What will I need to know for Midterm 2?

Let's focus on <u>hypothesis testing</u> for right now.

According to WebAssign (Brase & Brase, 2019):

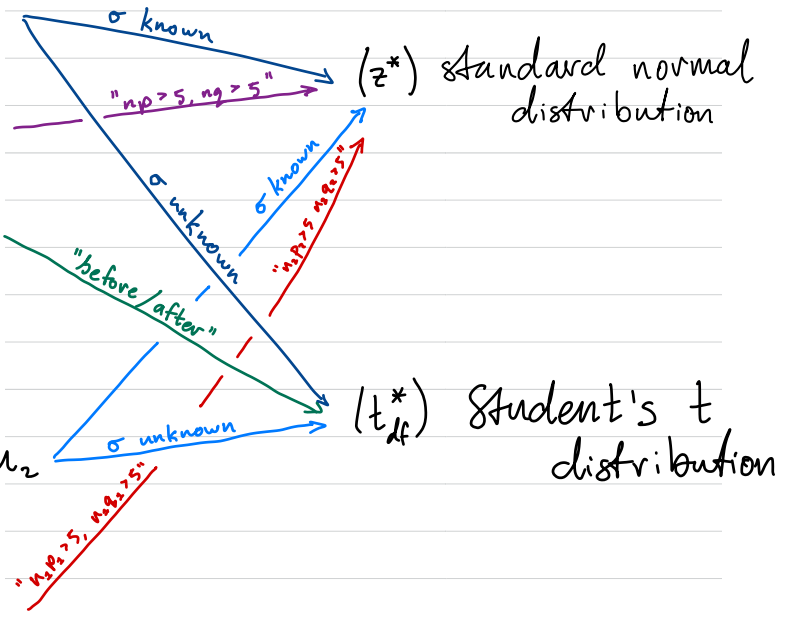
8.2 population mean — μ

8.3 population proportion — p

8.4 difference of means (from "paired data" — dependent populations) — d

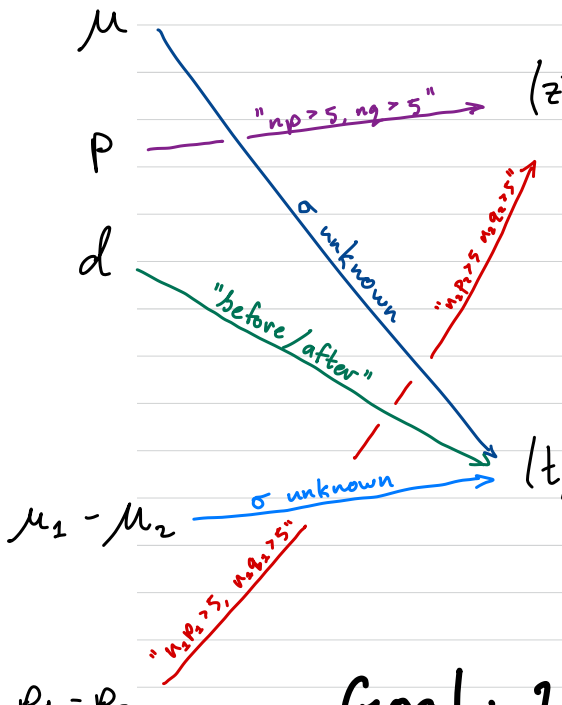
8.5 difference of means (from independent populations) — $\mu_1 - \mu_2$

8.5 difference of proportions (from indep. populations) — $p_1 - p_2$



But in the "real world", σ is never known!

So? We'll "eat dessert first."



(z^*) standard normal distribution

- p
- $p_A - p_B$

"categorical" data (proportions)

(t^*_{df}) Student's t distribution

- μ
- $\mu_A - \mu_B$
- d

"numerical" data (means)

Goal: Unify procedures for hypothesis tests.

$$\text{test statistic} = \frac{\text{point estimate} - \text{null value}}{\text{standard error}}$$

Plan

TODAY
- Summarize hypothesis testing framework for **both** categorical and numerical data
- Inference using the standard norm. distr. ⇝ p
- Inference using Student's t-distribution. ⇝ μ

MONDAY · Difference of two proportions ⟶ $p_A - p_B$

WEDNESDAY · Difference of two means ⟶ $\mu_A - \mu_B$

FRIDAY · Paired numerical data ⟶ d

MONDAY · Practice midterm / calculator review

WEDNESDAY · Midterm II

If you're reading from Open Intro,
we just finished Ch. 5 on "foundations of inference".
Ch. 6 is all about categorical data.
Ch. 7 covers numerical data (and Student's t-distr).

