Statistical Methods

Nels Grevstad

Metropolitan State University of Denver

ngrevsta@msudenver.edu

September 12, 2019

Topics

Objectives

Objectives:

- Identify potentially confounding variables (in observational studies).
- Carry out a paired t test for two population means.
- Compute and interpret a paired t CI for the difference between two population means.

Identifying Causality: Experiments vs Observational Studies

- Many studies are carried out to examine whether two variables (called ***explanatory*** and ***response*** variables) are **related** to each other. For example:

Identifying Causality: Experiments vs Observational Studies

- Many studies are carried out to examine whether two variables (called ***explanatory*** and ***response*** variables) are **related** to each other. For example:
  - Does a person's income (response) depend on their gender (explanatory variable)?

Identifying Causality: Experiments vs Observational Studies

- Many studies are carried out to examine whether two variables (called ***explanatory*** and ***response*** variables) are **related** to each other. For example:
  - Does a person's income (response) depend on their gender (explanatory variable)?
  - Does a person's risk of colon cancer (response) depend on their diet (explanatory variable)?

- Such studies can be either of two types:

- Such studies can be either of two types:
  - ***Observational study***: The investigator merely **observes** whether the two variables vary together.

- Such studies can be either of two types:
  - ***Observational study***: The investigator merely **observes** whether the two variables vary together.

    No attempt is made to **induce changes** in the response variable.

- Such studies can be either of two types:
  - ***Observational study***: The investigator merely **observes** whether the two variables vary together.

    No attempt is made to **induce changes** in the response variable.
  - ***Experiment***: **Treatments** are **imposed** on individuals.

- Such studies can be either of two types:
  - ***Observational study***: The investigator merely **observes** whether the two variables vary together.

    No attempt is made to **induce changes** in the response variable.
  - ***Experiment***: **Treatments** are **imposed** on individuals.

    A **deliberate attempt** is made to **induce changes** in the response variable.

- An **observational study** (by itself) ***can't*** establish **cause and effect**.

- An **observational study** (by itself) ***can't*** establish **cause and effect**.

  Such studies suffer from the possible presence of variables whose effects on the response are ***confounded*** with the effect (if any) of the explanatory variable.

Exercise

An **observational study** showed that people who eat foods rich in antioxidants (such as fruits and vegetables) have lower rates of colon cancer than those who don't eat such foods.

Exercise

An **observational study** showed that people who eat foods rich in antioxidants (such as fruits and vegetables) have lower rates of colon cancer than those who don't eat such foods.

a) Can we conclude that eating such foods **reduces** the risk of colon cancer?

Exercise

An **observational study** showed that people who eat foods rich in antioxidants (such as fruits and vegetables) have lower rates of colon cancer than those who don't eat such foods.

a) Can we conclude that eating such foods **reduces** the risk of colon cancer?

b) List a few possible **confounding** variables that might explain the lower rates of colon cancer.

Exercise

An **observational study** showed that people who eat foods rich in antioxidants (such as fruits and vegetables) have lower rates of colon cancer than those who don't eat such foods.

a) Can we conclude that eating such foods **reduces** the risk of colon cancer?

b) List a few possible **confounding** variables that might explain the lower rates of colon cancer.

   Hint: Try to identify **other** ways in which people who eat lots of fruits and vegetables might differ from people who don't.

- To establish **cause and effect**, we need to carry out an **experiment.**

Example

In a clinical trial (**experiment** on human subjects) to investigate whether dietary antioxidants (vitamins A, C, and E) would lower colon cancer rates, **864** subjects were **randomized** to four treatment groups given different amounts of antioxidants:

Example

In a clinical trial (**experiment** on human subjects) to investigate whether dietary antioxidants (vitamins A, C, and E) would lower colon cancer rates, **864** subjects were **randomized** to four treatment groups given different amounts of antioxidants:

Group 1: Daily beta carotene (vitamin A)
Group 2: Daily vitamins C and E
Group 3: All three vitamins daily
Group 4: No vitamin supplements.

Example

In a clinical trial (**experiment** on human subjects) to investigate whether dietary antioxidants (vitamins A, C, and E) would lower colon cancer rates, **864** subjects were **randomized** to four treatment groups given different amounts of antioxidants:

Group 1: Daily beta carotene (vitamin A)
Group 2: Daily vitamins C and E
Group 3: All three vitamins daily
Group 4: No vitamin supplements.

