivers using and not using a phone in the sample should be independent.
- C) Distributions of reaction times of drivers in both groups should not be extremely skewed.
- D) Both sample sizes should be the same.

Test statistics

The test statistic for inference on the difference of two means where σ_1 and σ_2 are unknown is the T statistic.

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

where

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{and} \quad df = \min(n_1 - 1, n_2 - 1)$$

Note: The calculation of the df is actually much more complicated. For simplicity we'll use the above formula to estimate the true df when conducting the analysis by hand.

Test statistic

	Treatment Phone	Control No Phone
\bar{x}	0.546	0.356
s	0.213	0.245
n	30	30

in context...

Test statistic

	Treatment Phone	Control No Phone
\bar{x}	0.546	0.356
s	0.213	0.245
n	30	30

in context...

$$T = \frac{\text{point estimate} - \text{null value}}{SE}$$

Test statistic

	Treatment Phone	Control No Phone
\bar{x}	0.546	0.356
s	0.213	0.245
n	30	30

in context...

$$\begin{aligned} T &= \frac{\text{point estimate} - \text{null value}}{SE} \\ &= \frac{(0.356 - 0.546) - 0}{\sqrt{\frac{0.213^2}{30} + \frac{0.245^2}{30}}} \end{aligned}$$

Test statistic

	Treatment Phone	Control No Phone
\bar{x}	0.546	0.356
s	0.213	0.245
n	30	30

in context...

$$\begin{aligned} T &= \frac{\text{point estimate} - \text{null value}}{SE} \\ &= \frac{(0.356 - 0.546) - 0}{\sqrt{\frac{0.213^2}{30} + \frac{0.245^2}{30}}} \\ &= \frac{-0.19}{0.0593} \end{aligned}$$

Test statistic

	Treatment Phone	Control No Phone
\bar{x}	0.546	0.356
s	0.213	0.245
n	30	30

in context...

$$\begin{aligned} T &= \frac{\text{point estimate} - \text{null value}}{SE} \\ &= \frac{(0.356 - 0.546) - 0}{\sqrt{\frac{0.213^2}{30} + \frac{0.245^2}{30}}} \\ &= \frac{-0.19}{0.0593} \\ &= -3.20 \end{aligned}$$

Test statistic

- ▶ Coincidentally, in the experiment the number of participants in both groups is the same.
- ▶ Thus, $df = \min(30 - 1, 30 - 1) = \min(29, 29) = 29$.

p-value

Which of the following is the correct p-value for this hypothesis test?

$$T = -3.20 \quad df = 29$$

- A) Between 0.0005 and 0.001
- B) Between 0.001 and 0.0025
- C) Between 0.002 and 0.005
- D) Between 0.01 and 0.02

p-value

Which of the following is the correct p-value for this hypothesis test?

$$T = -3.20 \quad df = 29$$

- A) Between 0.0005 and 0.001
- B) Between 0.001 and 0.0025
- C) Between 0.002 and 0.005
- D) Between 0.01 and 0.02

```
pt(q = -3.20, df = 29)
```

```
## [1] 0.001659221
```

Synthesis

What is the conclusion of the hypothesis test? How (if at all) would this conclusion change your behavior if you went diamond shopping?

Synthesis

What is the conclusion of the hypothesis test? How (if at all) would this conclusion change your behavior if you went diamond shopping?

- ▶ p-value is small so reject H_0 . The data provide convincing evidence to suggest that the average reaction time of drivers not using a phone is faster than the drivers using a phone while driving.
- ▶ Try not to use your phone while driving because you may never know when you would require the fast reaction time to avoid an accident.

Equivalent confidence level

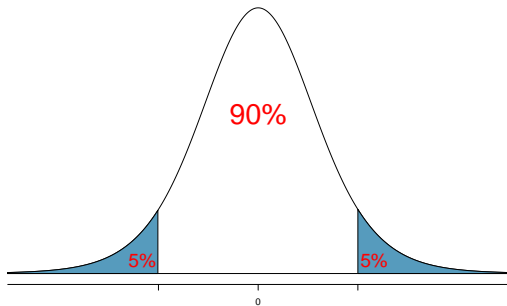
What is the equivalent confidence level for a one-sided hypothesis test at $\alpha = 0.05$?

- A) 90
- B) 92.5
- C) 95
- D) 97.5

Equivalent confidence level

What is the equivalent confidence level for a one-sided hypothesis test at $\alpha = 0.05$?

