STAT 111

Recitation 10

Mo Huang

Email: mohuang@wharton.upenn.edu

Office Hours: Wednesdays 3:00 - 4:00 pm, JMHH F96

Slides: `github.com/mohuangx/STAT111-Spring2019`

April 12, 2019

Two-sample t-test

- Example: Suppose we have two classes, let's call 201 and 202, and we want to see if there is a difference in height between these two classes. How do we test this?

Two-sample t-test

- Example: Suppose we have two classes, let's call 201 and 202, and we want to see if there is a difference in height between these two classes. How do we test this?
- Answer: Let $X_{11}, \ldots, X_{1m}$ represent the heights of the m students in 201 with mean μ_1 and let $X_{21}, \ldots, X_{2n}$ represent the heights of the n students in 202 with mean μ_2. We can test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$.

Two-sample t-test

- Example: Suppose we have two classes, let's call 201 and 202, and we want to see if there is a difference in height between these two classes. How do we test this?
- Answer: Let $X_{11}, \ldots, X_{1m}$ represent the heights of the m students in 201 with mean μ_1 and let $X_{21}, \ldots, X_{2n}$ represent the heights of the n students in 202 with mean μ_2. We can test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$.
- This is an example of a two-sample t-test!
- Two-sample t-test: Let $X_{11}, \ldots, X_{1m}$ be i.i.d random variables with (unknown) mean μ_1 and (unknown) variance σ^2. Let $X_{21}, \ldots, X_{2n}$ be i.i.d random variables with (unknown) mean μ_2 and (unknown) variance σ^2. We want to test whether or not $\mu_1 = \mu_2$.

Example

- Suppose we want to test whether there is a difference in height between class 201 ($X_{11}, \dots, X_{1m}$) and 202 ($X_{21}, \dots, X_{2n}$). We observe that $\bar{x}_1 = 66.7$, $s_1^2 = 10.5$, $m = 28$, and $\bar{x}_2 = 65.6$, $s_2^2 = 12.3$, $n = 34$.

Step 1 $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Two-sided test.

Example

- Suppose we want to test whether there is a difference in height between class 201 $(X_{11}, \ldots, X_{1m})$ and 202 $(X_{21}, \ldots, X_{2n})$. We observe that $\bar{x}_1 = 66.7$, $s_1^2 = 10.5$, $m = 28$, and $\bar{x}_2 = 65.6$, $s_2^2 = 12.3$, $n = 34$.

Step 1 $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Two-sided test.

Step 2 Choose $\alpha = 0.05$.

Example

- Suppose we want to test whether there is a difference in height between class 201 $(X_{11}, \ldots, X_{1m})$ and 202 $(X_{21}, \ldots, X_{2n})$. We observe that $\bar{x}_1 = 66.7$, $s_1^2 = 10.5$, $m = 28$, and $\bar{x}_2 = 65.6$, $s_2^2 = 12.3$, $n = 34$.

Step 1 $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Two-sided test.

Step 2 Choose $\alpha = 0.05$.

Step 3 Test-statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{m} + \frac{1}{n}}}, \text{ where } s = \sqrt{\frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}}$$

Example

- Suppose we want to test whether there is a difference in height between class 201 $(X_{11}, \ldots, X_{1m})$ and 202 $(X_{21}, \ldots, X_{2n})$. We observe that $\bar{x}_1 = 66.7$, $s_1^2 = 10.5$, $m = 28$, and $\bar{x}_2 = 65.6$, $s_2^2 = 12.3$, $n = 34$.

Step 1 $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Two-sided test.

Step 2 Choose $\alpha = 0.05$.

Step 3 Test-statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{m} + \frac{1}{n}}}, \text{ where } s = \sqrt{\frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}}$$

$$t = \frac{66.7 - 65.6}{3.390\sqrt{\frac{1}{28} + \frac{1}{34}}} = 1.254$$

Example

- Suppose we want to test whether there is a difference in height between class 201 ($X_{11}, \ldots, X_{1m}$) and 202 ($X_{21}, \ldots, X_{2n}$). We observe that $\bar{x}_1 = 66.7$, $s_1^2 = 10.5$, $m = 28$, and $\bar{x}_2 = 65.6$, $s_2^2 = 12.3$, $n = 34$.

Step 1 $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Two-sided test.

Step 2 Choose $\alpha = 0.05$.

Step 3 Test-statistic is $t = 1.254$.

