8 Two-Sample Hypothesis Tests

MTH 3240 Environmental Statistics

Spring 2020

Objectives

Objectives:

- Carry out a two-sample t test for the difference between two population means.
- Compute and interpret two-sample t confidence interval for the difference between two population means.

Introduction

- We're often interested in testing for a difference between **two** population means μ_x and μ_y.

Introduction

- We're often interested in testing for a difference between **two** population means μ_x and μ_y.
- Here are three examples.

Example

In a **control-impact** study, we might suspect that the mean contaminant level is higher at the **impact site** than at the **control site**. The hypotheses would be

$$H_0 : \mu_x = \mu_y$$
$$H_a : \mu_x > \mu_y$$

where μ_x and μ_y are the true (unknown) **population mean** contaminant levels at the **impact** and **control sites**, respectively.

Example

In a **control-impact** study, we might suspect that the mean contaminant level is higher at the **impact site** than at the **control site**. The hypotheses would be

$$\begin{aligned} H_0 : \mu_x &= \mu_y \\ H_a : \mu_x &> \mu_y \end{aligned}$$

where $\boldsymbol{\mu_x}$ and $\boldsymbol{\mu_y}$ are the true (unknown) **population mean** contaminant levels at the **impact** and **control sites**, respectively.

$\boldsymbol{H_0}$ says there's **no difference** between the two sites' means, and $\boldsymbol{H_a}$ says the **impact site's** mean is **higher.**

Example

In a **before-after** study, we might suspect that the site became contaminated as a result of the impact event. The hypotheses would be

$$\begin{aligned} H_0 : \mu_x &= \mu_y \\ H_a : \mu_x &> \mu_y \end{aligned}$$

where $\boldsymbol{\mu_x}$ and $\boldsymbol{\mu_y}$ are the true (unknown) mean contaminant levels **after** and **before** the impact event, respectively.

Example

In a **before-after** study, we might suspect that the site became contaminated as a result of the impact event. The hypotheses would be

$$\begin{aligned} H_0: \mu_x &= \mu_y \\ H_a: \mu_x &> \mu_y \end{aligned}$$

where μ_x and μ_y are the true (unknown) mean contaminant levels **after** and **before** the impact event, respectively.

H_0 says the impact event had **no effect** on the contaminant levels, and H_a says it **increased** them.

Example

In an **experiment**, we might randomly assign experimental units to **treatment** and **control** conditions and compare their responses. To decide if there's **any difference** in the effects of the two conditions, we'd test

$$\begin{aligned} H_0 : \mu_x &= \mu_y \\ H_a : \mu_x &\neq \mu_y \end{aligned}$$

where $\boldsymbol{\mu_x}$ and $\boldsymbol{\mu_y}$ are the true (unknown) **population mean responses** to the two conditions.

Example

In an **experiment**, we might randomly assign experimental units to **treatment** and **control** conditions and compare their responses. To decide if there's **any difference** in the effects of the two conditions, we'd test

$$
\begin{aligned}
H_0 : \mu_x &= \mu_y \\
H_a : \mu_x &\neq \mu_y
\end{aligned}
$$

where μ_x and μ_y are the true (unknown) **population mean responses** to the two conditions.

H_0 says there's **no difference** in the mean responses for the two conditions, and H_a says there's **a difference**.

- We'll look at two tests for comparing two population means:

  1. The **two-sample t test**
  2. The **rank sum test**

- We'll look at two tests for comparing two population means:

  1. The **two-sample t test**
  2. The **rank sum test**

  The t **test** requires a **normality** assumption (or large sample sizes), but the **rank sum test** is a **nonparametric** test (doesn't require normality).

Two-Sample t Test

- For the ***two-sample*** t ***test***, we suppose we have random samples

$$X_1, X_2, \ldots, X_{n_x} \qquad \text{and} \qquad Y_1, Y_2, \ldots, Y_{n_y}$$

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

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

Two-Sample t Test

- For the ***two-sample*** t ***test***, we suppose we have random samples

$$X_1, X_2, \ldots, X_{n_x} \qquad \text{and} \qquad Y_1, Y_2, \ldots, Y_{n_y}$$

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

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

- The **sample sizes**

$$n_x \qquad \text{and} \qquad n_y$$

don't have to be the same.

