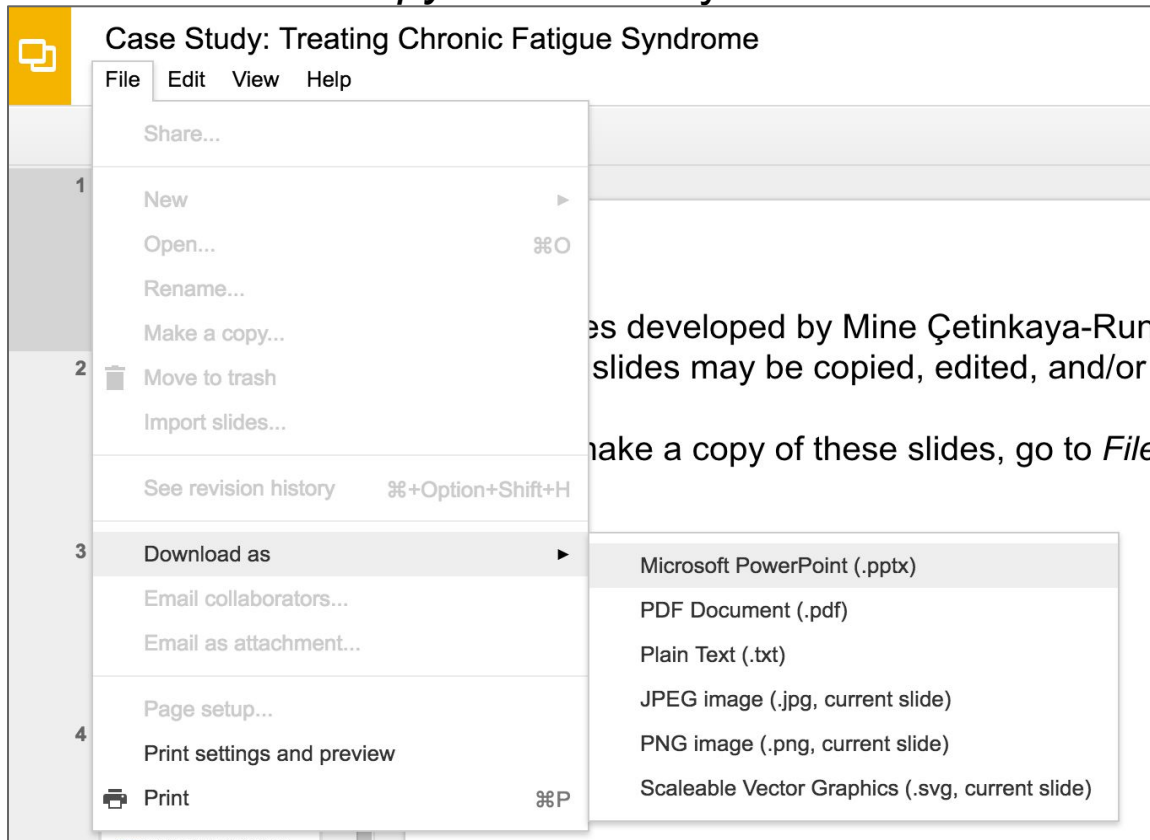


Slides developed by Mine Çetinkaya-Rundel of OpenIntro
Translated from LaTeX to Google Slides by Curry W. Hilton of OpenIntro.
The slides may be copied, edited, and/or shared via the [CC BY-SA license](#)

To make a copy of these slides, go to *File > Download as > [option]*, as shown below. Or if you are logged into a Google account, you can choose *Make a copy...* to create your own version in Google Drive.



