9 Paired Samples Hypothesis Tests

MTH 3240 Environmental Statistics

Spring 2020

Objectives

Objectives:

- Distinguish matched pairs study designs from independent samples designs.
- Carry out a paired t test for the difference between two population means.
- Compute and interpret paired t confidence interval for the difference between two population means.

Matched Pairs Studies

- In ***matched pairs study designs***, two samples are collected in such a way that each individual in one sample **matches** with one in the other sample.

Matched Pairs Studies

- In ***matched pairs study designs***, two samples are collected in such a way that each individual in one sample **matches** with one in the other sample.
- In environmental studies, they arise when a variable is measured at **concurrent time points** or **coincident spatial locations** under each of **two conditions.**

Example

Dredging a waterway refers to the removing sediment from its bottom, for example to deepen the waterway, improve water circulation, etc.

Example

Dredging a waterway refers to the removing sediment from its bottom, for example to deepen the waterway, improve water circulation, etc.

But dredging can alter sediment composition in ways that are detrimental to benthic (bottom dwelling) organisms.

A **matched pairs design** was used in a **before-after** study of the **impact** of **dredging** on **sediment composition** in the Rio Grande Harbor, Brazil.

A **matched pairs design** was used in a **before-after** study of the **impact** of **dredging** on **sediment composition** in the Rio Grande Harbor, Brazil.

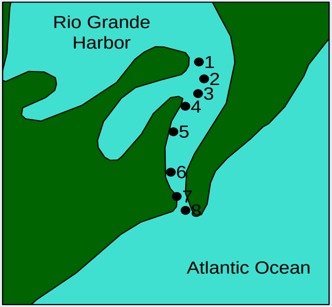
The **percent clay** in sediment was measured at $n = 8$ sites in the harbor **before** it was dredged and again at the **same eight sites after** dredging.

A **matched pairs design** was used in a **before-after** study of the **impact** of **dredging** on **sediment composition** in the Rio Grande Harbor, Brazil.

The **percent clay** in sediment was measured at $n = 8$ sites in the harbor **before** it was dredged and again at the **same eight sites after** dredging.

Each of the eight sites forms a **matched pair** of **before** and **after** measurements.

Locations of Sampling Stations in Dredging Impact Assessment Study



The data, including the ***differences*** for the eight sites, are below.

	Clay Percent		
Site	After Dredging	Before Dredging	Difference
1	53.8	61.3	-7.5
2	38.4	60.8	-22.4
3	54.1	49.4	4.7
4	55.7	56.2	-0.5
5	42.0	58.6	-16.6
6	48.1	57.1	-9.0
7	48.7	55.4	-6.7
8	19.3	48.3	-29.0
	$\bar{X} = 45.0$	$\bar{Y} = 55.9$	$\bar{D} = -10.9$
			$S_d = 11.2$.

Example

To assess the **impact** of **forest clear-cutting** on an adjacent stream's **water quality**, **nitrate** (mg/L) was measured on each of $n = 11$ days both **upstream** and **downstream** of a clear-cutting operation in Ireland.

Example

To assess the **impact** of **forest clear-cutting** on an adjacent stream's **water quality**, **nitrate** (mg/L) was measured on each of $n = 11$ days both **upstream** and **downstream** of a clear-cutting operation in Ireland.

Each of the 11 days forms a **matched pair** of **upstream** and **downstream** nitrate measurements.

Example

To assess the **impact** of **forest clear-cutting** on an adjacent stream's **water quality**, **nitrate** (mg/L) was measured on each of $n = 11$ days both **upstream** and **downstream** of a clear-cutting operation in Ireland.

Each of the 11 days forms a **matched pair** of **upstream** and **downstream** nitrate measurements.

The table below shows the data and the ***differences***.