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

Null Hypothesis:

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

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

Alternative Hypothesis:

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

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

- We'll usually write these hypotheses as

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

and

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**)

- We'll usually write these hypotheses as

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

and

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**)

- The difference $\boldsymbol{\mu_x - \mu_y}$ between the population means is sometimes called the ***effect size***.

- We'll usually write these hypotheses as

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

and

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**)

- The difference $\boldsymbol{\mu_x - \mu_y}$ between the population means is sometimes called the ***effect size***.

  If μ_x and μ_y are true (unknown) mean responses to **treatment** and **control** conditions in an experiment, $\boldsymbol{H_0}$ says the treatment has **no effect.**

- We'll denote the **sample means** by

$$\bar{X} \qquad \text{and} \qquad \bar{Y}$$

and the **sample standard deviations** by

$$S_x \qquad \text{and} \qquad S_y$$

- We'll denote the **sample means** by

$$\bar{X} \qquad \text{and} \qquad \bar{Y}$$

  and the **sample standard deviations** by

$$S_x \qquad \text{and} \qquad S_y$$

- We **estimate** the **effect size** by the difference between the sample means, $\bar{X} - \bar{Y}$.

Two-Sample t Test Statistic:

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

where

$$S_{\bar{X}-\bar{Y}} \;=\; \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}\,.$$

Two-Sample t Test Statistic:

$$t = \frac{\bar{X} - \bar{Y} - 0}{S_{\bar{X}-\bar{Y}}} = \frac{\bar{X} - \bar{Y}}{S_{\bar{X}-\bar{Y}}}$$

where

$$S_{\bar{X}-\bar{Y}} = \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}\,.$$

- The denominator of t, $S_{\bar{X}-\bar{Y}}$, is an estimate of the **standard error** of the statistic $\bar{X} - \bar{Y}$.

Two-Sample t Test Statistic:

$$t = \frac{\bar{X} - \bar{Y} - 0}{S_{\bar{X}-\bar{Y}}} = \frac{\bar{X} - \bar{Y}}{S_{\bar{X}-\bar{Y}}}$$

where

$$S_{\bar{X}-\bar{Y}} = \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}\,.$$

- The denominator of t, $\boldsymbol{S_{\bar{X}-\bar{Y}}}$, is an estimate of the **standard error** of the statistic $\boldsymbol{\bar{X} - \bar{Y}}$.
- So t indicates how many **standard errors** the **estimated effect size** $\bar{X} - \bar{Y}$ is **away** from **zero**.

- $\bar{X} - \bar{Y}$ is an estimate of $\mu_x - \mu_y$, so ...

- $\bar{X} - \bar{Y}$ is an estimate of $\mu_x - \mu_y$, so ...
  - **If H_0 was true**, ...

- $\bar{X} - \bar{Y}$ is an estimate of $\mu_x - \mu_y$, so ...
  - **If H_0 was true**, ...

    ... we'd expect $\bar{X} - \bar{Y}$ to be close to **zero**.

- $\bar{X} - \bar{Y}$ is an estimate of $\mu_x - \mu_y$, so ...
  - **If H_0 was true**, ...

    ... we'd expect $\bar{X} - \bar{Y}$ to be close to **zero**.
  - But **if H_a was true**, ...

- $\bar{X} - \bar{Y}$ is an estimate of $\mu_x - \mu_y$, so ...
  - **If H_0 was true**, ...

    ... we'd expect $\bar{X} - \bar{Y}$ to be close to **zero**.
  - But **if H_a was true**, ...

    ... we'd expect $\bar{X} - \bar{Y}$ to differ from zero in the direction specified by H_a.

- $\bar{X} - \bar{Y}$ is an estimate of $\mu_x - \mu_y$, so ...
  - **If H_0 was true**, ...

    ... we'd expect $\bar{X} - \bar{Y}$ to be close to **zero**.
  - But **if H_a was true**, ...