Step 4 Find the critical region.

Example

- Suppose we want to test whether there is a difference in height between class 201 $(X_{11}, \ldots, X_{1m})$ and 202 $(X_{21}, \ldots, X_{2n})$. We observe that $\bar{x}_1 = 66.7$, $s_1^2 = 10.5$, $m = 28$, and $\bar{x}_2 = 65.6$, $s_2^2 = 12.3$, $n = 34$.

Step 1 $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Two-sided test.

Step 2 Choose $\alpha = 0.05$.

Step 3 Test-statistic is $t = 1.254$.

Step 4 Find the critical region.

How many degrees of freedom do we have?

Example

- Suppose we want to test whether there is a difference in height between class 201 ($X_{11}, \ldots, X_{1m}$) and 202 ($X_{21}, \ldots, X_{2n}$). We observe that $\bar{x}_1 = 66.7$, $s_1^2 = 10.5$, $m = 28$, and $\bar{x}_2 = 65.6$, $s_2^2 = 12.3$, $n = 34$.

Step 1 $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Two-sided test.

Step 2 Choose $\alpha = 0.05$.

Step 3 Test-statistic is $t = 1.254$.

Step 4 Find the critical region.

How many degrees of freedom do we have? $m + n - 2 = 60$

Example

- Suppose we want to test whether there is a difference in height between class 201 ($X_{11}, \ldots, X_{1m}$) and 202 ($X_{21}, \ldots, X_{2n}$). We observe that $\bar{x}_1 = 66.7$, $s_1^2 = 10.5$, $m = 28$, and $\bar{x}_2 = 65.6$, $s_2^2 = 12.3$, $n = 34$.

Step 1 $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Two-sided test.

Step 2 Choose $\alpha = 0.05$.

Step 3 Test-statistic is $t = 1.254$.

Step 4 Find the critical region.

How many degrees of freedom do we have? $m + n - 2 = 60$

So we need to look at t_{60}. What is the critical region?

Example

- Suppose we want to test whether there is a difference in height between class 201 ($X_{11}, \ldots, X_{1m}$) and 202 ($X_{21}, \ldots, X_{2n}$). We observe that $\bar{x}_1 = 66.7$, $s_1^2 = 10.5$, $m = 28$, and $\bar{x}_2 = 65.6$, $s_2^2 = 12.3$, $n = 34$.

Step 1 $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Two-sided test.

Step 2 Choose $\alpha = 0.05$.

Step 3 Test-statistic is $t = 1.254$.

Step 4 Find the critical region.

How many degrees of freedom do we have? $m + n - 2 = 60$

So we need to look at t_{60}. What is the critical region?

$t \geq t_{60,0.025} = 2.000$ and $t \leq -t_{60,0.025} = -2.000$.

Example

- Suppose we want to test whether there is a difference in height between class 201 ($X_{11}, \dots, X_{1m}$) and 202 ($X_{21}, \dots, X_{2n}$). We observe that $\bar{x}_1 = 66.7$, $s_1^2 = 10.5$, $m = 28$, and $\bar{x}_2 = 65.6$, $s_2^2 = 12.3$, $n = 34$.

Step 1 $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Two-sided test.

Step 2 Choose $\alpha = 0.05$.

Step 3 Test-statistic is $t = 1.254$.

Step 4 Find the critical region: $t \geq 2.000$ and $t \leq -2.000$

Step 5 Do we reject H_0?

Example

- Suppose we want to test whether there is a difference in height between class 201 $(X_{11}, \ldots, X_{1m})$ and 202 $(X_{21}, \ldots, X_{2n})$. We observe that $\bar{x}_1 = 66.7$, $s_1^2 = 10.5$, $m = 28$, and $\bar{x}_2 = 65.6$, $s_2^2 = 12.3$, $n = 34$.

Step 1 $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Two-sided test.

Step 2 Choose $\alpha = 0.05$.

Step 3 Test-statistic is $t = 1.254$.

Step 4 Find the critical region: $t \geq 2.000$ and $t \leq -2.000$

Step 5 Do we reject H_0? No, $t = 1.254$ is not in the critical region.

Paired two sample t test

- Suppose we have two samples where there is a natural pairing of data between the two samples. Let μ_d be the mean difference between the two samples.

- For example, we have n patients and we are interested in determining if a drug decreases cholesterol levels. We collect cholesterol levels before $(x_{11}, \ldots, x_{1n})$ and after $(x_{21}, \ldots, x_{2n})$ administering the drug.