Nitrate Concentration

Date	Upstream	Downstream	Difference
08/15/97	1147.4	995.3	152.1
08/18/97	1412.2	1303.6	108.6
08/31/97	1613.9	1923.3	-309.4
09/18/97	763.3	747.8	15.5
11/04/97	1031.4	1082.9	-51.5
11/07/97	1093.2	1938.7	-845.5
02/27/98	390.8	338.8	52.0
07/14/98	909.8	776.8	133.0
08/25/98	1033.0	676.8	356.2
09/30/98	897.5	1291.0	-393.5
10/29/98	2314.0	1232.9	1081.1
	$\bar{X} = 1146.0$	$\bar{Y} = 1118.9$	$\bar{D} = 27.1$
			$S_d = 480.7$.

- With **matched pairs studies**, we **control** for variables that **aren't measured** in the study by holding them **constant within each pair**.

- With **matched pairs studies**, we **control** for variables that **aren't measured** in the study by holding them **constant within each pair**.

  In the previous example, we **control** for **natural day-to-day variation** in variables that affect the stream's nitrate concentration by pairing **upstream** and **downstream** measurements **by day**.

- With **matched pairs studies**, we **control** for variables that **aren't measured** in the study by holding them **constant within each pair.**

  In the previous example, we **control** for **natural day-to-day variation** in variables that affect the stream's nitrate concentration by pairing **upstream** and **downstream** measurements **by day.**

  In the dredging example, we **control** for **spatial variation** in variables that affect the the harbor's clay percent by pairing **before** and **after** measurements **by location.**

- **Within** a **matched pair**, the two measured values are usually **similar** (compared to values for *unmatched* individuals).

- The *two-sample* t *test* is **not appropriate** for **paired samples ...**

- The *two-sample* t *test* is **not appropriate** for **paired samples ...**

  ... because it requires that the two samples be drawn **independently** of each other, ...

- The *two-sample t test* is **not appropriate** for **paired samples ...**

  ... because it requires that the two samples be drawn **independently** of each other, ...

  .... but in a **matched pairs study**, they're drawn in **pairs** (i.e. **not** independently).

Exercise

For each of the following studies, decide whether a **matched pairs** design or an **independent samples** design was used.

Exercise

For each of the following studies, decide whether a **matched pairs** design or an **independent samples** design was used.

a) In 1878 Charles Darwin performed an experiment to determine if the height of a *Zea mays* (corn) plant is affected by whether the plant is cross-fertilized or self-fertilized. In each of **15 pots**, two plants were grown, one **self-fertilized** and the other **cross-fertilized**, and their heights later measured.

a) In a study of the impact of agriculture on surface water quality, phosphorus was measured during each of **33 rainstorm events** at the outlets of two adjacent watersheds, one of which contains **farms** and the other **no farms.**

a) In a study of the impact of agriculture on surface water quality, phosphorus was measured during each of **33 rainstorm events** at the outlets of two adjacent watersheds, one of which contains **farms** and the other **no farms.**

b) In a study of the health hazards for workers in **two types** of swine confinement **buildings**, dried fecal matter was measured in the air of **12** randomly selected **finishing buildings**, where 50 - 100 kg animals are housed, and also in a random sample of **11 nursery buildings**, which house smaller animals.

c) On April 7, 2000, an **oil pipeline** owned by the Potomac Electric Power Company **ruptured**, spilling 126,000 gallons of oil into marsh areas of Swanson Creek, Maryland. To **assess** the **impact** on benthic (bottom dwelling) communities, benthos were evaluated at **10** randomly selected locations in **Swanson Creek** near the spill and at **10** other randomly selected locations in the nearby, undisturbed **Hunting Creek**.

c) On April 7, 2000, an **oil pipeline** owned by the Potomac Electric Power Company **ruptured**, spilling 126,000 gallons of oil into marsh areas of Swanson Creek, Maryland. To **assess** the **impact** on benthic (bottom dwelling) communities, benthos were evaluated at **10** randomly selected locations in **Swanson Creek** near the spill and at **10** other randomly selected locations in the nearby, undisturbed **Hunting Creek**.

d) In a study of the long-term **effect** on shoreline biology of **effluent** from an oil refinery at Littlewick Bay, Wales, barnacle densities were measured at **10** shoreline **locations near the refinery** in **1974** and again at the **same 10 locations** in **1981**.

e) A study was carried out by the University of Toronto to **assess** the **impact** of **effluent** from a sewage treatment plant near Orangeville, Ontario, Canada into the Credit River.