Inference using the standard norm. distr. ⇝ p

(Quick recap)

SAMPLING DISTRIBUTION OF $\hat{p}$

The sampling distribution for $\hat{p}$ based on a sample of size n from a population with a true proportion p is nearly normal when:

1. The sample's observations are independent, e.g. are from a simple random sample.
2. We expected to see at least 10 successes and 10 failures in the sample, i.e. $np \geq 10$ and $n(1-p) \geq 10$. This is called the **success-failure condition**.

When these conditions are met, then the sampling distribution of $\hat{p}$ is nearly normal with mean p and standard error $SE = \sqrt{\frac{p(1-p)}{n}}$.

Typically we don't know the true proportion p, so we substitute some value to check conditions and estimate the standard error. For confidence intervals, the sample proportion $\hat{p}$ is used to check the success-failure condition and compute the standard error. For hypothesis tests, typically the null value – that is, the proportion claimed in the null hypothesis – is used in place of p.

confidence int'ls
$\hat{p} \approx p$ allowed

hypothesis tests
use $p_0 \approx p$, not $\hat{p}$!

CONFIDENCE INTERVAL FOR A SINGLE PROPORTION

Once you've determined a one-proportion confidence interval would be helpful for an application, there are four steps to constructing the interval:

Prepare. Identify $\hat{p}$ and n, and determine what confidence level you wish to use.

Check. Verify the conditions to ensure $\hat{p}$ is nearly normal. For one-proportion confidence intervals, use $\hat{p}$ in place of p to check the success-failure condition.

Calculate. If the conditions hold, compute SE using $\hat{p}$, find $z^\star$, and construct the interval.

Conclude. Interpret the confidence interval in the context of the problem.

HYPOTHESIS TESTING FOR A SINGLE PROPORTION

Once you've determined a one-proportion hypothesis test is the correct procedure, there are four steps to completing the test:

Prepare. Identify the parameter of interest, list hypotheses, identify the significance level, and identify $\hat{p}$ and n.

do not substitute $\hat{p} \approx p$!

Check. Verify conditions to ensure $\hat{p}$ is nearly normal under H_0. For one-proportion hypothesis tests, use the null value to check the success-failure condition.

Calculate. If the conditions hold, compute the standard error, again using p_0, compute the Z-score, and identify the p-value.

Conclude. Evaluate the hypothesis test by comparing the p-value to α, and provide a conclusion in the context of the problem.

$$Z = \frac{\text{point estimate} - \text{null value}}{\text{standard error}} = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

$$SE = \sqrt{\frac{0.80 \times 0.20}{670}} = 0.0154$$

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

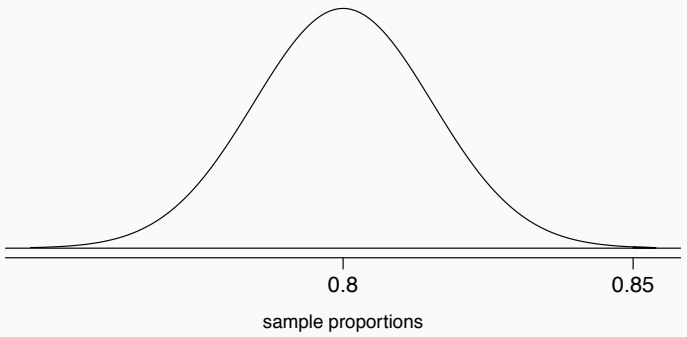
$$SE = \sqrt{\frac{0.80 \times 0.20}{670}} = 0.0154$$

$$Z = \frac{0.85 - 0.80}{0.0154} = 3.25$$

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

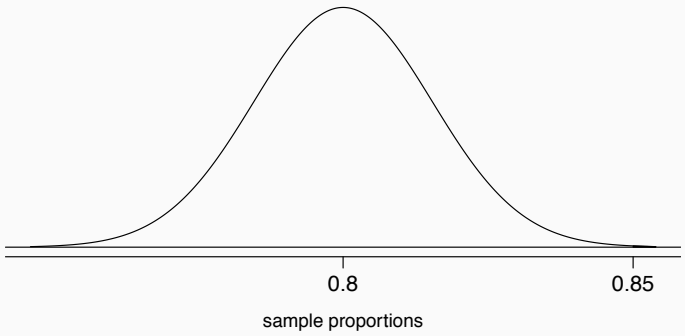
$$\begin{aligned} SE &= \sqrt{\frac{0.80 \times 0.20}{670}} = 0.0154 \\ Z &= \frac{0.85 - 0.80}{0.0154} = 3.25 \\ p - value &= 1 - 0.9994 = 0.0006 \end{aligned}$$



The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

$$\begin{aligned} SE &= \sqrt{\frac{0.80 \times 0.20}{670}} = 0.0154 \\ Z &= \frac{0.85 - 0.80}{0.0154} = 3.25 \\ p - value &= 1 - 0.9994 = 0.0006 \end{aligned}$$



Since the p-value is low, we reject H_0. The data provide convincing evidence that more than 80% of Americans have a good intuition on experimental design.