After four years, researchers were surprised to find no significant difference in colon cancer among these groups.

- Note:

  1. **Randomization** produces groups that are **similar** with respect to variables whose effects might **otherwise** be **confounded** with antioxidant intake (e.g. amount of exercise, smoking and drinking status, etc.).

- Note:

  1. **Randomization** produces groups that are **similar** with respect to variables whose effects might **otherwise** be **confounded** with antioxidant intake (e.g. amount of exercise, smoking and drinking status, etc.).
  2. **Before** imposing treatments, any differences across groups in propensity for developing colon cancer would be due to **chance**.

- Note:

  1. **Randomization** produces groups that are **similar** with respect to variables whose effects might **otherwise** be **confounded** with antioxidant intake (e.g. amount of exercise, smoking and drinking status, etc.).

  2. **Before** imposing treatments, any differences across groups in propensity for developing colon cancer would be due to **chance**.

  3. Therefore, **after** imposing treatments, any **statistically significant differences** in colon cancer rates could be attributed the **effects** of the treatments (antioxidants).

Paired t Test for Two Population Means μ_1 and μ_2

Paired Samples Study Designs

- The ***paired*** t ***test*** is used with two samples collected using a ***paired samples study design***.

Exercise

An **experiment** is to be carried out to compare the amount of wear using two different materials for soles of boys' shoes.

Exercise

An **experiment** is to be carried out to compare the amount of wear using two different materials for soles of boys' shoes.

Independent Samples Study Design: Randomly split twenty boys into two treatment groups of size ten, one receiving shoes with **material A** and the other shoes with **material B**.

Exercise

An **experiment** is to be carried out to compare the amount of wear using two different materials for soles of boys' shoes.

Independent Samples Study Design: Randomly split twenty boys into two treatment groups of size ten, one receiving shoes with **material A** and the other shoes with **material B.**

Paired Samples Study Design: Give each of ten boys one shoe made with **material A** and the other with **material B**. Randomly choose which shoe (left or right) gets which material.

Exercise

An **experiment** is to be carried out to compare the amount of wear using two different materials for soles of boys' shoes.

Independent Samples Study Design: Randomly split twenty boys into two treatment groups of size ten, one receiving shoes with **material A** and the other shoes with **material B.**

Paired Samples Study Design: Give each of ten boys one shoe made with **material A** and the other with **material B**. Randomly choose which shoe (left or right) gets which material.

Which study design is preferred? Why?

- Because boys spend different amounts of time on their feet and run, walk, and play differently, the amount of **shoe wear** will **vary** from one boy to the next.

- Because boys spend different amounts of time on their feet and run, walk, and play differently, the amount of **shoe wear** will **vary** from one boy to the next.
- In the **paired samples study design**, each boy serves as his own ***control*** (or **comparison**) – the amount of time he spends on his feet and the way he runs, walks, and plays affects both feet equally.

- Because boys spend different amounts of time on their feet and run, walk, and play differently, the amount of **shoe wear** will **vary** from one boy to the next.
- In the **paired samples study design**, each boy serves as his own ***control*** (or **comparison**) – the amount of time he spends on his feet and the way he runs, walks, and plays affects both feet equally.

Paired t Test

- Suppose we have two samples from a **paired samples study design.**

Paired t Test

- Suppose we have two samples from a **paired samples study design.**
- We'll see how to use the samples to decide if the two **population means** μ_1 and μ_2 are different.

Paired t Test

- Suppose we have two samples from a **paired samples study design.**
- We'll see how to use the samples to decide if the two **population means** μ_1 and μ_2 are different.
- We denote the first sample by $X_1, X_2, \ldots, X_n$ and the second by $Y_1, Y_2, \ldots, Y_n$, where each X_i is **paired** with it's corresponding Y_i.