Fecal coliform was measured on **each of several days** 1.5 km **upstream** of the plant and on those *same days* 2.5 km **downstream**.

- We'll look at three tests for **paired samples**:

  1. The **paired t test**
  2. The **signed rank test**
  3. The **sign test for paired samples**

- We'll look at three tests for **paired samples**:

  1. The **paired t test**
  2. The **signed rank test**
  3. The **sign test for paired samples**

  The **paired t test** requires a **normality** assumption (or large sample sizes).

  The **signed rank test** and **sign test** are **nonparametric** tests (i.e. they **don't** rely on a normality assumption).

Paired t Test

- For the ***paired t test***, we suppose we have **paired samples**

$$X_1, X_2, \ldots, X_n \qquad \text{and} \qquad Y_1, Y_2, \ldots, Y_n$$

from **two populations** whose **means** are

$$\mu_x \qquad \text{and} \qquad \mu_y$$

Paired t Test

- For the ***paired t test***, we suppose we have **paired samples**

$$X_1, X_2, \ldots, X_n \qquad \text{and} \qquad Y_1, Y_2, \ldots, Y_n$$

from **two populations** whose **means** are

$$\mu_x \qquad \text{and} \qquad \mu_y$$

Here, X_1 and Y_1 are a **pair**, X_2 and Y_2 are a **pair**, and so on.

Paired t Test

- For the ***paired*** t ***test***, we suppose we have **paired samples**

  $$\boldsymbol{X_1, X_2, \ldots, X_n} \qquad \text{and} \qquad \boldsymbol{Y_1, Y_2, \ldots, Y_n}$$

  from **two populations** whose **means** are

  $$\boldsymbol{\mu_x} \qquad \text{and} \qquad \boldsymbol{\mu_y}$$

  Here, X_1 and Y_1 are a **pair**, X_2 and Y_2 are a **pair**, and so on.

- The sample size n is the **same** for the two samples.

- The **null hypothesis** is that there's **no difference** between μ_x and μ_y.

Null Hypothesis:

$$H_0 : \mu_x - \mu_y \ = \ 0.$$

- The **null hypothesis** is that there's **no difference** between $\boldsymbol{\mu_x}$ and $\boldsymbol{\mu_y}$.

Null Hypothesis:

$$H_0 : \mu_x - \mu_y \; = \; 0.$$

(Same hypothesis as for the two-sample t test.)

- The **alternative hypothesis** is one of the following.

Alternative Hypothesis:

1. $H_a : \mu_x - \mu_y > 0$ (**upper-tailed test**)
2. $H_a : \mu_x - \mu_y < 0$ (**lower-tailed test**)
3. $H_a : \mu_x - \mu_y \neq 0$ (**two-tailed test**)

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

- The **alternative hypothesis** is one of the following.

Alternative Hypothesis:

1. $H_a : \mu_x - \mu_y > 0$ (**upper-tailed test**)
2. $H_a : \mu_x - \mu_y < 0$ (**lower-tailed test**)
3. $H_a : \mu_x - \mu_y \neq 0$ (**two-tailed test**)

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

(Same hypotheses as for the two-sample t test.)

- We'll denote the ***differences*** by $\boldsymbol{D_1, D_2, \ldots, D_n}$, that is,

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

- We'll act as though these **differences** are a **random sample** from a ***population of differences*** whose mean is μ_d.