    ... we'd expect $\bar{X} - \bar{Y}$ to differ from zero in the direction specified by H_a.
- Thus ...

- $\bar{X} - \bar{Y}$ is an estimate of $\mu_x - \mu_y$, so ...
  - **If H_0 was true**, ...

    ... we'd expect $\bar{X} - \bar{Y}$ to be close to **zero**.
  - But **if H_a was true**, ...

    ... we'd expect $\bar{X} - \bar{Y}$ to differ from zero in the direction specified by H_a.
- Thus ...
  1. t will be approximately **zero** (most likely) if H_0 is true.
  2. It will **differ from zero** (most likely) in the direction specified by H_a if H_a is true.

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

- Now suppose we have random samples from two populations.

- Now suppose we have random samples from two populations.

  If either

  1. The populations are both normal, or
  2. The sample sizes n_x and n_y are both large,

- Now suppose we have random samples from two populations.

  If either

  1. The populations are both normal, or
  2. The sample sizes n_x and n_y are both large,

  the **null distribution** is as follows.

Sampling Distribution of t Under H_0: If t is the two-sample t test statistic, then when

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

is true,

$$t \sim t(\text{df}),$$

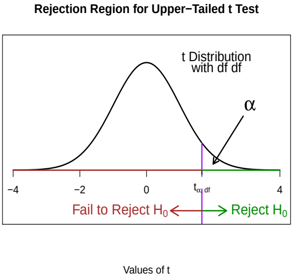
where the **df** are

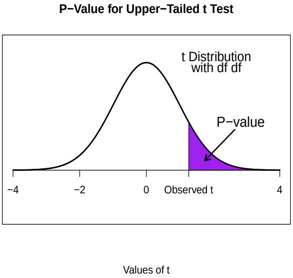
$$\text{df} = \frac{\left(\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}\right)^2}{\frac{(S_x^2/n_x)^2}{n_x - 1} + \frac{(S_y^2/n_y)^2}{n_y - 1}},$$

(which is rounded *down* to the nearest integer).

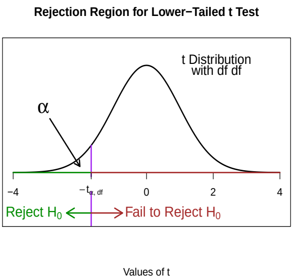
- **P-values** and **rejection regions** are obtained from the appropriate tail(s) of the t(**df**) **distribution**, as shown on the next slides.

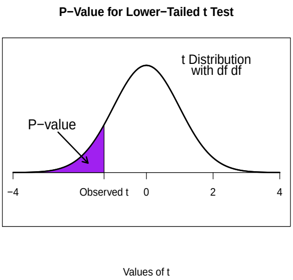
1. $H_a : \mu_x - \mu_y > 0$ (Upper-Tailed Test)



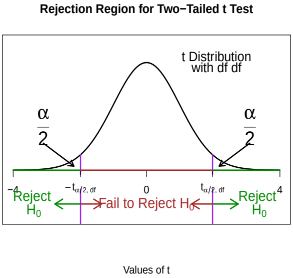


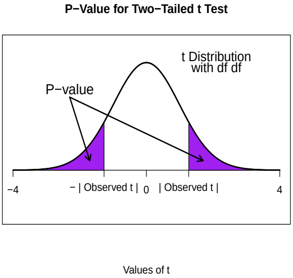
2. $H_a : \mu_x - \mu_y < 0$ (Lower-Tailed Test)





3. $H_a : \mu_x - \mu_y \neq 0$ (Two-Tailed Test)





Two-Sample t Test for μ_x and μ_y

Assumptions: The data $x_1, x_2, \ldots, x_{n_x}$ and $y_1, y_2, \ldots, y_{n_y}$ are independent random samples from two populations and either the populations are normal or n_x and n_y are large.

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

Test statistic value: $t = \frac{\bar{x}-\bar{y}}{\sqrt{s_x^2/n_x + s_y^2/n_y}}$.