- The **null hypothesis** is that no difference between the population means μ_1 and μ_2:

Null Hypothesis:

$$H_0 : \mu_1 - \mu_2 \; = \; 0$$

- The **alternative hypothesis** will depend on what we're trying to "prove":

Alternative Hypothesis: The alternative hypothesis will be one of

1. $H_a : \mu_1 - \mu_2 > 0$ (**one-sided, upper-tailed**)
2. $H_a : \mu_1 - \mu_2 < 0$ (**one-sided, lower-tailed**)
3. $H_a : \mu_1 - \mu_2 \neq 0$ (**two-sided, two-tailed**)

depending on what we're trying to verify using the data.

- Consider the n **differences**

$$\begin{aligned} D_1 &= X_1 - Y_1 \\ D_2 &= X_2 - Y_2 \\ &\vdots \\ D_n &= X_n - Y_n \end{aligned}$$

Example

Here are the data on amount of wear in soles of boys shoes.

Boy	Material B	Material A	Difference
1	14.0	13.2	0.8
2	8.8	8.2	0.6
3	11.2	10.9	0.3
4	14.2	14.3	-0.1
5	11.8	10.7	1.1
6	6.4	6.6	-0.2
7	9.8	9.5	0.3
8	11.3	10.8	0.5
9	9.3	8.8	0.5
10	13.6	13.3	0.3
	$\bar{X} = 11.04$	$\bar{Y} = 10.63$	$\bar{D} = 0.41$
	$s_x = 2.52$	$s_y = 2.45$	$s_D = 0.39$

- We considered $D_1, D_2, \ldots, D_n$ to be a **single random sample** from a ***population of differences*** whose **mean** is $\boldsymbol{\mu_d}$.

Equivalent Ways of Stating the Hypotheses

Proposition

μ_d is related to μ_1 and μ_2 as follows.

$$\mu_d = \mu_1 - \mu_2.$$

- The above fact holds because

$$\mu_d \;=\; E(D_i) \;=\; E(X_i - Y_i)\,,$$

and $X_i - Y_i$ is a **linear combination** of X_i and Y_i, so

$$E(X_i - Y_i) \;=\; E(X_i) - E(Y_i) \;=\; \mu_1 - \mu_2\,.$$

- Hypotheses about $\mu_1 - \mu_2$ can be written in terms of μ_d:

Hypothesis about μ_1 and μ_2	**Equivalent Hypothesis about μ_d**
$H_0 : \mu_1 - \mu_2 = 0$	$H_0 : \mu_d = 0$
$H_a : \mu_1 - \mu_2 > 0$	$H_a : \mu_d > 0$
$H_a : \mu_1 - \mu_2 < 0$	$H_a : \mu_d < 0$
$H_a : \mu_1 - \mu_2 \neq 0$	$H_a : \mu_d \neq 0$

Paired t Test Statistic for $\mu_1 - \mu_2$ (or μ_d):

$$T = \frac{\bar{D} - \mu_d}{S_D/\sqrt{n}},$$

where $\bar{D}$ and S_D are the **sample mean** and **sample standard deviation** of the **differences** $D_1, D_2, \ldots, D_n$.

Paired t Test Statistic for $\mu_1 - \mu_2$ (or μ_d):

$$T \;=\; \frac{\bar{D} - \mu_d}{S_D/\sqrt{n}},$$

where $\bar{D}$ and S_D are the **sample mean** and **sample standard deviation** of the **differences** $D_1, D_2, \ldots, D_n$.

- t is just the **one-sample t test statistic** for a test of μ_d.

- Now suppose either the sample of **differences** is from a $N(\mu_d, \sigma_d)$ population or n is *large*.

- Now suppose either the sample of **differences** is from a $N(\mu_d, \sigma_d)$ population or n is *large*.

  In this case, the sampling distribution of the test statistic is as follows.

Sampling Distribution of the Test Statistic Under H_0:
If t is the paired t test statistic, then when

$$H_0 : \mu_1 - \mu_2 = 0 \qquad \text{(or } H_0 : \mu_d = 0)$$

is true,

$$t \sim t(n-1).$$

- The $t(n-1)$ curve gives us:

- The $t(n-1)$ curve gives us:
    - The ***rejection region*** as the **extreme 100α% of t values** (in the direction(s) specified by H_a).

- The $t(n-1)$ curve gives us:
  - The ***rejection region*** as the **extreme 100α% of t values** (in the direction(s) specified by H_a).
  - The ***p-value*** as the **tail area(s) beyond the observed t value** (in the direction(s) specified by H_a).

