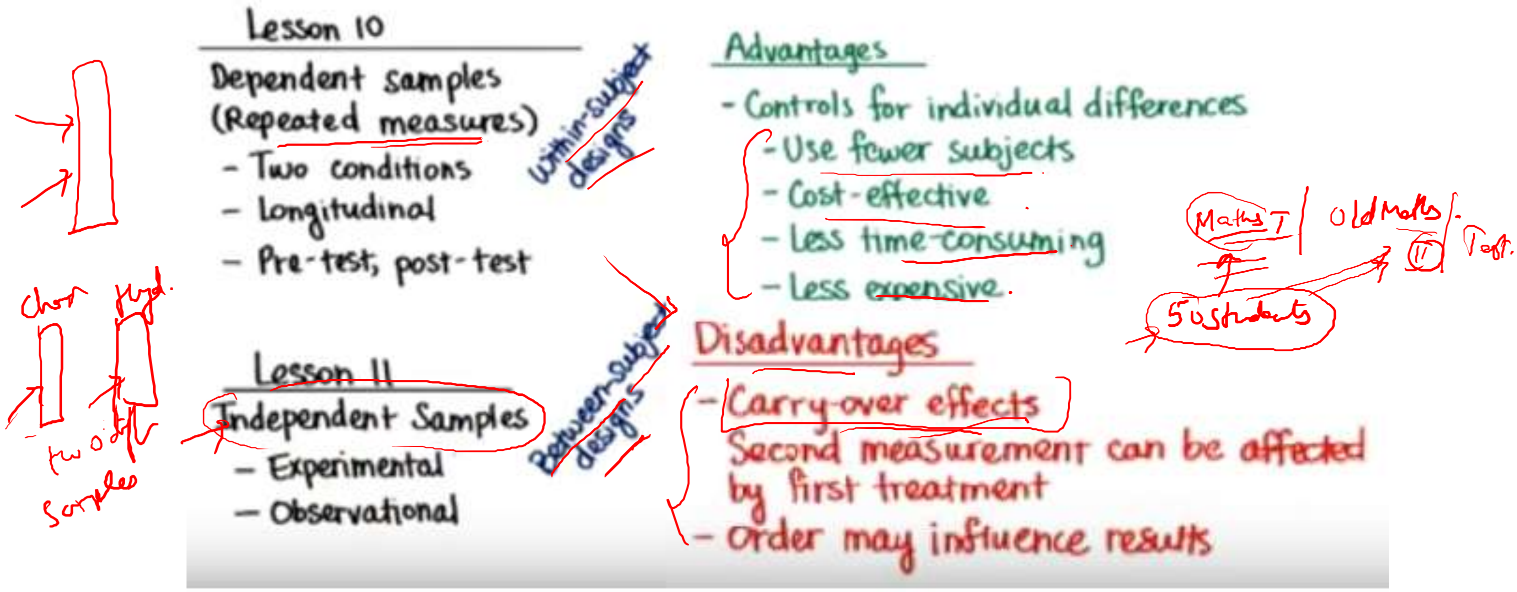
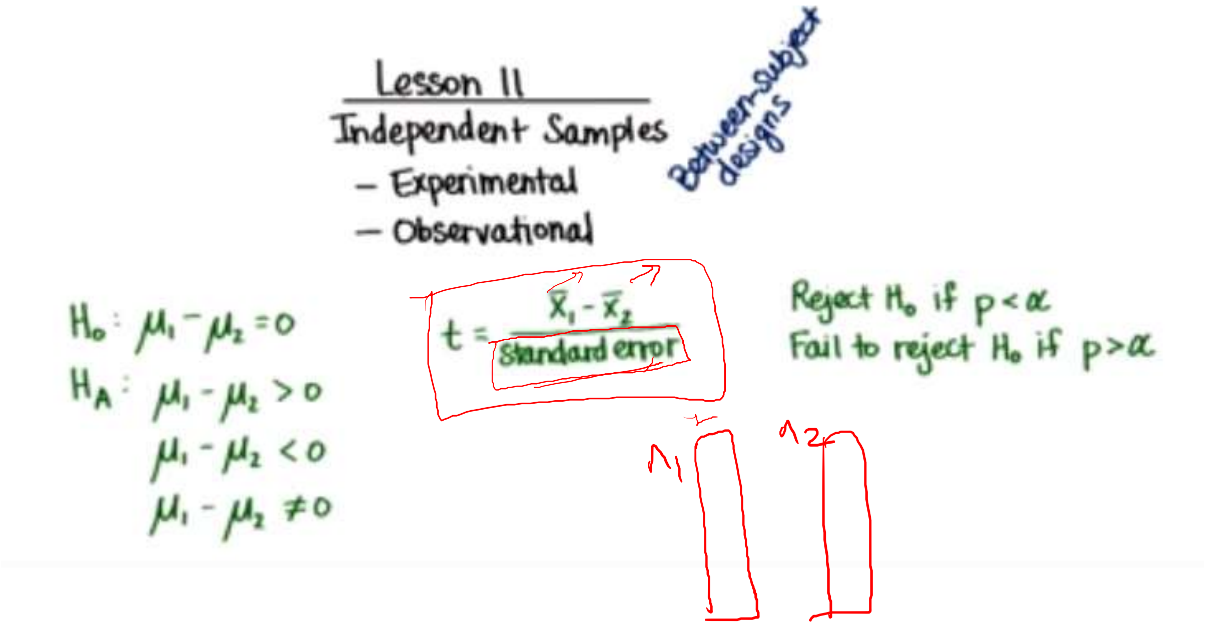
Notes for Students – Lesson 11