Paired Data

Slides developed by Mine Çetinkaya-Rundel of OpenIntro

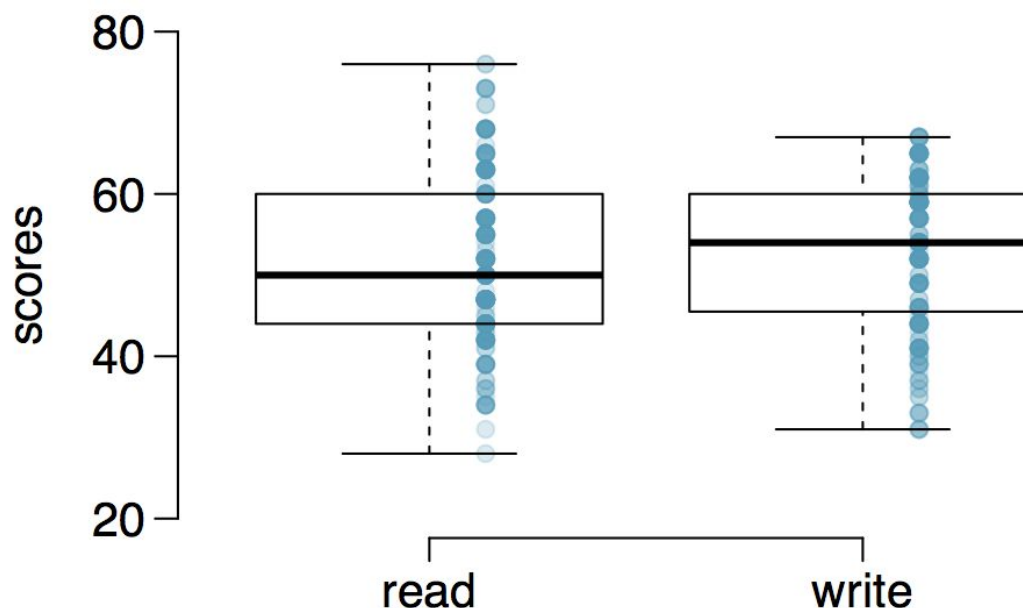
Translated from LaTeX to Google Slides by Curry W. Hilton of OpenIntro.

The slides may be copied, edited, and/or shared via the [CC BY-SA license](#)

Some images may be included under fair use guidelines (educational purposes)

Paired observations

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?



Paired observations

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

Paired observations

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

Analyzing paired data

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

Analyzing 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

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

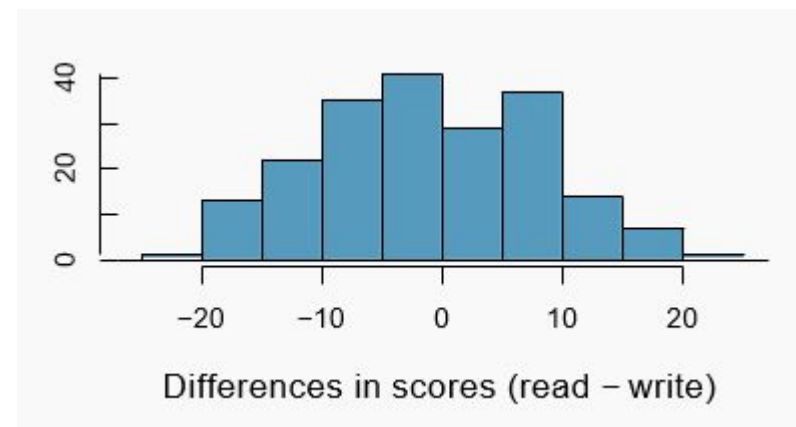
Analyzing paired data

- 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	5
2	86	44	33	11
3	141	63	44	19
4	172	47	52	-5
:	:	:	:	:
200	137	63	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 : Average traffic flow on Friday 6th and 13th are equal.

$$\mu_{diff} = 0$$

H_A : Average traffic flow on Friday 6th and 13th are different.

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

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 the 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