- We want to test $H_0 : \mu_d = 0$.

- Consider $d_i = x_{2i} - x_{1i}$, the difference in measurement between sample 2 and sample 1 for subject i.
  - Estimate of μ_d: $\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$
  - Estimate of σ^2: $s_d^2 = \frac{d_1^2 + d_2^2 + \cdots + d_n^2 - n(\bar{d})^2}{n-1}$

- The test statistic is

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

Example

- Suppose we have 10 patients and we are interested in determining if a drug decreases cholesterol levels. We collect the following cholesterol levels before and after administering the drug:

Patient	1	2	3	4	5	6	7	8	9	10
Before	204	243	253	212	239	241	256	267	231	251
After	200	235	256	200	232	210	249	270	233	243
Difference	-4	-8	3	-12	-7	-31	-7	3	2	-8

Step 1 $H_0 : \mu_d = 0$ vs. $H_1 : \mu_d < 0$. One-sided test.

Example

- Suppose we have 10 patients and we are interested in determining if a drug decreases cholesterol levels. We collect the following cholesterol levels before and after administering the drug:

Patient	1	2	3	4	5	6	7	8	9	10
Before	204	243	253	212	239	241	256	267	231	251
After	200	235	256	200	232	210	249	270	233	243
Difference	-4	-8	3	-12	-7	-31	-7	3	2	-8

Step 1 $H_0 : \mu_d = 0$ vs. $H_1 : \mu_d < 0$. One-sided test.

Step 2 Choose $\alpha = 0.01$.

Example

- Suppose we have 10 patients and we are interested in determining if a drug decreases cholesterol levels. We collect the following cholesterol levels before and after administering the drug:

Patient	1	2	3	4	5	6	7	8	9	10
Before	204	243	253	212	239	241	256	267	231	251
After	200	235	256	200	232	210	249	270	233	243
Difference	-4	-8	3	-12	-7	-31	-7	3	2	-8

Step 1 $H_0 : \mu_d = 0$ vs. $H_1 : \mu_d < 0$. One-sided test.

Step 2 Choose $\alpha = 0.01$.

Step 3 Test-statistic is

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

Example

- Suppose we have 10 patients and we are interested in determining if a drug decreases cholesterol levels. We collect the following cholesterol levels before and after administering the drug:

Patient	1	2	3	4	5	6	7	8	9	10
Before	204	243	253	212	239	241	256	267	231	251
After	200	235	256	200	232	210	249	270	233	243
Difference	-4	-8	3	-12	-7	-31	-7	3	2	-8

Step 1 $H_0 : \mu_d = 0$ vs. $H_1 : \mu_d < 0$. One-sided test.

Step 2 Choose $\alpha = 0.01$.

Step 3 Test-statistic is

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

$$\bar{d} = -6.9 \qquad s_d = 9.96$$

Example

- Suppose we have 10 patients and we are interested in determining if a drug decreases cholesterol levels. We collect the following cholesterol levels before and after administering the drug:

Patient	1	2	3	4	5	6	7	8	9	10
Before	204	243	253	212	239	241	256	267	231	251
After	200	235	256	200	232	210	249	270	233	243
Difference	-4	-8	3	-12	-7	-31	-7	3	2	-8

Step 1 $H_0 : \mu_d = 0$ vs. $H_1 : \mu_d < 0$. One-sided test.

Step 2 Choose $\alpha = 0.01$.

Step 3 Test-statistic is

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

$$\bar{d} = -6.9 \qquad s_d = 9.96$$

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{-6.9}{9.96/\sqrt{10}} = -2.191$$

Example

- Suppose we have 10 patients and we are interested in determining if a drug decreases cholesterol levels. We collect the following cholesterol levels before and after administering the drug:

Patient	1	2	3	4	5	6	7	8	9	10
Before	204	243	253	212	239	241	256	267	231	251
After	200	235	256	200	232	210	249	270	233	243
Difference	-4	-8	3	-12	-7	-31	-7	3	2	-8

Step 1 $H_0 : \mu_d = 0$ vs. $H_1 : \mu_d < 0$. One-sided test.

Step 2 Choose $\alpha = 0.01$.

Step 3 Test-statistic is $t = -2.191$.

Step 4 Find the critical region.