- We'll act as though these **differences** are a **random sample** from a ***population of differences*** whose mean is μ_d.
- The ***paired*** t ***test*** is just a **one-sample** t **test** for μ_d based on the **differences.**

Fact:

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

where $\bar{X}$ and $\bar{Y}$ are the means of the X and Y samples and $\bar{D}$ is the mean of the sample of differences.

Fact:

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

where $\bar{X}$ and $\bar{Y}$ are the means of the X and Y samples and $\bar{D}$ is the mean of the sample of differences.

"The mean of the differences equals the difference between the means"

Fact:

$$\mu_d = \mu_x - \mu_y.$$

where $\boldsymbol{\mu_x}$ and $\boldsymbol{\mu_y}$ are the means of the X and Y populations and $\boldsymbol{\mu_d}$ is the mean of the population of differences.

Fact:

$$\mu_d = \mu_x - \mu_y.$$

where μ_x and μ_y are the means of the X and Y populations and μ_d is the mean of the population of differences.

”The mean of the differences is the difference between the means”:

Thus the hypotheses can be restated in terms of $\boldsymbol{\mu_d}$ as:

	Hypothesis About $\boldsymbol{\mu_x - \mu_y}$	**Equivalent Hypothesis About $\boldsymbol{\mu_d}$**
Null	$H_0 : \mu_x - \mu_y = 0$	$H_0 : \mu_d = 0$
Alternatives	$H_a : \mu_x - \mu_y > 0$	$H_a : \mu_d > 0$
	$H_a : \mu_x - \mu_y < 0$	$H_a : \mu_d < 0$
	$H_a : \mu_x - \mu_y \neq 0$	$H_a : \mu_d \neq 0$

- In a **matched pairs study**, the ***effect size*** is

$$\mu_x - \mu_y = \mu_d.$$

- In a **matched pairs study**, the ***effect size*** is

$$\mu_x - \mu_y = \mu_d.$$

It's **estimated** by

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

Paired t Test Statistic:

$$t \;=\; \frac{\bar{D} \;-\; 0}{S_{\bar{D}}} \;=\; \frac{\bar{D}}{S_{\bar{D}}}$$

where

$$S_{\bar{D}} \;=\; \frac{S_d}{\sqrt{n}}\,,$$

and $\bar{D}$ and S_d are the sample mean and sample standard deviation of the **differences**.

Paired t Test Statistic:

$$t = \frac{\bar{D} - 0}{S_{\bar{D}}} = \frac{\bar{D}}{S_{\bar{D}}}$$

where

$$S_{\bar{D}} = \frac{S_d}{\sqrt{n}},$$

and $\bar{D}$ and S_d are the sample mean and sample standard deviation of the **differences**.

- Note that t is just the **one-sample t test statistic** based on the **differences.**

1. *Large positive* values of t provide evidence in favor of $H_a : \mu_x - \mu_y > 0$ (or $H_a : \mu_d > 0$).
2. *Large negative* values of t provide evidence in favor of $H_a : \mu_x - \mu_y < 0$ (or $H_a : \mu_d < 0$).
3. *Both large positive and large negative* values of t provide evidence in favor of $H_a : \mu_x - \mu_y \neq 0$ (or $H_a : \mu_d \neq 0$).

- Now suppose either the sample of **differences** is from **normal population** or n is **large.**

- Now suppose either the sample of **differences** is from **normal population** or n is **large.**

  In this case, the **null distribution** is as follows.

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

$$H_0 : \mu_x - \mu_y \;=\; 0 \qquad \text{(or equivalently } H_0 : \mu_d \;=\; 0\text{)}$$

is true,

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

- **P-values** and **rejection regions** are obtained from the appropriate tail(s) of the $t(n-1)$ **distribution.**

Paired t Test for μ_d

Assumptions: $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ are two random samples that are paired and either the differences $d_1, d_2, \ldots, d_n$ form a single sample from a *normal* population or n is large.

Null hypothesis: $H_0 : \mu_d = 0$.

Test statistic value: $t = \frac{\bar{d}}{s_d/\sqrt{n}}$.