- **Comment**: Sometimes we want to test

$$H_0 : \mu_d = \Delta_0$$

where Δ_0 is some non-zero value. In this case the test statistic is

$$T = \frac{\bar{D} - \Delta_0}{S_D/\sqrt{n}},$$

which follows a $t(n-1)$ distribution when H_0 is true.

Exercise

Here (again) are the data on amount of wear in soles of boys shoes.

Boy	Material B	Material A	Difference
1	14.0	13.2	0.8
2	8.8	8.2	0.6
3	11.2	10.9	0.3
4	14.2	14.3	-0.1
5	11.8	10.7	1.1
6	6.4	6.6	-0.2
7	9.8	9.5	0.3
8	11.3	10.8	0.5
9	9.3	8.8	0.5
10	13.6	13.3	0.3
	$\bar{X} = 11.04$	$\bar{Y} = 10.63$	$\bar{D} = 0.41$
	$s_x = 2.52$	$s_y = 2.45$	$s_D = 0.39$

A histogram of the $n = 10$ **differences** suggests the **normality assumption** is tenable.

a) Carry out a **paired t test**, with $\alpha = 0.05$, to decide if there's **any difference** in wear for the two the materials.

a) Carry out a **paired** t **test**, with $\alpha = 0.05$, to decide if there's **any difference** in wear for the two the materials.

Hint: You should get t **= 3.324** and **p-value = 0.0089**.

a) Carry out a **paired t test**, with $\alpha = 0.05$, to decide if there's **any difference** in wear for the two the materials.

   Hint: You should get t **= 3.324** and **p-value = 0.0089**.

b) If you found a difference in Part *a*, which material is preferred?

- The **differences** $D_1, D_2, \ldots, D_n$ will be **normally distributed** if the X_i's and Y_i's are drawn from **normal** populations.

- The **differences** $D_1, D_2, \ldots, D_n$ will be **normally distributed** if the X_i's and Y_i's are drawn from **normal** populations.

Proposition

Suppose $X_i \sim \mathrm{N}(\mu_1, \sigma_1)$ and $Y_i \sim \mathrm{N}(\mu_2, \sigma_2)$. Let

$$D_i = X_i - Y_i\,.$$

Then

$$D_i \sim \mathrm{N}(\mu_d,\ \sigma_d)$$

where $\mu_d = \mu_1 - \mu_2$ (and σ_d is discussed later).

- The above fact holds because **linear combinations** of **normally distributed** random variables are **normally distributed** (and $X_i - Y_i$ is a linear combination of X_i and Y_i).

(Optional Section) Advantage of Matched Pairs Study Designs

Proposition

It can be shown that

$$\bar{D} = \bar{X} - \bar{Y},$$

where $\bar{D}$ is the sample mean of $D_1, D_2, \ldots, D_n$ and $\bar{X}$ and $\bar{Y}$ are the sample means of $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$.

- Compare the **paired** t and **two-sample** t **test statistics** (when $m = n$):

- Compare the **paired** t and **two-sample** t **test statistics** (when $m = n$):

Paired t Test Statistic:

$$T = \frac{\bar{X} - \bar{Y} - 0}{S_D/\sqrt{n}}$$

Two-Sample t Test Statistic:

$$T = \frac{\bar{X} - \bar{Y} - 0}{\sqrt{S_1^2/n + S_2^2/n}}$$

It can be shown that usually $S_D^2 < S_1^2 + S_2^2$

- Compare the **paired** t and **two-sample** t **test statistics** (when $m = n$):