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

Two-Sample t Test for μ_x and μ_y

Alternative hypothesis	P-value = area under t distribution with d.f. given by df:	Rejection region = t values such that:*
$H_a : \mu_x - \mu_y > 0$	to the right of t	$t > t_{\alpha,\text{df}}$
$H_a : \mu_x - \mu_y < 0$	to the left of t	$t < -t_{\alpha,\text{df}}$
$H_a : \mu_x - \mu_y \neq 0$	to the left of $-\|t\|$ and right of $\|t\|$	$t > t_{\alpha/2,\text{df}}$ or $t < -t_{\alpha/2,\text{df}}$

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

Exercise

To assess the impact of a wastewater treatment plant's effluent discharge into the Febros River, Portugal, gudgeon fish (*Gobio gobio*) were sampled **upstream** and **downstream** of the plant and their **weights** (g) measured.

Exercise

To assess the impact of a wastewater treatment plant's effluent discharge into the Febros River, Portugal, gudgeon fish (*Gobio gobio*) were sampled **upstream** and **downstream** of the plant and their **weights** (g) measured.

The water was warmer downstream than upstream, possibly due to the effluent discharge, and fish are more active and have higher metabolic activity in warmer water, which can lead to growth.

Exercise

To assess the impact of a wastewater treatment plant's effluent discharge into the Febros River, Portugal, gudgeon fish (*Gobio gobio*) were sampled **upstream** and **downstream** of the plant and their **weights** (g) measured.

The water was warmer downstream than upstream, possibly due to the effluent discharge, and fish are more active and have higher metabolic activity in warmer water, which can lead to growth.

Also, the nutrient load was higher downstream, and this can lead to more available food for the fish.

For these reasons, the researchers hypothesized that the fish would be **larger downstream than upstream**.

For these reasons, the researchers hypothesized that the fish would be **larger downstream than upstream**.

The summary statistics are below.

For these reasons, the researchers hypothesized that the fish would be **larger downstream than upstream.**

The summary statistics are below.

Gudgeon Weights

Upstream	Downstream
$n_x = 23$	$n_y = 22$
$\bar{X} = 12.12$	$\bar{Y} = 16.93$
$S_x = 2.57$	$S_y = 4.17$

For these reasons, the researchers hypothesized that the fish would be **larger downstream than upstream.**

The summary statistics are below.

Gudgeon Weights

	Upstream	Downstream
	$n_x = 23$	$n_y = 22$
	$\bar{X} = 12.12$	$\bar{Y} = 16.93$
	$S_x = 2.57$	$S_y = 4.17$

Carry out a **two-sample t test** to decide if the population mean gudgeon weight is **greater downstream than upstream.** Use a level of significance $\alpha = 0.05$.

Hints: The degrees of freedom are

$$\text{df} = \frac{\left(\frac{2.57^2}{23} + \frac{4.17^2}{22}\right)^2}{\frac{(2.57^2/23)^2}{23-1} + \frac{(4.17^2/22)^2}{22-1}} = \mathbf{34},$$

and you should get $t = \mathbf{-4.63}$ and **p-value = 0.0000**.

Two-Sample t Confidence Interval

- We can compute a ***two-sample*** t ***confidence interval*** for the true (unknown) **effect size** $\mu_x - \mu_y$.

Two-Sample t Confidence Interval

- We can compute a ***two-sample t confidence interval*** for the true (unknown) **effect size $\mu_x - \mu_y$.**

Two-Sample t CI: A $100(1-\alpha)\%$ ***two-sample t CI for*** $\mu_x - \mu_y$ is

$$\bar{X} - \bar{Y} \ \pm \ t_{\alpha/2,\text{df}}\, S_{\bar{X}-\bar{Y}}$$

where

$$S_{\bar{X}-\bar{Y}} \ = \ \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}} \, ,$$

and the **df** are the same as for the two-sample t test.

- The CI is valid if either
  1. The two populations are both **normal**, or
  2. The sample sizes n_x and n_y are both **large**.

- The CI is valid if either
  1. The two populations are both **normal**, or
  2. The sample sizes n_x and n_y are both **large**.
- We can be $100(1-\alpha)\%$ confident that the true (unknown) effect size $\mu_x - \mu_y$ will be contained in the interval.