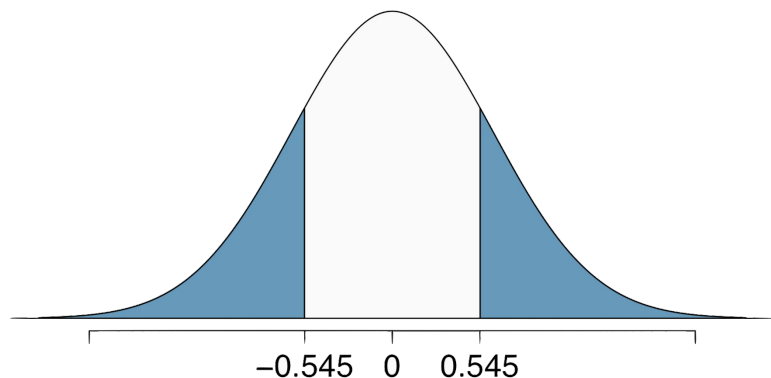
Checking assumptions & conditions

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-statistics and the p-value

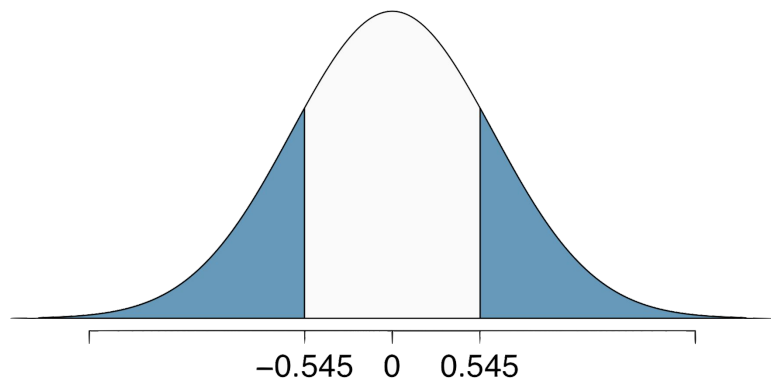
The observed average difference between the two scores is -0.545 points and the standard deviation of the difference is 8.887 points. Do these data provide convincing evidence of a difference between the average scores on the two exams?
Use $\alpha = 0.05$



Calculating the test-statistics 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 $\alpha = 0.05$



$$T = \frac{-0.545 - 0}{\frac{8.887}{\sqrt{200}}}$$

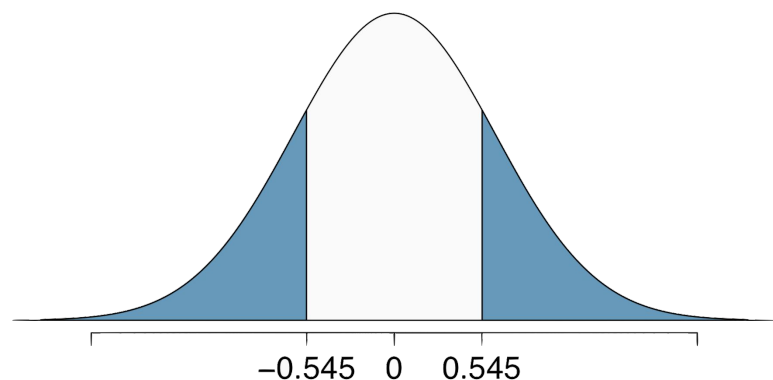
$$T = \frac{-0.545}{0.628} = -0.87$$

$$df = 200 - 1 = 199$$

Calculating the test-statistics 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 $\alpha = 0.05$



$$T = \frac{-0.545 - 0}{\frac{8.887}{\sqrt{200}}}$$

$$T = \frac{-0.545}{0.628} = -0.87$$

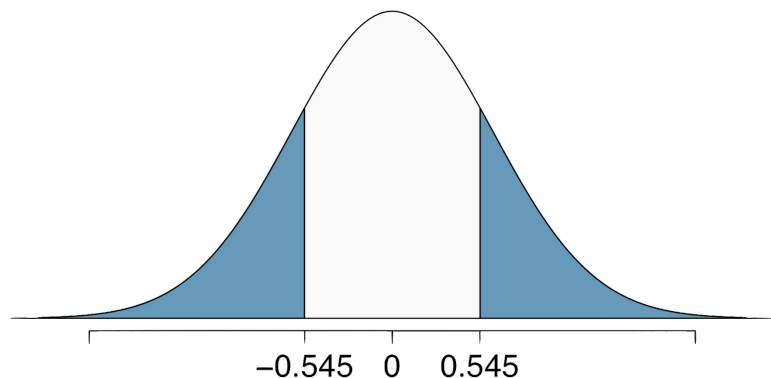
$$df = 200 - 1 = 199$$

$$p - value = 0.1927 \times 2 = 0.3854$$

Calculating the test-statistics 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 $\alpha = 0.05$



$$T = \frac{-0.545 - 0}{\frac{8.887}{\sqrt{200}}}$$

$$T = \frac{-0.545}{0.628} = -0.87$$

$$df = 200 - 1 = 199$$

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

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 a 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 a 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

$$\begin{aligned} -0.545 \pm 1.97 \frac{8.887}{\sqrt{200}} &= -0.545 \pm 1.87 \times 0.628 \\ &= -0.545 \pm 1.24 \\ &= (-1.785, 0.695) \end{aligned}$$

Find more resources at openintro.org/os, including

- Slides
- Videos
- Statistical Software Labs
- Discussion Forums (free support for students and teachers)
- Learning Objectives

Teachers only content is also available for [Verified Teachers](#), including

- Exercise solutions
- Sample exams
- Ability to request a free desk copy for a course
- Statistics Teachers email group

Questions? [Contact us](#).