Introduction to Statistics

Nels Grevstad

Metropolitan State University of Denver

ngrevsta@msudenver.edu

October 27, 2019

Topics

Objectives

Objectives:

- Carry out a one-mean t test for a population mean μ when the population standard deviation σ is unknown.

Hypothesis Test for μ when σ is Unknown (9.5)

Introduction to the One-Mean t Test

- Usually the population standard deviation σ ***isn't*** **known**, so we **can't** use the *one-mean z test* for μ.

Hypothesis Test for μ when σ is Unknown (9.5)

Introduction to the One-Mean t Test

- Usually the population standard deviation σ ***isn't* known**, so we **can't** use the *one-mean z test* for μ.

  Instead, we **estimate** σ by the sample standard deviation s, and then use the ***one-mean t test*** described ahead.

Hypothesis Test for μ when σ is Unknown (9.5)

Introduction to the One-Mean t Test

- Usually the population standard deviation σ ***isn't*** **known**, so we **can't** use the *one-mean z test* for μ.

  Instead, we **estimate** σ by the sample standard deviation s, and then use the ***one-mean t test*** described ahead.

  The test is used when we have a **random sample** from a **population**.

  We'll use the sample to decide if the population mean μ is different from some **claimed value** μ_0.

- The **null hypothesis** is:

Null Hypothesis:

$$H_0 : \mu = \mu_0$$

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

Alternative Hypothesis: The alternative hypothesis will be one of

1. $H_a : \mu > \mu_0$ (**one-sided, upper-tailed test**)
2. $H_a : \mu < \mu_0$ (**one-sided, lower-tailed test**)
3. $H_a : \mu \neq \mu_0$ (**two-sided, two-tailed test**)

depending on what we're trying to substantiate in our study.

- The **test statistic** for the ***one-mean t test for μ*** is

One-Mean T Test Statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

When

$$H_0 : \mu = \mu_0$$

is true, the **sampling distribution** of the test statistic t is a t distribution with $n - 1$ df.

- The **test statistic** for the ***one-mean*** t ***test for*** μ is

One-Mean T Test Statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

When

$$H_0 : \mu = \mu_0$$

is true, the **sampling distribution** of the test statistic t is a t distribution with $n-1$ df.

(The *one-mean* t *test* is valid if either the sample is from a **normal** population or the sample size n is **large** ($n \geq 30$).)

- t measures (approximately) how many **standard errors** $\bar{x}$ is **away from** the claimed value μ_0.

1. Values of t **close to zero** provide almost ***no* evidence against** the **null hypothesis** $H_0 : \mu = \mu_0$.

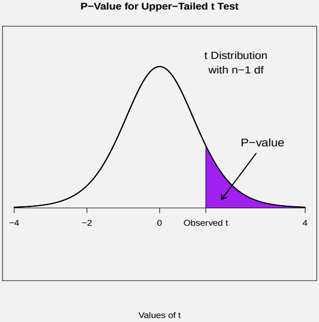
2. **Positive** values of t provide **evidence against the null hypothesis in favor of** $H_a : \mu > \mu_0$.

3. **Negative** values of t provide **evidence against the null hypothesis in favor of** $H_a : \mu < \mu_0$.

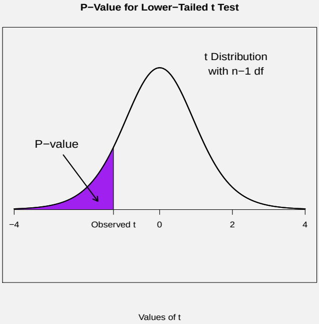
4. **Positive *and* negative** values of t provide **evidence against the null hypothesis in favor of** $H_a : \mu \neq \mu_0$.

- The ***p-value*** is the probability that just by chance (under H_0) we'd get a test statistic value as far from zero, in the direction predicted by H_a, as the observed value.

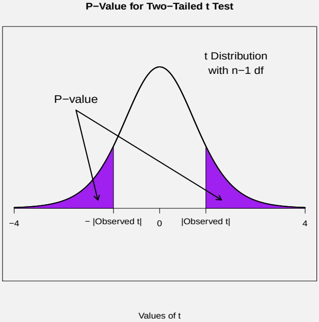
1. **P-value** = Area to the **right** of the observed t if the alternative hypothesis is $H_a : \mu > \mu_0$.



2. **P-value** = Area to the **left** of the observed t if the alternative hypothesis is $H_a : \mu < \mu_0$.