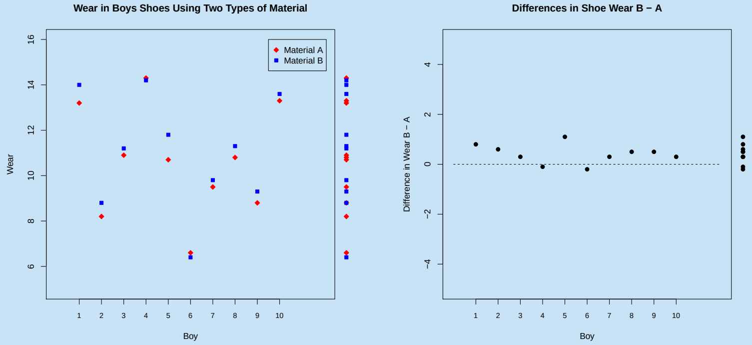
Paired t Test Statistic:	**Two-Sample t Test Statistic**:
$T = \frac{\bar{X}-\bar{Y}-0}{S_D/\sqrt{n}}$	$T = \frac{\bar{X}-\bar{Y}-0}{\sqrt{S_1^2/n+S_2^2/n}}$

It can be shown that usually $S_D^2 < S_1^2 + S_2^2$

(because $\sigma_d^2 = \sigma_1^2 + \sigma_2^2 - \rho\sigma_1\sigma_2$, where $0 < \rho < 1$).

Example

The figure below shows two representations of the data from the study comparing two materials for soles of boys shoes.



Compare:

Material A:	**Material B:**	**Differences**
$S_1^2 = 6.35$	$S_2^2 = 6.00$	$S_D^2 = 0.15$

Compare:

Two-Sample t:	**Paired t:**
$t = 0.37$	$t = 3.32$
df $= 17$	df $= 9$
p-value $= 0.7165$	p-value $= 0.0089$

Paired t Confidence Interval for $\mu_1 - \mu_2$

Paired t CI: When the differences $D_1, D_2, \ldots, D_n$ in **paired samples** can be treated as a sample from a population whose mean is μ_d $(= \mu_1 - \mu_2)$ a $100(1-\alpha)\%$ ***paired t confidence interval for $\mu_1 - \mu_2$*** is:

$$\bar{D} \pm t_{\alpha/2,n-1} \cdot \frac{S_D}{\sqrt{n}}.$$

- The CI is valid if either the sample of **differences** is from a **normal** population or n **is large**.

- The CI is valid if either the sample of **differences** is from a **normal** population or n **is large**.
- In either case, we can be $100(1-\alpha)\%$ confident that $\mu_1 - \mu_2$ will be contained in the CI.

- The CI is valid if either the sample of **differences** is from a **normal** population or n **is large**.
- In either case, we can be $100(1-\alpha)\%$ confident that $\mu_1 - \mu_2$ will be contained in the CI.
- The CI is just the **one-sample t CI** based on the **differences**.

Exercise

For the boys' shoe wear data set, the **sample mean** and **standard deviation** of the **differences** was

$$\bar{D} = 0.41 \qquad \text{and} \qquad S_D = 0.39.$$

a) Give the (point) **estimate** of the true difference in means $\mu_1 - \mu_2$ (or μ_d).

Exercise

For the boys' shoe wear data set, the **sample mean** and **standard deviation** of the **differences** was

$$\bar{D} = 0.41 \qquad \text{and} \qquad S_D = 0.39.$$

a) Give the (point) **estimate** of the true difference in means $\mu_1 - \mu_2$ (or μ_d).

b) Compute a 95% **paired t CI** for $\mu_1 - \mu_2$ (or μ_d).

Exercise

For the boys' shoe wear data set, the **sample mean** and **standard deviation** of the **differences** was

$$\bar{D} = 0.41 \qquad \text{and} \qquad S_D = 0.39.$$

a) Give the (point) **estimate** of the true difference in means $\mu_1 - \mu_2$ (or μ_d).

b) Compute a 95% **paired t CI** for $\mu_1 - \mu_2$ (or μ_d).

Hint: Using $t_{0.025,9} = 2.262$, you should get **(0.116, 0.704)**.

Exercise

For the boys' shoe wear data set, the **sample mean** and **standard deviation** of the **differences** was

$$\bar{D} = 0.41 \qquad \text{and} \qquad S_D = 0.39.$$

a) Give the (point) **estimate** of the true difference in means $\mu_1 - \mu_2$ (or μ_d).

b) Compute a 95% **paired t CI** for $\mu_1 - \mu_2$ (or μ_d).

   Hint: Using $t_{0.025,9} = 2.262$, you should get **(0.116, 0.704)**.

c) Does the CI contain the value **zero**? What does this say about the shoe wear for the two materials?