Decision rule: Reject H_0 if p-value $< \alpha$ or t is in rejection region.

Null hypothesis: $H_0 : \mu_d = 0$.

Test statistic value: $t = \frac{\bar{d}}{s_d/\sqrt{n}}$.

Decision rule: Reject H_0 if p-value $< \alpha$ or t is in rejection region.

Alternative hypothesis	P-value = area under t-distribution with $n-1$ d.f.:	Rejection region = t values such that:*
$H_a : \mu_d > 0$	to the right of t	$t > t_{\alpha,n-1}$
$H_a : \mu_d < 0$	to the left of t	$t < -t_{\alpha,n-1}$
$H_a : \mu_d \neq 0$	to the left of $-\lvert t\rvert$ and right of $\lvert t\rvert$	$t > t_{\alpha/2,n-1}$ or $t < -t_{\alpha/2,n-1}$

* $t_{\alpha,n-1}$ is the $100(1-\alpha)$th percentile of the t distribution with $n-1$ d.f.

Exercise

For the study of the impact of dredging on the Brazilian harbor, we want to decide if there was **any change** in the sediment's clay percentage.

Exercise

For the study of the impact of dredging on the Brazilian harbor, we want to decide if there was **any change** in the sediment's clay percentage.

The summary statistics for the $n = 8$ differences (clay percentage **after** dredging minus **before**) are

$$\bar{D} \ = \ -10.9 \qquad \text{and} \qquad S_d \ = \ 11.2$$

The hypotheses are

$$H_0 : \mu_x - \mu_y \ = \ 0 \qquad \text{(or equivalently } H_0 : \mu_d \ = \ 0)$$

$$H_a : \mu_x - \mu_y \neq 0 \qquad \text{(or equivalently } H_0 : \mu_d \ \neq \ 0)$$

The hypotheses are

$$H_0 : \mu_x - \mu_y \ = \ 0 \qquad \text{(or equivalently } H_0 : \mu_d \ = \ 0)$$

$$H_a : \mu_x - \mu_y \neq 0 \qquad \text{(or equivalently } H_0 : \mu_d \ \neq \ 0)$$

where $\boldsymbol{\mu_x}$ and $\boldsymbol{\mu_y}$ are the true population mean **clay percents after** and **before** dredging, respectively, and $\boldsymbol{\mu_d}$ is population mean **difference.**

Carry out the **paired t test** using level of significance $\alpha = 0.05$.

Carry out the **paired t test** using level of significance $\alpha = 0.05$.

Hints: You should get $\boldsymbol{t = -2.75}$ and **p-value** = 2(0.013) = **0.026**.

Paired t Confidence Interval

- Recall that in a **matched pairs study**, we **estimate** the ***effect size*** $\mu_x - \mu_y$ (or μ_d) by $\bar{X} - \bar{Y}$ (or $\bar{D}$).

Paired t Confidence Interval

- Recall that in a **matched pairs study**, we **estimate** the ***effect size*** $\mu_x - \mu_y$ (or μ_d) by $\bar{X} - \bar{Y}$ (or $\bar{D}$).
- By attaching a **margin of error** to the estimate, we get a **CI**.

Paired t CI: A $100(1-\alpha)\%$ ***paired t CI for $\mu_x - \mu_y$*** (or μ_d) is

$$\bar{D} \pm t_{\alpha/2,n-1} \, S_{\bar{D}} \, ,$$

where

$$S_{\bar{D}} \; = \; \frac{S_d}{\sqrt{n}} \, .$$

Paired t CI: A $100(1-\alpha)\%$ ***paired t CI for $\boldsymbol{\mu_x - \mu_y}$*** (or $\boldsymbol{\mu_d}$) is

$$\bar{D} \pm t_{\alpha/2,n-1}\, S_{\bar{D}}\,,$$

where

$$S_{\bar{D}} \;=\; \frac{S_d}{\sqrt{n}}\,.$$

- Note that this is just the **one-sample t CI** based on the **differences.**

- The CI is valid if either
    1. The population of **differences** is **normal**, or
    2. The sample size n is **large**.