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

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

Margin of Error: For the two-sample t CI,

$$\text{Margin of Error} \;=\; t_{\alpha/2,\text{df}}\, S_{\bar{X}-\bar{Y}}$$

where

$$S_{\bar{X}-\bar{Y}} \;=\; \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}\,.$$

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

Margin of Error: For the two-sample t CI,

$$\text{Margin of Error} \; = \; t_{\alpha/2,\text{df}} \, S_{\bar{X}-\bar{Y}}$$

where

$$S_{\bar{X}-\bar{Y}} \; = \; \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}\,.$$

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

Example

For the study of impact of the wastewater treatment plant's discharge into the Febros River on fish weights, recall that the summary statistics are:

Gudgeon Weights

Upstream	Downstream
$n_x = 23$	$n_y = 22$
$\bar{X} = 12.12$	$\bar{Y} = 16.93$
$S_x = 2.57$	$S_y = 4.17$

The **estimate** of the true (unknown) **effect size** $\mu_x - \mu_y$ is

$$\bar{X} - \bar{Y} = 12.12 - 16.93 = \mathbf{-4.81},$$

The **estimate** of the true (unknown) **effect size** $\mu_x - \mu_y$ is

$$\bar{X} - \bar{Y} = 12.12 - 16.93 = \mathbf{-4.81},$$

i.e. we estimate that gudgeon weights are **4.81 g** heavier **downstream** than **upstream**, on average.

The estimated **standard error** of the statistic $\bar{X} - \bar{Y}$ is

The estimated **standard error** of the statistic $\bar{X} - \bar{Y}$ is

$$S_{\bar{X}-\bar{Y}} = \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}} = \sqrt{\frac{2.57^2}{23} + \frac{4.17^2}{22}} = \mathbf{1.04}.$$

The estimated **standard error** of the statistic $\bar{X} - \bar{Y}$ is

$$S_{\bar{X}-\bar{Y}} = \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}} = \sqrt{\frac{2.57^2}{23} + \frac{4.17^2}{22}} = \mathbf{1.04}.$$

so **95% two-sample t CI** for $\boldsymbol{\mu_x - \mu_y}$ is

The estimated **standard error** of the statistic $\bar{X} - \bar{Y}$ is

$$S_{\bar{X}-\bar{Y}} = \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}} = \sqrt{\frac{2.57^2}{23} + \frac{4.17^2}{22}} = \mathbf{1.04}.$$

so **95% two-sample t CI** for $\boldsymbol{\mu_x - \mu_y}$ is

$$\begin{aligned} \bar{X} - \bar{Y} \pm t_{\alpha/2,\text{df}}\, S_{\bar{X}-\bar{Y}} &= -4.81 \pm 2.032\,(1.04) \\ &= -4.81 \pm 2.11 \\ &= \mathbf{(-6.92,\ -2.70)} \end{aligned}$$

The estimated **standard error** of the statistic $\bar{X} - \bar{Y}$ is

$$S_{\bar{X}-\bar{Y}} = \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}} = \sqrt{\frac{2.57^2}{23} + \frac{4.17^2}{22}} = \mathbf{1.04}.$$

so **95% two-sample t CI** for $\boldsymbol{\mu_x - \mu_y}$ is

$$\begin{aligned} \bar{X} - \bar{Y} \pm t_{\alpha/2,\text{df}}\, S_{\bar{X}-\bar{Y}} &= -4.81 \pm 2.032\,(1.04) \\ &= -4.81 \pm 2.11 \\ &= \mathbf{(-6.92,\ -2.70)} \end{aligned}$$

(where the t critical value $t_{0.025,\text{df}} = 2.032$ was obtained from a t distribution table and the df is 34 from the previous exercise).

We're **95% confident** that the true (unknown) effect of the effluent discharge on gudgeon weights is an **increase** of between **2.70** and **6.92 g**.

We're **95% confident** that the true (unknown) effect of the effluent discharge on gudgeon weights is an **increase** of between **2.70** and **6.92 g**.

From the previous slide, the **margin of error** in the **estimate** of the effect size is **2.11 g**.