Statistical Methods

Nels Grevstad

Metropolitan State University of Denver

ngrevsta@msudenver.edu

August 29, 2019

Topics

1. Equivalency Between CIs and Hypothesis Tests
2. Two-Sample t Test for Two Population Means μ_1 and μ_2
3. Two-Sample t Confidence Interval for $\mu_1 - \mu_2$

Objectives

Objectives:

- State the equivalency between confidence intervals and hypothesis tests.
- Carry out a two-sample t test for two population means.
- Compute and interpret a two-sample t CI for the difference between two population means.

Equivalency Between CIs and Hypotheses Tests

- **CIs** can be used to **test hypotheses.**

Equivalency Between CIs and Hypotheses Tests

- **CIs** can be used to **test hypotheses.**

Using CIs to Test Hypotheses: A CI for a parameter θ with **level of confidence** $100(1-\alpha)$**%** can be used to test the hypotheses

$$\begin{aligned} H_0: \theta &= \theta_0 \\ H_a: \theta &\neq \theta_0 \end{aligned}$$

with **significance level** α by invoking the following decision rule:

Reject H_0 if the CI **doesn't** contain θ_0
Fail to reject H_0 if it **does** contain θ_0

Reject H_0 if the CI **doesn't** contain θ_0
Fail to reject H_0 if it **does** contain θ_0

- The **CI** approach will always reach the **same conclusion** as the associated **hypothesis test.**

Exercise

A **95% confidence interval** for a population mean μ is

$$(21,\ 34)\,.$$

Exercise

A **95% confidence interval** for a population mean μ is

$$(21,\ 34)\,.$$

a) Could you reject the null hypothesis in a **test** of $H_0 : \mu = 23$ versus $H_a : \mu \neq 23$ at the **5% significance level**? Explain.

Exercise

A **95% confidence interval** for a population mean μ is

$$(21, \ 34).$$

a) Could you reject the null hypothesis in a **test** of $H_0 : \mu = 23$ versus $H_a : \mu \neq 23$ at the **5% significance level**? Explain.

b) Could you reject the null hypothesis in a **test** of $H_0 : \mu = 19$ versus $H_a : \mu \neq 19$ at the **5% significance level**? Explain.

Exercise

For a test of

$$\begin{aligned} H_0 : \boldsymbol{\mu} &= \mathbf{50} \\ H_a : \boldsymbol{\mu} &\neq \mathbf{50} \end{aligned}$$

the **p-value** is **0.16.** Thus H_0 would **not be rejected** at either the **5%** or **10% significance levels**.

Exercise

For a test of

$$\begin{aligned} H_0 : \boldsymbol{\mu} &= \mathbf{50} \\ H_a : \boldsymbol{\mu} &\neq \mathbf{50} \end{aligned}$$

the **p-value** is **0.16.** Thus H_0 would **not be rejected** at either the **5%** or **10% significance levels**.

a) Would a **95% confidence interval** for μ contain the value **50**? Explain.

Exercise

For a test of

$$\begin{aligned} H_0 : \boldsymbol{\mu} &= \mathbf{50} \\ H_a : \boldsymbol{\mu} &\neq \mathbf{50} \end{aligned}$$

the **p-value** is **0.16.** Thus H_0 would **not be rejected** at either the **5%** or **10% significance levels**.

a) Would a **95% confidence interval** for μ contain the value **50**? Explain.

b) Would a **90% confidence interval** for μ contain the value **50**? Explain.

- To see why the **CI** approach and **hypothesis test** reach the **same conclusion**, consider a **one-sample t test** of

$$\begin{aligned} H_0 : \mu &= \mu_0 \\ H_a : \mu &\neq \mu_0 \end{aligned}$$

The **rejection region approach**, with $\alpha = 0.05$, says

Reject H_0 if $t < -t_{0.025,\, n-1}$ or $t > t_{0.025,\, n-1}$.
Fail to reject H_0 otherwise.