- The CI is valid if either
  1. The population of **differences** is **normal**, or
  2. The sample size n is **large**.
- We can be $100(1-\alpha)\%$ confident that the true (unknown) effect size $\mu_x - \mu_y$ (or μ_d) will be contained in the interval.

- The "plus or minus" part is the ***margin of error***.

- The ”plus or minus” part is the ***margin of error***.

Margin of Error: For the paired t CI,

$$\text{Margin of Error} \ = \ t_{\alpha/2,n-1}\, S_{\bar{D}}\,,$$

where

$$S_{\bar{D}} \ = \ \frac{S_d}{\sqrt{n}}\,.$$

- The ”plus or minus” part is the ***margin of error***.

Margin of Error: For the paired t CI,

$$\text{Margin of Error} = t_{\alpha/2,n-1}\, S_{\bar{D}}\,,$$

where

$$S_{\bar{D}} = \frac{S_d}{\sqrt{n}}\,.$$

- A **smaller margin of error** indicates that $\bar{X} - \bar{Y}$ (or $\bar{D}$) is a **more precise estimate** of the (unknown) **effect size** $\mu_x - \mu_y$ (or μ_d).

Example

For the study of dredging in the Brazilian harbor, recall that the summary statistics for the $n = 8$ **differences** are

$$\bar{D} = -10.9 \qquad \text{and} \qquad S_d = 11.2.$$

The **estimated size** of the **effect** of dredging on the sediment's **clay percentage** is

$$\bar{X} - \bar{Y} \;=\; \bar{D} \;=\; \mathbf{-10.9},$$

The **estimated size** of the **effect** of dredging on the sediment's **clay percentage** is

$$\bar{X} - \bar{Y} \;=\; \bar{D} \;=\; \mathbf{-10.9},$$

i.e. a **decrease** of **10.9** percentage points in the clay.

The **estimated size** of the **effect** of dredging on the sediment's **clay percentage** is

$$\bar{X} - \bar{Y} \;=\; \bar{D} \;=\; \mathbf{-10.9},$$

i.e. a **decrease** of **10.9** percentage points in the clay.

The **standard error** of the estimate $\bar{D}$ is

$$S_{\bar{D}} \;=\; \frac{S_d}{\sqrt{n}} \;=\; \frac{11.2}{\sqrt{8}} \;=\; \mathbf{3.96}.$$

A **95% paired t Ci** for the true (unknown) **effect size $\mu_x - \mu_y$** (or μ_d) is

A **95% paired t Ci** for the true (unknown) **effect size $\mu_x - \mu_y$** (or μ_d) is

$$\begin{aligned}\bar{D} \pm t_{\alpha/2,n-1} S_{\bar{D}} &= -10.9 \pm 2.36\,(3.96) \\ &= -10.9 \pm 9.35 \\ &= \mathbf{(-20.25,\ -1.55)}\end{aligned}$$

(where the t critical value $t_{0.025,7} = 2.36$ was obtained from a t distribution table using $n - 1 = 7$ df).

A **95% paired t Ci** for the true (unknown) **effect size $\boldsymbol{\mu_x - \mu_y}$** (or $\boldsymbol{\mu_d}$) is

$$\begin{aligned} \bar{D} \pm t_{\alpha/2,n-1}\, S_{\bar{D}} &= -10.9 \pm 2.36\,(3.96) \\ &= -10.9 \pm 9.35 \\ &= \mathbf{(-20.25,\ -1.55)} \end{aligned}$$

(where the t critical value $t_{0.025,7} = 2.36$ was obtained from a t distribution table using $n - 1 = 7$ df).

The **margin of error** in the estimate is **9.35** percentage points.

We can be **95% confident** that the true (unknown) **effect** of dredging is a **decrease** in clay of between **1.55** and **20.25** percentage points.