3. **P-value** = Area to the **left** of $-|t|$ **and right** of $|t|$ if the alternative hypothesis is $H_a : \mu \neq \mu_0$.



- As always, after choosing a **level of significance** α, we **reject** $\boldsymbol{H_0}$ if **p-value** $< \alpha$, otherwise fail reject H_0.

- As always, after choosing a **level of significance** α, we **reject** H_0 if **p-value** $< \alpha$, otherwise fail reject H_0.
- The next example illustrates a **two-tailed** t **test**.

- As always, after choosing a **level of significance** α, we **reject** H_0 if **p-value** $< \alpha$, otherwise fail reject H_0.
- The next example illustrates a **two-tailed** t **test.**

  The two the come after it illustrate **one-tailed tests.**

Example

A quality control engineer monitors a machine that puts cereal into boxes.

Example

A quality control engineer monitors a machine that puts cereal into boxes.

According to the label, each box is supposed to contain **16** oz of cereal.

Example

A quality control engineer monitors a machine that puts cereal into boxes.

According to the label, each box is supposed to contain **16** oz of cereal.

The machine will need to be adjusted if the boxes are systematically being **under-filled** or **over-filled**.

To decide if the boxes are being **under-filled or overfilled**, the engineer will test the **hypotheses**

$$H_0: \mu = 16$$
$$H_a: \mu \neq 16$$

where μ is the true (unknown) population mean weight.

To decide if the boxes are being **under-filled or overfilled**, the engineer will test the **hypotheses**

$$\begin{aligned} H_0: \ \mu &= 16 \\ H_a: \ \mu &\neq 16 \end{aligned}$$

where μ is the true (unknown) population mean weight.

A random sample of **ten** boxes gives

$$\bar{x} = \mathbf{16.6} \qquad \text{and} \qquad s = \mathbf{0.9}.$$

Thus the observed **test statistic** is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Thus the observed **test statistic** is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{16.6 - 16}{0.9/\sqrt{10}}$$

Thus the observed **test statistic** is

$$\begin{aligned} t &= \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \\ &= \frac{16.6-16}{0.9/\sqrt{10}} \\ &= \mathbf{2.11}. \end{aligned}$$

Thus the **sample mean** weight, $\bar{x} = \mathbf{16.6}$, is about **2.11 standard errors above 16** ounces.

Thus the observed **test statistic** is

$$\begin{aligned} t &= \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \\ &= \frac{16.6-16}{0.9/\sqrt{10}} \\ &= \mathbf{2.11}. \end{aligned}$$

Thus the **sample mean** weight, $\bar{x} = \mathbf{16.6}$, is about **2.11 standard errors above 16** ounces.

The **p-value** is the **probability** that we'd get a t value this far away from zero (in either direction) by chance **if** the **population mean** μ was **16**.

From the **two tail** areas of the t **distribution** with $n - 1 = 9$ **df**, to the **right** of **2.11** and **left** of **-2.11**,

$$\textbf{p-value} = 2 \times 0.033 = \mathbf{0.066}.$$

(The value 0.033 was obtained from the **table** showing **areas to the right of** t.)

From the **two tail** areas of the t **distribution** with $n - 1 = 9$ **df**, to the **right** of **2.11** and **left** of **-2.11**,

$$\textbf{p-value} \;=\; 2 \times 0.033 \;=\; \mathbf{0.066}.$$

(The value 0.033 was obtained from the **table** showing **areas to the right of** t.)

Thus we'd get a result like the one we got **6.6%** of the time **even if** the **population mean** μ was **16** ounces.

Using a **level of significance** $\alpha = \mathbf{0.05}$, the **decision rule** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

Using a **level of significance** $\alpha = 0.05$, the **decision rule** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

Because $\mathbf{0.066 \geq 0.05}$, we **fail to reject** $\boldsymbol{H_0}$.

Using a **level of significance** $\alpha = 0.05$, the **decision rule** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

Because $\mathbf{0.066 \geq 0.05}$, we **fail to reject H_0.**

There's **no statistically significant evidence** that the population mean cereal box weight μ is different from 16 ounces.

Using a **level of significance** $\alpha = 0.05$, the **decision rule** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

Because $0.066 \geq 0.05$, we **fail to reject** H_0.

There's **no statistically significant evidence** that the population mean cereal box weight μ is different from 16 ounces.

The result that the engineer got (by taking a random sample) can be explained by chance variation (sampling error).

Exercise

A study was designed to find out if **dietary calcium reduces blood pressure.**

Exercise

A study was designed to find out if **dietary calcium reduces blood pressure**.