- In other words, we **fail to reject H_0** if

$$t_{0.025,\,n-1} \;\leq\; t \;\leq\; t_{0.025,\,n-1}\,.$$

- In other words, we **fail to reject H_0** if

$$t_{0.025,\,n-1} \;\leq\; t \;\leq\; t_{0.025,\,n-1}\,.$$

Plugging

$$t \;=\; \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

in for t above, and solving for μ_0, we **fail to reject H_0** if

- In other words, we **fail to reject H_0** if

$$t_{0.025,\, n-1} \;\le\; t \;\le\; t_{0.025,\, n-1}\,.$$

Plugging

$$t \;=\; \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

in for t above, and solving for μ_0, we **fail to reject H_0** if

$$\bar{X} \;-\; t_{0.025,\, n-1}\frac{S}{\sqrt{n}} \;\le\; \mu_0 \;\le\; \bar{X} \;+\; t_{0.025,\, n-1}\frac{S}{\sqrt{n}}\,,$$

- In other words, we **fail to reject H_0** if

$$t_{0.025,\,n\,-\,1} \;\leq\; t \;\leq\; t_{0.025,\,n\,-\,1}\,.$$

Plugging

$$t \;=\; \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

in for t above, and solving for μ_0, we **fail to reject H_0** if

$$\bar{X} \;-\; t_{0.025,\,n\,-\,1}\frac{S}{\sqrt{n}} \;\leq\; \mu_0 \;\leq\; \bar{X} \;+\; t_{0.025,\,n\,-\,1}\frac{S}{\sqrt{n}}\,,$$

i.e. if μ_0 is in the **CI.**

Two-Sample t Test for Two Population Means μ_1 and μ_2 (9.1, 9.2)

- Suppose we have random samples of sizes m and n from **two populations.**

Two-Sample t Test for Two Population Means μ_1 and μ_2 (9.1, 9.2)

- Suppose we have random samples of sizes m and n from **two populations.**
- We'll see how to use the samples to decide if the **population means** μ_1 and μ_2 are different.

Two-Sample t Test for Two Population Means μ_1 and μ_2 (9.1, 9.2)

- Suppose we have random samples of sizes m and n from **two populations.**
- We'll see how to use the samples to decide if the **population means** μ_1 and μ_2 are different.

  The appropriate test is called the ***two-sample*** t ***test for*** $\mu_1 - \mu_2$.

- 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.

The Sampling Distribution of $\bar{X} - \bar{Y}$

- Suppose $X_1, X_2, \ldots, X_m$ are a random sample from a **population** whose **mean** is $\boldsymbol{\mu_1}$ and $Y_1, Y_2, \ldots, Y_n$ are random sample from a **population** whose **mean** is $\boldsymbol{\mu_2}$.

The Sampling Distribution of $\bar{X} - \bar{Y}$

- Suppose $X_1, X_2, \ldots, X_m$ are a random sample from a **population** whose **mean** is μ_1 and $Y_1, Y_2, \ldots, Y_n$ are random sample from a **population** whose **mean** is μ_2.
- The difference $\bar{X} - \bar{Y}$ between the two **sample means** is an **estimator** of $\mu_1 - \mu_2$.

Proposition

If $X_1, X_2, \ldots, X_m$ are a random sample from a $N(\mu_1, \sigma_1)$ distribution and $Y_1, Y_2, \ldots, Y_n$ are random sample from a $N(\mu_2, \sigma_2)$ distribution, and the two samples are drawn *independently* of each other. Then

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2,\ \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right). \qquad (1)$$

It follows that

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim N(0, 1). \qquad (2)$$

Furthermore, (1) and (2) still hold (at least approximately) even if the samples are from **non-normal** populations as long as the sample sizes m and n are both **large**.

- This follows because

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1}{\sqrt{m}}\right) \qquad \text{and} \qquad \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2}{\sqrt{n}}\right),$$

and so $\bar{X} - \bar{Y}$ is a linear combination of two independent normal random variables.