- A) 90%
- B) 92.5
- C) 95
- D) 97.5



Critical value

What is the appropriate t^* for a confidence interval for the average difference between using a phone and not using it while driving?

- A) 1.32
- B) 1.70
- C) 2.07
- D) 2.82

Critical value

What is the appropriate t^* for a confidence interval for the average difference between using a phone and not using it while driving?

A) 1.32

B) 1.70

C) 2.07

D) 2.82

```
qt(p = 0.95, df = 29)
```

```
## [1] 1.699127
```

Confidence interval

Calculate the interval, and interpret it in context

Confidence interval

Calculate the interval, and interpret it in context

point estimate $\pm ME$

Confidence interval

Calculate the interval, and interpret it in context

point estimate $\pm ME$

$$(\bar{x}_C - \bar{x}_T) \pm t_{df}^* \times SE = (0.356 - 0.546) \pm 1.70 \times 0.0593$$

Confidence interval

Calculate the interval, and interpret it in context

point estimate $\pm ME$

$$\begin{aligned}(\bar{x}_C - \bar{x}_T) \pm t_{df}^* \times SE &= (0.356 - 0.546) \pm 1.70 \times 0.0593 \\ &= -0.19 \pm 0.1008\end{aligned}$$

Confidence interval

Calculate the interval, and interpret it in context

point estimate $\pm ME$

$$\begin{aligned}(\bar{x}_C - \bar{x}_T) \pm t_{df}^* \times SE &= (0.356 - 0.546) \pm 1.70 \times 0.0593 \\&= -0.19 \pm 0.1008 \\&= (-0.2908, -0.0892)\end{aligned}$$

Confidence interval

Calculate the interval, and interpret it in context

point estimate $\pm ME$

$$\begin{aligned}(\bar{x}_C - \bar{x}_T) \pm t_{df}^* \times SE &= (0.356 - 0.546) \pm 1.70 \times 0.0593 \\&= -0.19 \pm 0.1008 \\&= (-0.2908, -0.0892)\end{aligned}$$

We are 90% confident that the average reaction time of a driver not using a phone is 0.0892 to 0.2908 seconds faster than the average reaction time of a driver using a phone.

Recap: Inference using difference of two small sample means

- ▶ If σ_1 or σ_2 is unknown, difference between the sample means follow a t -distribution with $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_1}}$.

Recap: Inference using difference of two small sample means

- ▶ If σ_1 or σ_2 is unknown, difference between the sample means follow a t -distribution with $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_1}}$.
- ▶ Conditions:
 - ▶ Independence within groups (often verified by a random sample, and if sampling without replacement, $n < 10\%$ of population) and between groups.
 - ▶ No extreme skew in either group.

Recap: Inference using difference of two small sample means

- ▶ If σ_1 or σ_2 is unknown, difference between the sample means follow a t -distribution with $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

- ▶ Conditions:

- ▶ Independence within groups (often verified by a random sample, and if sampling without replacement, $n < 10\%$ of population) and between groups.
- ▶ No extreme skew in either group.

- ▶ Hypothesis testing:

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}, \text{ where } df = \min(n_1 - 1, n_2 - 1).$$

Recap: Inference using difference of two small sample means

- ▶ If σ_1 or σ_2 is unknown, difference between the sample means follow a t -distribution with $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_1}}$.

- ▶ Conditions:

- ▶ Independence within groups (often verified by a random sample, and if sampling without replacement, $n < 10\%$ of population) and between groups.
- ▶ No extreme skew in either group.

- ▶ Hypothesis testing:

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}, \text{ where } df = \min(n_1 - 1, n_2 - 1).$$

- ▶ Confidence interval:

$$\text{point estimate} \pm t_{df}^* \times SE$$

Computing the power for a 2-sample test

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		
	H_A true		

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		Type 1 Error, α
	H_A true		

- Type 1 error is rejecting H_0 when you shouldn't have, and the probability of doing so is α (significance level)

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		Type 1 Error, α
	H_A true	Type 2 Error, β	

- ▶ Type 1 error is rejecting H_0 when you shouldn't have, and the probability of doing so is α (significance level)
- ▶ Type 2 error is failing to reject H_0 when you should have, and the probability of doing so is β (a little more complicated to calculate)

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$1 - \alpha$	Type 1 Error, α
	H_A true	Type 2 Error, β	