Ten men were given a calcium supplement for 12 weeks. Blood pressure was measured both ***before*** and ***after*** the 12-week period.

Exercise

A study was designed to find out if **dietary calcium reduces blood pressure.**

Ten men were given a calcium supplement for 12 weeks. Blood pressure was measured both ***before*** and ***after*** the 12-week period.

The data are the ***changes*** in blood pressure for the 10 men:

-7   4   -18   -17   3   5   -1   -10   -11   2

A **negative** value means the **blood pressure decreased.**

The sample mean and standard deviation of the blood pressure ***changes*** are

$$\bar{x} = -5.00 \qquad \text{and} \qquad s = 8.74$$

The sample mean and standard deviation of the blood pressure ***changes*** are

$$\bar{x} = -5.00 \qquad \text{and} \qquad s = 8.74$$

We'll carry out a **one-mean t test** to decide if calcium **lowers blood pressure.**

The sample mean and standard deviation of the blood pressure ***changes*** are

$$\bar{x} = -5.00 \qquad \text{and} \qquad s = 8.74$$

We'll carry out a **one-mean t test** to decide if calcium **lowers blood pressure.**

a) State the **hypotheses** that should be tested.

The sample mean and standard deviation of the blood pressure ***changes*** are

$$\bar{x} = -5.00 \qquad \text{and} \qquad s = 8.74$$

We'll carry out a **one-mean t test** to decide if calcium **lowers blood pressure**.

a) State the **hypotheses** that should be tested.

b) Calculate the **test statistic**.

The sample mean and standard deviation of the blood pressure ***changes*** are

$$\bar{x} = -5.00 \qquad \text{and} \qquad s = 8.74$$

We'll carry out a **one-mean t test** to decide if calcium **lowers blood pressure**.

a) State the **hypotheses** that should be tested.

b) Calculate the **test statistic**.

Hint: You should get **-1.81**.

The sample mean and standard deviation of the blood pressure ***changes*** are

$$\bar{x} = -5.00 \qquad \text{and} \qquad s = 8.74$$

We'll carry out a **one-mean t test** to decide if calcium **lowers blood pressure**.

a) State the **hypotheses** that should be tested.

b) Calculate the **test statistic.**

Hint: You should get **-1.81**.

c) Determine the **p-value.**

The sample mean and standard deviation of the blood pressure ***changes*** are

$$\bar{x} = -5.00 \qquad \text{and} \qquad s = 8.74$$

We'll carry out a **one-mean t test** to decide if calcium **lowers blood pressure**.

a) State the **hypotheses** that should be tested.

b) Calculate the **test statistic**.

Hint: You should get **-1.81**.

c) Determine the **p-value**.

Hint: You should get **0.053**.

d) State the **conclusion** using a **level of significance** $\boldsymbol{\alpha = 0.05}$, in which case the **decision criterion** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

d) State the **conclusion** using a **level of significance** $\boldsymbol{\alpha = 0.05}$, in which case the **decision criterion** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

e) **Interpret** the result: Is there **statistically significant** evidence that calcium **lowers blood pressure**?

d) State the **conclusion** using a **level of significance** $\boldsymbol{\alpha = 0.05}$, in which case the **decision criterion** is

Reject H_0 if p-value < 0.05.
Fail to reject H_0 if p-value ≥ 0.05.

e) **Interpret** the result: Is there **statistically significant** evidence that calcium **lowers blood pressure**?

f) If instead we had used a **significance level** $\boldsymbol{\alpha = 0.10}$, would the conclusion have been different? Explain

Exercise

In the U.S., the mean number of newly decayed teeth an individual gets per year is **0.30**.

Exercise

In the U.S., the mean number of newly decayed teeth an individual gets per year is **0.30**.

To investigate the **claim** that **sugar increases the decay rate of teeth**, **25** adults were examined and then given a sugar solution to supplement all their meals.

Exercise

In the U.S., the mean number of newly decayed teeth an individual gets per year is **0.30**.

To investigate the **claim** that **sugar increases the decay rate of teeth**, **25** adults were examined and then given a sugar solution to supplement all their meals.

After one year, the mean and standard deviation of the number of newly decayed teeth for the group were

$$\bar{x} = \mathbf{0.60} \qquad \text{and} \qquad s = \mathbf{0.45}.$$

We'll carry out a **one-mean t test** to decide if there's statistically significant evidence that the true (unknown) **population mean** tooth decay rate μ for people given a sugar solution is **greater than 0.30**.