Proposition

Under the assumptions stated in the last proposition, the random variable

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S_1^2/m + S_2^2/n}} \sim t(\nu)$$

(at least approximately), where S_1 and S_2 are the **sample standard deviations** and the **df** ν are given by

$$\nu = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}}, \tag{3}$$

which should be rounded *down* to the nearest integer.

Two-Sample t Test Statistic for $\mu_1 - \mu_2$:

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

Two-Sample t Test Statistic for $\mu_1 - \mu_2$:

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

- t measures how many standard errors $\bar{X} - \bar{Y}$ is away from $\mathbf{0}$.

Two-Sample t Test Statistic for $\mu_1 - \mu_2$:

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

- t measures how many standard errors $\bar{X} - \bar{Y}$ is away from $\mathbf{0}$.
- $\bar{X} - \bar{Y}$ is an estimator of the unknown difference $\mu_1 - \mu_2$, so ...

Two-Sample t Test Statistic for $\mu_1 - \mu_2$:

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

- t measures how many standard errors $\bar{X} - \bar{Y}$ is away from $\mathbf{0}$.
- $\bar{X} - \bar{Y}$ is an estimator of the unknown difference $\boldsymbol{\mu_1 - \mu_2}$, so ...
    1. t will be approximately **zero** (most likely) if $\mu_1 - \mu_2 = 0$.
    2. It will be **positive** (most likely) if $\mu_1 - \mu_2 > 0$.
    3. It will be **negative** (most likely) if $\mu_1 - \mu_2 < 0$.

1. **Large positive** values of t provide **evidence against H_0 in favor of**
   $H_a : \mu_1 - \mu_2 > 0$.
2. **Large negative** values of t provide **evidence against H_0 in favor of**
   $H_a : \mu_1 - \mu_2 < 0$.
3. **Large positive *and* large negative** values of t provide **evidence against H_0 in favor of**
   $H_a : \mu_1 - \mu_2 \neq 0$.

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

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

is true,

$$t \sim t(\nu).$$

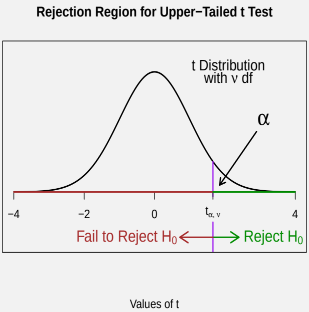
- The $t(\nu)$ curve gives us:

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

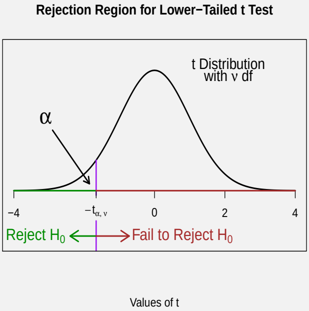
- The $t(\nu)$ curve gives us:
  - The ***rejection region*** as the **extreme 100α% of t values** (in the direction(s) specified by H_a).
  - The ***p-value*** as the **tail area(s) beyond the observed t value** (in the direction(s) specified by H_a).

Rejection Region: The **rejection region** is the **set of t values** in the tail of the $t(\nu)$ curve:

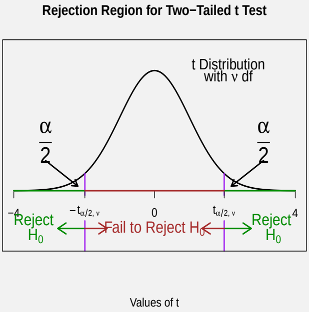
1. To the **right of** $t_{\alpha,\nu}$ if the alternative hypothesis is $H_a : \mu > \mu_0$:



2. To the **left of** $-t_{\alpha,\nu}$ if the alternative hypothesis is $H_a : \mu < \mu_0$:

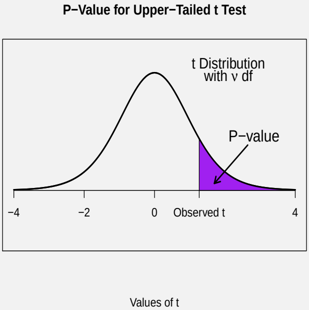


3. To the **left of** $-t_{\alpha/2,\nu}$ **and right of** $t_{\alpha/2,\nu}$ if the alternative hypothesis is $H_a : \mu \neq \mu_0$:

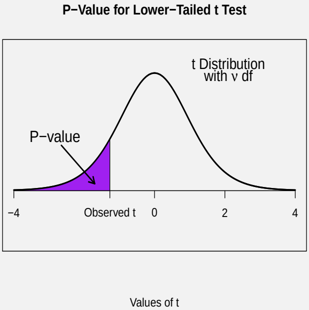


P-Value: The **p-value** is the **tail area** under the $t(\nu)$ curve:

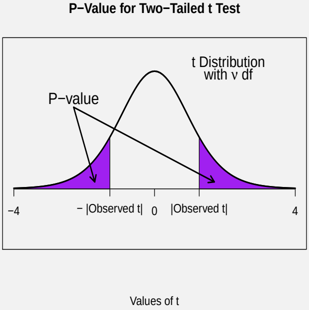
1. To the **right** of the **observed** t if the alternative hypothesis is $H_a : \mu > \mu_0$:



2. To the **left** of the **observed** t if the alternative hypothesis is $H_a : \mu < \mu_0$:



3. To the **left of** $-|t|$ **and right of** $|t|$ if the alternative hypothesis is $H_a: \mu \neq \mu_0$:



Exercise

Angry passengers, congested streets, time schedules, and noise and air pollution can lead to stress and premature retirement in urban bus drivers.

Exercise

Angry passengers, congested streets, time schedules, and noise and air pollution can lead to stress and premature retirement in urban bus drivers.

An intervention program designed by the Stockholm Transit District was implemented to improve the work conditions of the city's bus drivers.

Exercise

Angry passengers, congested streets, time schedules, and noise and air pollution can lead to stress and premature retirement in urban bus drivers.

An intervention program designed by the Stockholm Transit District was implemented to improve the work conditions of the city's bus drivers.

Drivers were assigned to **improved** routes (**intervention**) or **normal** routes (**control**), and various physiological and psychological data were recorded for each driver.

Shown below are the data on the heart rates, in beats per minute:

Intervention		**Control**						
68	66	74	52	67	63	77	57	80
74	58	77	53	76	54	73	54	60
69	63	77	63	60	68	64	66	71
68	73	66	55	71	84	63	73	59
64	76	68	64	82				

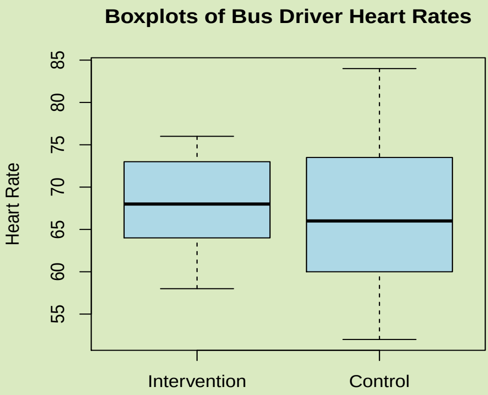
Here are the summary statistics:

Intervention			Control		
m	$=$	10	n	$=$	31
$\bar{x}$	$=$	67.90	$\bar{y}$	$=$	66.81
s_1	$=$	5.49	s_2	$=$	9.04

Here are the summary statistics:

Intervention	Control
$m = 10$	$n = 31$
$\bar{x} = 67.90$	$\bar{y} = 66.81$
$s_1 = 5.49$	$s_2 = 9.04$

Side-by-side boxplots are shown on the next slide.