- ▶ Type 1 error is rejecting H_0 when you shouldn't have, and the probability of doing so is α (significance level)
- ▶ Type 2 error is failing to reject H_0 when you should have, and the probability of doing so is β (a little more complicated to calculate)
- ▶ **Power** of a test is the probability of correctly rejecting H_0 , and the probability of doing so is $1 - \beta$

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$1 - \alpha$	Type 1 Error, α
	H_A true	Type 2 Error, β	Power, $1 - \beta$

- ▶ Type 1 error is rejecting H_0 when you shouldn't have, and the probability of doing so is α (significance level)
- ▶ Type 2 error is failing to reject H_0 when you should have, and the probability of doing so is β (a little more complicated to calculate)
- ▶ **Power** of a test is the probability of correctly rejecting H_0 , and the probability of doing so is $1 - \beta$
- ▶ In hypothesis testing, we want to keep α and β low, but there are inherent trade-offs.

Type 2 error rate

If the alternative hypothesis is actually true, what is the chance that we make a Type 2 Error, i.e. we fail to reject the null hypothesis even when we should reject it?

- ▶ The answer is not obvious.
- ▶ If the true population average is very close to the null hypothesis value, it will be difficult to detect a difference (and reject H_0).
- ▶ If the true population average is very different from the null hypothesis value, it will be easier to detect a difference.
- ▶ Clearly, β depends on the **effect size** (δ).

Example - Blood Pressure (BP), hypotheses

Suppose a pharmaceutical company has developed a new drug for lowering blood pressure, and they are preparing a clinical trial to test the drug's effectiveness. They recruit people who are taking a particular standard blood pressure medication, and half of the subjects are given the new drug (treatment) and the other half continue to take their current medication through generic-looking pills to ensure blinding (control). What are the hypotheses for a two-sided hypothesis test in this context?

Example - Blood Pressure (BP), hypotheses

Suppose a pharmaceutical company has developed a new drug for lowering blood pressure, and they are preparing a clinical trial to test the drug's effectiveness. They recruit people who are taking a particular standard blood pressure medication, and half of the subjects are given the new drug (treatment) and the other half continue to take their current medication through generic-looking pills to ensure blinding (control). What are the hypotheses for a two-sided hypothesis test in this context?

$$H_0 : \mu_{treatment} - \mu_{control} = 0$$

$$H_A : \mu_{treatment} - \mu_{control} \neq 0$$

Example - BP, standard error

Suppose researchers would like to run the clinical trial on patients with systolic blood pressures between 140 and 180 mmHg.

Suppose previously published studies suggest that the standard deviation of the patients' blood pressures will be about 12 mmHg and the distribution of patient blood pressures will be approximately symmetric. If we had 100 patients per group, what would be the approximate standard error for difference in sample means of the treatment and control groups?

Example - BP, standard error

Suppose researchers would like to run the clinical trial on patients with systolic blood pressures between 140 and 180 mmHg.

Suppose previously published studies suggest that the standard deviation of the patients' blood pressures will be about 12 mmHg and the distribution of patient blood pressures will be approximately symmetric. If we had 100 patients per group, what would be the approximate standard error for difference in sample means of the treatment and control groups?

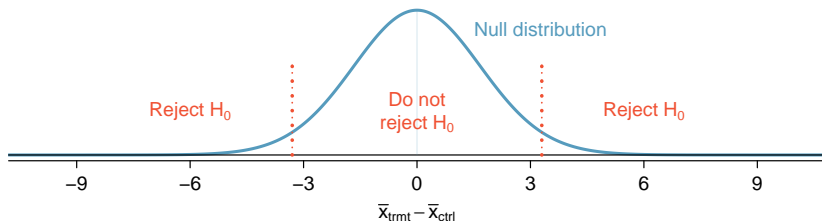
$$SE = \sqrt{\frac{12^2}{100} + \frac{12^2}{100}} = 1.70$$

Example - BP, minimum effect size required to reject H_0

For what values of the difference between the observed averages of blood pressure in treatment and control groups (effect size) would we reject the null hypothesis at the 5% significance level?

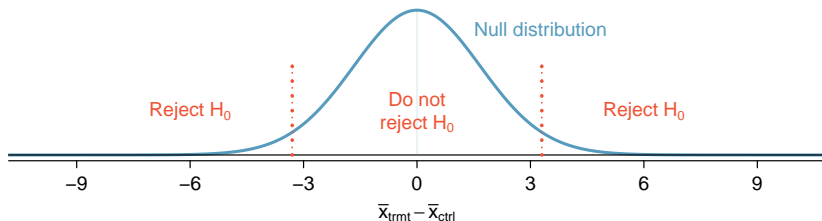
Example - BP, minimum effect size required to reject H_0

For what values of the difference between the observed averages of blood pressure in treatment and control groups (effect size) would we reject the null hypothesis at the 5% significance level?