Example

- Suppose we have 10 patients and we are interested in determining if a drug decreases cholesterol levels. We collect the following cholesterol levels before and after administering the drug:

Patient	1	2	3	4	5	6	7	8	9	10
Before	204	243	253	212	239	241	256	267	231	251
After	200	235	256	200	232	210	249	270	233	243
Difference	-4	-8	3	-12	-7	-31	-7	3	2	-8

Step 1 $H_0 : \mu_d = 0$ vs. $H_1 : \mu_d < 0$. One-sided test.

Step 2 Choose $\alpha = 0.01$.

Step 3 Test-statistic is $t = -2.191$.

Step 4 Find the critical region.

How many degrees of freedom do we have?

Example

- Suppose we have 10 patients and we are interested in determining if a drug decreases cholesterol levels. We collect the following cholesterol levels before and after administering the drug:

Patient	1	2	3	4	5	6	7	8	9	10
Before	204	243	253	212	239	241	256	267	231	251
After	200	235	256	200	232	210	249	270	233	243
Difference	-4	-8	3	-12	-7	-31	-7	3	2	-8

Step 1 $H_0 : \mu_d = 0$ vs. $H_1 : \mu_d < 0$. One-sided test.

Step 2 Choose $\alpha = 0.01$.

Step 3 Test-statistic is $t = -2.191$.

Step 4 Find the critical region.

How many degrees of freedom do we have? $n - 1 = 9$

Example

- Suppose we have 10 patients and we are interested in determining if a drug decreases cholesterol levels. We collect the following cholesterol levels before and after administering the drug:

Patient	1	2	3	4	5	6	7	8	9	10
Before	204	243	253	212	239	241	256	267	231	251
After	200	235	256	200	232	210	249	270	233	243
Difference	-4	-8	3	-12	-7	-31	-7	3	2	-8

Step 1 $H_0 : \mu_d = 0$ vs. $H_1 : \mu_d < 0$. One-sided test.

Step 2 Choose $\alpha = 0.01$.

Step 3 Test-statistic is $t = -2.191$.

Step 4 Find the critical region.

How many degrees of freedom do we have? $n - 1 = 9$

So we need to look at t_9. What is the critical region?

Example

- Suppose we have 10 patients and we are interested in determining if a drug decreases cholesterol levels. We collect the following cholesterol levels before and after administering the drug:

Patient	1	2	3	4	5	6	7	8	9	10
Before	204	243	253	212	239	241	256	267	231	251
After	200	235	256	200	232	210	249	270	233	243
Difference	-4	-8	3	-12	-7	-31	-7	3	2	-8

Step 1 $H_0 : \mu_d = 0$ vs. $H_1 : \mu_d < 0$. One-sided test.

Step 2 Choose $\alpha = 0.01$.

Step 3 Test-statistic is $t = -2.191$.

Step 4 Find the critical region.

How many degrees of freedom do we have? $n - 1 = 9$

So we need to look at t_9. What is the critical region?

$t \leq -t_{9,0.01} = -2.821$.

Example

- Suppose we have 10 patients and we are interested in determining if a drug decreases cholesterol levels. We collect the following cholesterol levels before and after administering the drug:

Patient	1	2	3	4	5	6	7	8	9	10
Before	204	243	253	212	239	241	256	267	231	251
After	200	235	256	200	232	210	249	270	233	243
Difference	-4	-8	3	-12	-7	-31	-7	3	2	-8

Step 1 $H_0 : \mu_d = 0$ vs. $H_1 : \mu_d < 0$. One-sided test.

Step 2 Choose $\alpha = 0.01$.

Step 3 Test-statistic is $t = -2.191$.

Step 4 Find the critical region: $t \leq -2.821$

Step 5 Do we reject H_0?

Example

- Suppose we have 10 patients and we are interested in determining if a drug decreases cholesterol levels. We collect the following cholesterol levels before and after administering the drug:

Patient	1	2	3	4	5	6	7	8	9	10
Before	204	243	253	212	239	241	256	267	231	251
After	200	235	256	200	232	210	249	270	233	243
Difference	-4	-8	3	-12	-7	-31	-7	3	2	-8

Step 1 $H_0 : \mu_d = 0$ vs. $H_1 : \mu_d < 0$. One-sided test.

Step 2 Choose $\alpha = 0.01$.

Step 3 Test-statistic is $t = -2.191$.

Step 4 Find the critical region: $t \leq -2.821$

Step 5 Do we reject H_0? No, $t = -2.191$ is not in the critical region.