Everything is exactly the same



Standard Error

Q1.

https://docs.google.com/spreadsheets/d/1Uu4dlVitf

$$t\text{-statistic} = \frac{\text{Difference between means}}{\text{Standard error}}$$

$H_0: \mu_1 - \mu_2 = 0$

$\mu_A: \mu_1 - \mu_2 \neq 0$

Subtract two normally distributed samples:

$SD = \sqrt{S_1^2 + S_2^2}$

independent samples

$S.E = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

$$\text{Standard error} = \frac{S}{\sqrt{n}} = \frac{\sqrt{S_1^2 + S_2^2}}{\sqrt{n}} = \sqrt{\frac{S_1^2 + S_2^2}{n}} = \boxed{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$df = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$$

CqNT4n1LBUdt93N_S5PJdDlvtiyyrPYlE0/edit?usp=sharing

You and your friends want to go out to eat, but you don't want to pay a lot. You decide to either go to Gettysburg or Wilma. You look online and find the average meal prices at 18 restaurants in Gettysburg and 14 restaurants in Wilma.

$H_0: \mu_G = \mu_W$

$H_A: \mu_G \neq \mu_W$

$t = \frac{\bar{x}_1 - \bar{x}_2}{s/\sqrt{n}}$

$\frac{s}{\sqrt{n}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

What do we need to know to compare these Samples?

- [ ] The population mean for food prices at all restaurants in Gettysburg and Wilma
- [x] The Sample averages
- [ ] The population standard deviation for food prices at all restaurants in Gettysburg and Wilma
- [x] The size of each sample
- [x] The sample standard deviations

Q2.Part 1

What is the average meal price for each sample of restaurants?

Gettysburg:

Wilma:

Q2.Part 2

You and your friends want to go out to eat, but you don't want to pay a lot. You decide to either go to Gettysburg or Wilma. You look online and find the average meal prices at 18 restaurants in Gettysburg and 14 restaurants in Wilma.

$H_0: \mu_G = \mu_W \qquad \bar{x}_G = 8.94$

$H_A: \mu_G \neq \mu_W \qquad \bar{x}_W = 11.14$

What is the sample standard deviation for each sample of restaurants?

Gettysburg:

Wilma:

Q2. Part 3

You and your friends want to go out to eat, but you don't want to pay a lot. You decide to either go to Gettysburg or Wilma. You look online and find the average meal prices at 18 restaurants in Gettysburg and 14 restaurants in Wilma.

$H_0: \mu_G = \mu_W \qquad \bar{x}_G = 8.94 \qquad s_G = 2.65$

$H_A: \mu_G \neq \mu_W \qquad \bar{x}_W = 11.14 \qquad s_W = 2.18$

Calculate the standard error.

$$S_{\bar{x}_G - \bar{x}_W} = \sqrt{\frac{s_G^2}{n_G} + \frac{s_W^2}{n_W}} =$$

Q2. Part 4

You and your friends want to go out to eat, but you don't want to pay a lot. You decide to either go to Gettysburg or Wilma. You look online and find the average meal prices at 18 restaurants in Gettysburg and 14 restaurants in Wilma.

$H_0: \mu_G = \mu_W$ $\quad \bar{x}_G = 8.94 \quad s_G = 2.65$

$H_A: \mu_G \neq \mu_W$ $\quad \bar{x}_W = 11.14 \quad s_W = 2.18$

$s_{\bar{x}_G - \bar{x}_W} = 0.85$

Which is correct for calculating the t-statistic?

- $\dfrac{\bar{x}_G - \bar{x}_W}{s_{\bar{x}_G - \bar{x}_W}}$
- $\dfrac{\bar{x}_W - \bar{x}_G}{s_{\bar{x}_G - \bar{x}_W}}$
- Both are correct

Q2 Part 5

Calculate the t-statistic both ways.

$$t = \frac{\bar{x}_G - \bar{x}_W}{s_{\bar{x}_G - \bar{x}_W}} \qquad t = \frac{\bar{x}_W - \bar{x}_G}{s_{\bar{x}_G - \bar{x}_W}}$$

$$= \qquad\qquad =$$

Q2 Part 6

You and your friends want to go out to eat, but you don't want to pay a lot. You decide to either go to Gettysburg or Wilma. You look online and find the average meal prices at 18 restaurants in Gettysburg and 14 restaurants in Wilma.