We'll carry out a **one-mean t test** to decide if there's statistically significant evidence that the true (unknown) **population mean** tooth decay rate μ for people given a sugar solution is **greater than 0.30**.

a) State the **hypotheses** that should be tested.

We'll carry out a **one-mean t test** to decide if there's statistically significant evidence that the true (unknown) **population mean** tooth decay rate μ for people given a sugar solution is **greater than 0.30**.

a) State the **hypotheses** that should be tested.

b) Calculate the **test statistic**.

We'll carry out a **one-mean t test** to decide if there's statistically significant evidence that the true (unknown) **population mean** tooth decay rate μ for people given a sugar solution is **greater than 0.30**.

a) State the **hypotheses** that should be tested.

b) Calculate the **test statistic**.

Hint: You should get **3.33**.

We'll carry out a **one-mean t test** to decide if there's statistically significant evidence that the true (unknown) **population mean** tooth decay rate μ for people given a sugar solution is **greater than 0.30**.

a) State the **hypotheses** that should be tested.

b) Calculate the **test statistic**.

Hint: You should get **3.33**.

c) Determine the **p-value**.

We'll carry out a **one-mean t test** to decide if there's statistically significant evidence that the true (unknown) **population mean** tooth decay rate μ for people given a sugar solution is **greater than 0.30**.

a) State the **hypotheses** that should be tested.

b) Calculate the **test statistic**.

Hint: You should get **3.33**.

c) Determine the **p-value**.

Hint: You should get **0.002**.

Data Snooping: Don't Do It!

- The direction (left-tailed, right-tailed, or two-tailed) for H_a should be chosen **before** you examine the data.

Data Snooping: Don't Do It!

- The direction (left-tailed, right-tailed, or two-tailed) for H_a should be chosen **before** you examine the data.

  It should indicate what you **expect** to find when you carry out your study.

Data Snooping: Don't Do It!

- The direction (left-tailed, right-tailed, or two-tailed) for H_a should be chosen **before** you examine the data.

  It should indicate what you **expect** to find when you carry out your study.

  If you don't have a particular direction in mind for H_a *before* examining the data, *use the two-tailed test*.

- ***Data snooping*** refers to first examining the data *and then* choosing the direction for H_a that best matches what you *already* see in the data.

- ***Data snooping*** refers to first examining the data *and then* choosing the direction for H_a that best matches what you *already* see in the data.

  Data snooping is considered "**cheating**" because it leads to an **artificially small p-value** – one that's *half as large* as what it would've been if you'd done a two-tailed test.

- ***Data snooping*** refers to first examining the data *and then* choosing the direction for H_a that best matches what you *already* see in the data.

  Data snooping is considered "**cheating**" because it leads to an **artificially small p-value** – one that's *half as large* as what it would've been if you'd done a two-tailed test.

- The next example shows how **data snooping** can lead an **artificially small p-value** and in turn to **mistakenly rejecting H_0** when you shouldn't have.

Example

Suppose in the cereal-box weight study from the earlier example that the engineer "**cheats**" (**data snoops**) and decides, ***after*** noticing that $\bar{x} = \mathbf{16.6}$ is ***greater*** than **16**, to do an **upper-tailed test** of

$$\begin{aligned} H_0 : \mu &= 16 \\ H_a : \mu &> 16 \end{aligned}$$

Example

Suppose in the cereal-box weight study from the earlier example that the engineer **"cheats"** (**data snoops**) and decides, ***after*** noticing that $\bar{x} = \mathbf{16.6}$ is ***greater*** than **16**, to do an **upper-tailed test** of

$$\begin{aligned} H_0 : \; \mu &= 16 \\ H_a : \; \mu &> 16 \end{aligned}$$

What would the **p-value** be?

Example

Suppose in the cereal-box weight study from the earlier example that the engineer "**cheats**" (**data snoops**) and decides, ***after*** noticing that $\bar{x} = \mathbf{16.6}$ is ***greater*** than **16**, to do an **upper-tailed test** of

$$\begin{aligned} H_0 : \mu &= 16 \\ H_a : \mu &> 16 \end{aligned}$$

What would the **p-value** be?

Answer: **0.033** (just the **upper** tail area).

Would the **conclusion** (using $\alpha = 0.05$ still) **be different** from the one reached in the earlier example (when a two-tailed test was performed)?

Would the **conclusion** (using $\alpha = 0.05$ still) **be different** from the one reached in the earlier example (when a two-tailed test was performed)?

Answer: **Yes.** In this case, the **null hypothesis** would (mistakenly) be **rejected.**