Inference using Student's t-distribution. $\rightsquigarrow \mu$

(in detail)

CENTRAL LIMIT THEOREM FOR THE SAMPLE MEAN

When we collect a sufficiently large sample of n independent observations from a population with mean μ and standard deviation σ, the sampling distribution of $\bar{x}$ will be nearly normal with

$$\text{Mean} = \mu \qquad\qquad \text{Standard Error } (SE) = \frac{\sigma}{\sqrt{n}}$$

Before diving into confidence intervals and hypothesis tests using $\bar{x}$, we first need to cover two topics:

- When we modeled $\hat{p}$ using the normal distribution, certain conditions had to be satisfied. The conditions for working with $\bar{x}$ are a little more complex, and we'll spend ~~Section 7.1.2~~ the next slide discussing how to check conditions for inference.
- The standard error is dependent on the population standard deviation, σ. However, we rarely know σ, and instead we must estimate it. Because this estimation is itself imperfect, we use a new distribution called the t-distribution to fix this problem, which we discuss in ~~Section 7.1.3.~~ two slides

Two conditions are required to apply the Central Limit Theorem for a sample mean $\bar{x}$:

Independence. The sample observations must be independent, The most common way to satisfy this condition is when the sample is a simple random sample from the population. If the data come from a random process, analogous to rolling a die, this would also satisfy the independence condition.

Normality. When a sample is small, we also require that the sample observations come from a normally distributed population. We can relax this condition more and more for larger and larger sample sizes. This condition is obviously vague, making it difficult to evaluate, so next we introduce a couple rules of thumb to make checking this condition easier.

RULES OF THUMB: HOW TO PERFORM THE NORMALITY CHECK

There is no perfect way to check the normality condition, so instead we use two rules of thumb:

$n < 30$: If the sample size n is less than 30 and there are no clear outliers in the data, then we typically assume the data come from a nearly normal distribution to satisfy the condition.

$n \geq 30$: If the sample size n is at least 30 and there are no *particularly extreme* outliers, then we typically assume the sampling distribution of $\bar{x}$ is nearly normal, even if the underlying distribution of individual observations is not.

In practice, we cannot directly calculate the standard error for $\bar{x}$ since we do not know the population standard deviation, σ. We encountered a similar issue when computing the standard error for a sample proportion, which relied on the population proportion, p. Our solution in the proportion context was to use sample value in place of the population value when computing the standard error. We'll employ a similar strategy for computing the standard error of $\bar{x}$, using the sample standard deviation s in place of σ:

$$SE = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

This strategy tends to work well when we have a lot of data and can estimate σ using s accurately. However, the estimate is less precise with smaller samples, and this leads to problems when using the normal distribution to model $\bar{x}$.

FINDING A t-CONFIDENCE INTERVAL FOR THE MEAN

Based on a sample of n independent and nearly normal observations, a confidence interval for the population mean is

$$\text{point estimate} \pm t^\star_{df} \times SE \quad \rightarrow \quad \bar{x} \pm t^\star_{df} \times \frac{s}{\sqrt{n}}$$

where $\bar{x}$ is the sample mean, $t^\star_{df}$ corresponds to the confidence level and degrees of freedom df, and SE is the standard error as estimated by the sample.

Let's get our first taste of applying the t-distribution in the context of an example about the mercury content of dolphin muscle. Elevated mercury concentrations are an important problem for both dolphins and other animals, like humans, who occasionally eat them.

Figure 7.5: A Risso's dolphin.

Photo by Mike Baird (www.bairdphotos.com). CC BY 2.0 license.

n	$\bar{x}$	s	minimum	maximum
19	4.4	2.3	1.7	9.2

Figure 7.6: Summary of mercury content in the muscle of 19 Risso's dolphins from the Taiji area. Measurements are in micrograms of mercury per wet gram of muscle (μg/wet g).

EXAMPLE 7.6

Are the independence and normality conditions satisfied for this data set?

The observations are a simple random sample, therefore independence is reasonable. The summary statistics in Figure 7.6 do not suggest any clear outliers, since all observations are within 2.5 standard deviations of the mean. Based on this evidence, the normality condition seems reasonable.