Example - BP, minimum effect size required to reject H_0

For what values of the difference between the observed averages of blood pressure in treatment and control groups (effect size) would we reject the null hypothesis at the 5% significance level?



The difference should be at least

$$1.96 \times 1.70 = 3.332.$$

or at most

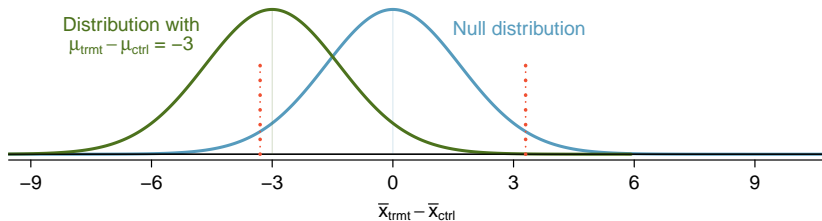
$$-1.96 \times 1.70 = -3.332.$$

Example - BP, power

Suppose that the company researchers care about finding any effect on blood pressure that is 3 mmHg or larger vs the standard medication. What is the power of the test that can detect this effect?

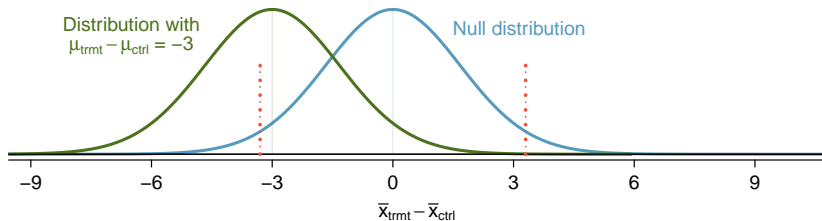
Example - BP, power

Suppose that the company researchers care about finding any effect on blood pressure that is 3 mmHg or larger vs the standard medication. What is the power of the test that can detect this effect?



Example - BP, power

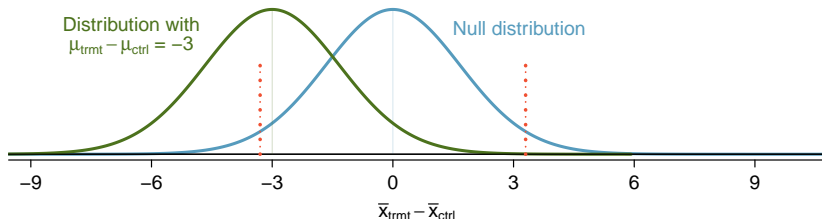
Suppose that the company researchers care about finding any effect on blood pressure that is 3 mmHg or larger vs the standard medication. What is the power of the test that can detect this effect?



$$Z = \frac{-3.332 - (-3)}{1.70} = 0.20$$

Example - BP, power

Suppose that the company researchers care about finding any effect on blood pressure that is 3 mmHg or larger vs the standard medication. What is the power of the test that can detect this effect?



$$Z = \frac{-3.332 - (-3)}{1.70} = 0.20$$

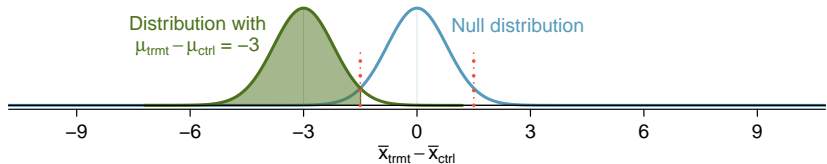
$$P(Z < -0.20) = 0.4207$$

Example - BP, required sample size for 80% power

What sample size will lead to a power of 80% for this test?

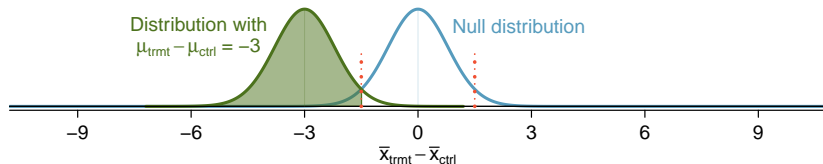
Example - BP, required sample size for 80% power

What sample size will lead to a power of 80% for this test?