Boxplots of Bus Driver Heart Rates
Heart Rate
55
60
65
70
75
80
85
Intervention
Control

a) Carry out a **two-sample t test** to decide if the intervention program **reduces the mean heart rate** of urban bus drivers in Stockholm. Use a **significance level** $\alpha = 0.05$.

a) Carry out a **two-sample t test** to decide if the intervention program **reduces the mean heart rate** of urban bus drivers in Stockholm. Use a **significance level $\alpha = 0.05$.**

Hint: The **df** are

$$\nu = \frac{\left(\frac{5.49^2}{10} + \frac{9.04^2}{31}\right)^2}{\frac{(5.49^2/10)^2}{10-1} + \frac{(9.04^2/31)^2}{31-1}} = 25.7,$$

which we round *down* to **25**, the **test statistic** ends up being $t = 0.46$ and the **p-value 0.675** (from R).

a) Carry out a **two-sample t test** to decide if the intervention program **reduces the mean heart rate** of urban bus drivers in Stockholm. Use a **significance level $\alpha = 0.05$.**

Hint: The **df** are

$$\nu = \frac{\left(\frac{5.49^2}{10} + \frac{9.04^2}{31}\right)^2}{\frac{(5.49^2/10)^2}{10-1} + \frac{(9.04^2/31)^2}{31-1}} = 25.7,$$

which we round *down* to **25**, the **test statistic** ends up being $t = 0.46$ and the **p-value 0.675** (from R).

b) Can you provide an explanation for the surprising results of the study?

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

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

for some (non-zero) value Δ_0. In this case, H_a also has Δ_0 in place of 0, and the test statistic is

$$T = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{S_1^2/m + S_2^2/n}}.$$

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

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

for some (non-zero) value Δ_0. In this case, H_a also has Δ_0 in place of 0, and the test statistic is

$$T = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{S_1^2/m + S_2^2/n}}.$$

P-values and rejection regions are exactly as described for the usual two-sample t test of

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

Two-Sample t Confidence Interval for $\mu_1 - \mu_2$

- The difference $\mu_1 - \mu_2$ is sometimes called the ***effect size***, and its estimate $\bar{X} - \bar{Y}$ the ***estimated effect size***.

Two-Sample t CI: Suppose $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ are independent random samples from populations whose means are μ_1 and μ_2. Then a $100(1-\alpha)\%$ ***two-sample t confidence interval for $\mu_1 - \mu_2$*** is

$$\bar{X} - \bar{Y} \pm t_{\alpha/2,\,\nu} \cdot \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}, \tag{4}$$

where the **df** ν is given by (3).

- The CI is valid if either the samples are from **normal** populations or m and n **are large.**

- The CI is valid if either the samples are from **normal** populations or m and n **are large**.
- In either case, we can be $100(1-\alpha)\%$ confident that $\mu_1 - \mu_2$ will be contained in the CI.

Exercise

Consider again the study on bus driver heart rates.

Exercise

Consider again the study on bus driver heart rates.

a) Give a (point) **estimate** for the **effect size** $\mu_1 - \mu_2$ of the intervention on mean heart rates.

Exercise

Consider again the study on bus driver heart rates.

a) Give a (point) **estimate** for the **effect size** $\mu_1 - \mu_2$ of the intervention on mean heart rates.

b) Compute and interpret a **95% CI** for $\mu_1 - \mu_2$.

Exercise

Consider again the study on bus driver heart rates.

a) Give a (point) **estimate** for the **effect size** $\mu_1 - \mu_2$ of the intervention on mean heart rates.

b) Compute and interpret a **95% CI** for $\mu_1 - \mu_2$.

Hints: The **df** are **25** (again) and the t **critical value** is $t_{0.025,\,25} = 2.060$.

Exercise

Consider again the study on bus driver heart rates.

a) Give a (point) **estimate** for the **effect size** $\mu_1 - \mu_2$ of the intervention on mean heart rates.

b) Compute and interpret a **95% CI** for $\mu_1 - \mu_2$.

Hints: The **df** are **25** (again) and the t **critical value** is $t_{0.025,\,25} = 2.060$.

c) Is **0** contained in the CI? What does that indicate about effect of the intervention on heart rates?