In the normal model, we used $z^\star$ and the standard error to determine the width of a confidence interval. We revise the confidence interval formula slightly when using the t-distribution:

$$\text{point estimate} \pm t^\star_{df} \times SE \qquad \rightarrow \qquad \bar{x} \pm t^\star_{df} \times \frac{s}{\sqrt{n}}$$

EXAMPLE 7.7

Using the summary statistics in Figure 7.6, compute the standard error for the average mercury content in the $n = 19$ dolphins.

We plug in s and n into the formula: $SE = s/\sqrt{n} = 2.3/\sqrt{19} = 0.528$.

The value $t^\star_{df}$ is a cutoff we obtain based on the confidence level and the t-distribution with df degrees of freedom. That cutoff is found in the same way as with a normal distribution: we find $t^\star_{df}$ such that the fraction of the t-distribution with df degrees of freedom within a distance $t^\star_{df}$ of 0 matches the confidence level of interest.

EXAMPLE 7.8

When $n = 19$, what is the appropriate degrees of freedom? Find $t^\star_{df}$ for this degrees of freedom and the confidence level of 95%

The degrees of freedom is easy to calculate: $df = n - 1 = 18$.

why?

Using statistical software, we find the cutoff where the upper tail is equal to 2.5%: $t^\star_{18} = 2.10$. The area below -2.10 will also be equal to 2.5%. That is, 95% of the t-distribution with $df = 18$ lies within 2.10 units of 0.

EXAMPLE 7.9

Compute and interpret the 95% confidence interval for the average mercury content in Risso's dolphins.

We can construct the confidence interval as

$$\bar{x} \pm t^\star_{18} \times SE \quad \rightarrow \quad 4.4 \pm 2.10 \times 0.528 \quad \rightarrow \quad (3.29, 5.51)$$

We are 95% confident the average mercury content of muscles in Risso's dolphins is between 3.29 and 5.51 μg/wet gram, which is considered extremely high.

HYPOTHESIS TESTING FOR A SINGLE MEAN

Once you've determined a one-mean hypothesis test is the correct procedure, there are four steps to completing the test:

Prepare. Identify the parameter of interest, list out hypotheses, identify the significance level, and identify $\bar{x}$, s, and n.

Check. Verify conditions to ensure $\bar{x}$ is nearly normal.

Calculate. If the conditions hold, compute SE, compute the T-score, and identify the p-value.

Conclude. Evaluate the hypothesis test by comparing the p-value to α, and provide a conclusion in the context of the problem.

Friday the 13th

Between 1990 - 1992 researchers in the UK collected data on traffic flow, accidents, and 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	6th	13th	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 13^{th}

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

Hypotheses

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

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

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

Hypotheses

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

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

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

Back to Friday the 13th

	type	date	6th	13th	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

$\downarrow$

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

$s_{diff} = 1176$

Finding the test statistic

Test statistic for inference on a small sample mean

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

Test statistic for inference on a small sample mean

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...

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

Finding the test statistic

Test statistic for inference on a small sample mean

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...

$$\begin{aligned} point\ estimate &= \bar{x}_{diff} = 1836 \\ SE &= \frac{s_{diff}}{\sqrt{n}} = \frac{1176}{\sqrt{10}} = 372 \end{aligned}$$

Finding the test statistic

Test statistic for inference on a small sample mean

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...

$$\begin{aligned} point\ estimate &= \bar{x}_{diff} = 1836 \\ SE &= \frac{s_{diff}}{\sqrt{n}} = \frac{1176}{\sqrt{10}} = 372 \\ T &= \frac{1836 - 0}{372} = 4.94 \end{aligned}$$

Finding the test statistic

Test statistic for inference on a small sample mean

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...

$$\begin{aligned} point\ estimate &= \bar{x}_{diff} = 1836 \\ SE &= \frac{s_{diff}}{\sqrt{n}} = \frac{1176}{\sqrt{10}} = 372 \\ T &= \frac{1836 - 0}{372} = 4.94 \\ df &= 10 - 1 = 9 \end{aligned}$$

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.
- Using R:

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

[1] 0.0008022394
```

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

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 6^{th} and 13^{th}.

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 $t^\star$ (as opposed to a $z^\star$).

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

Finding the critical t ($t^\star$)

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?

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

(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?

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

(a) $1836 \pm 1.96 \times 372$

(b) $1836 \pm 2.26 \times 372$ → *(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 2,677.
(b) on Friday 6th there are 995 to 2,677 fewer cars on the road than on the Friday 13th, on average.
(c) on Friday 6th there are 995 fewer to 2,677 more cars on the road than on the Friday 13th, on average.
(d) on Friday 13th there are 995 to 2,677 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?

Do you think the findings of this study suggests that people believe Friday 13^{th} 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 13^{th} 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 13^{th} 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.