Example - BP, required sample size for 80% power

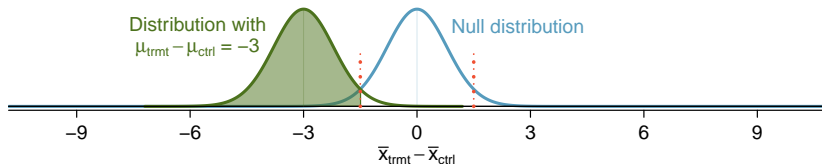
What sample size will lead to a power of 80% for this test?



$$SE = \frac{3}{2.8} = 1.07142$$

Example - BP, required sample size for 80% power

What sample size will lead to a power of 80% for this test?

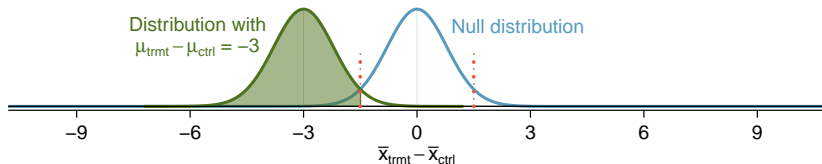


$$SE = \frac{3}{2.8} = 1.07142$$

$$1.07142 = \sqrt{\frac{12^2}{n} + \frac{12^2}{n}}$$

Example - BP, required sample size for 80% power

What sample size will lead to a power of 80% for this test?



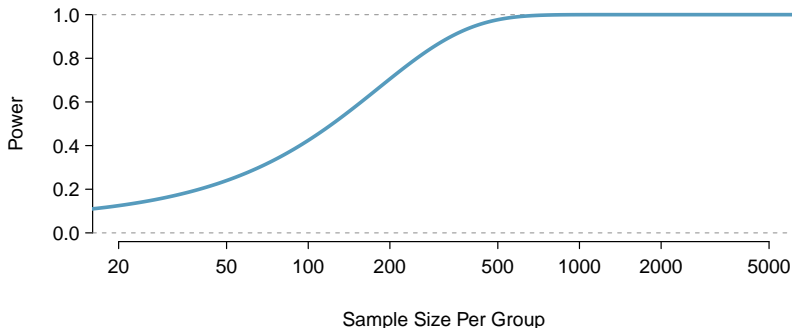
$$SE = \frac{3}{2.8} = 1.07142$$

$$1.07142 = \sqrt{\frac{12^2}{n} + \frac{12^2}{n}}$$

$$n = 250.88 \rightarrow n \geq 251$$

Recap

- ▶ Calculate required sample size for a desired level of power.
- ▶ Calculate power for a range of sample sizes, then choose the sample size that yields the target power (usually 80% or 90%).



Achieving desired power

There are several ways to increase power (and hence decrease type 2 error rate):

Achieving desired power

There are several ways to increase power (and hence decrease type 2 error rate):

1. Increase the sample size.

Achieving desired power

There are several ways to increase power (and hence decrease type 2 error rate):

1. Increase the sample size.
2. Decrease the standard deviation of the sample, which essentially has the same effect as increasing the sample size (it will decrease the standard error). With a smaller s we have a better chance of distinguishing the null value from the observed point estimate. This is difficult to ensure but cautious measurement process and limiting the population so that it is more homogeneous may help.

Achieving desired power

There are several ways to increase power (and hence decrease type 2 error rate):

1. Increase the sample size.
2. Decrease the standard deviation of the sample, which essentially has the same effect as increasing the sample size (it will decrease the standard error). With a smaller s we have a better chance of distinguishing the null value from the observed point estimate. This is difficult to ensure but cautious measurement process and limiting the population so that it is more homogeneous may help.
3. Increase α , which will make it more likely to reject H_0 (but note that this has the side effect of increasing the Type 1 error rate).

Achieving desired power

There are several ways to increase power (and hence decrease type 2 error rate):

1. Increase the sample size.
2. Decrease the standard deviation of the sample, which essentially has the same effect as increasing the sample size (it will decrease the standard error). With a smaller s we have a better chance of distinguishing the null value from the observed point estimate. This is difficult to ensure but cautious measurement process and limiting the population so that it is more homogeneous may help.
3. Increase α , which will make it more likely to reject H_0 (but note that this has the side effect of increasing the Type 1 error rate).
4. Consider a larger effect size. If the true mean of the population is in the alternative hypothesis but close to the null value, it will be harder to detect a difference.