$H_0: \mu_G = \mu_W$ $\bar{X}_G = 8.94$ $S_G = 2.65$ $S_{\bar{X}_G - \bar{X}_W} = 0.85$

$H_A: \mu_G \neq \mu_W$ $\bar{X}_W = 11.14$ $S_W = 2.18$ $t = \pm 2.58$

Find the t-critical values for a two-tailed test at $\alpha = 0.05$

± 2.042

(Hint: Degrees of freedom $= n_G + n_W - 2$)

Q2 Part 7

You and your friends want to go out to eat, but you don't want to pay a lot. You decide to either go to Gettysburg or Wilma. You look online and find the average meal prices at 18 restaurants in Gettysburg and 14 restaurants in Wilma.

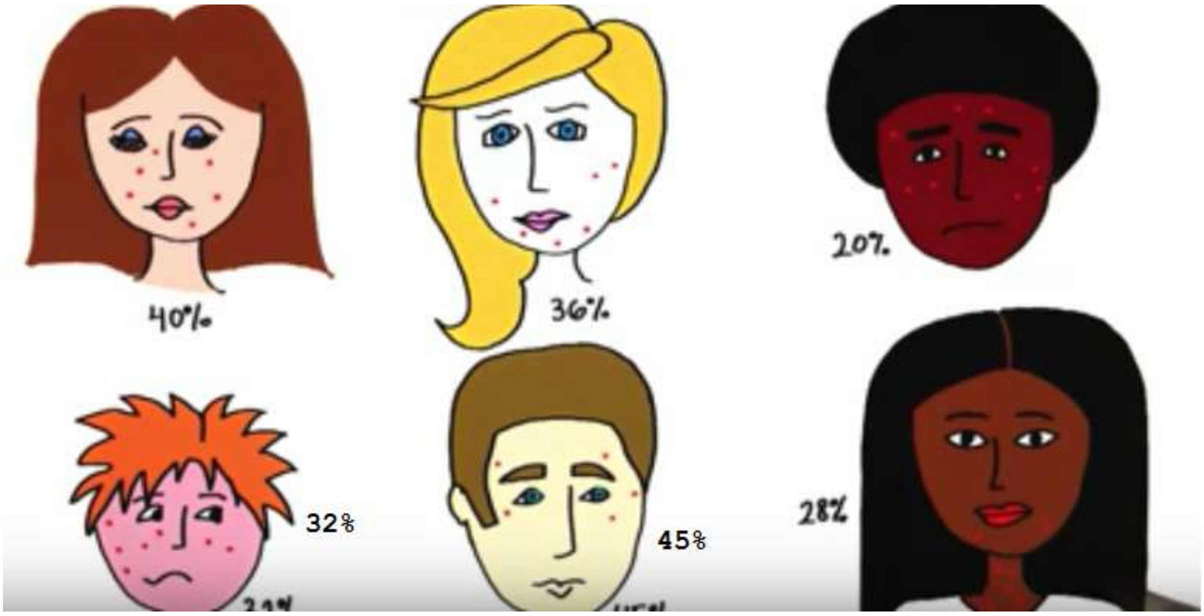
$H_0: \mu_G = \mu_W$ $\bar{X}_G = 8.94$ $S_G = 2.65$ $S_{\bar{X}_G - \bar{X}_W} = 0.85$

$H_A: \mu_G \neq \mu_W$ $\bar{X}_W = 11.14$ $S_W = 2.18$ $t = \pm 2.58$

t-critical values: ± 2.042

- Retain the null
- Reject the null ✓

Q3.



A dermatologist has developed a drug to get rid of acne. Let's call their drug drug A. He tested it out on six people.After four weeks, the following proportion of acne had disappeared from these subjects' faces, 40%, 36%, 20%, 32%, 45% and 28%.

A competing dermatologist also developed a drug to get rid of acne, which we'll call drug B. She tested it on five people. After four weeks, the following proportions of acne had disappeared from these subjects faces. 41%, 39%, 18%, 23%, and 35%.

We'll do a two-tailed test to see if the effects of these two drugs are significantly different.

Drug A	Drug B
40%	41%
36%	39%
20%	18%
32%	23%
45%	35%
28%	

$\bar{x}_A = 33.5\%$ $\quad$ $\bar{x}_B = 31.2\%$

$s_A = 8.89\%$ $\quad$ $s_B = 10.16\%$

$H_0: \mu_A = \mu_B$

$H_A: \mu_A \neq \mu_B$

Part 1

Calculate the t-statistic.

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} =$$

Part 2

Find t-critical values

Part 3

Accept or reject the null?

- $H_0: \mu_A = \mu_B$
- $H_A: \mu_A \neq \mu_B$

$t < t_{critical}$.

t-statistic = 0.4

t-critical values $\alpha = 0.05$

± 2.262

Q4.

https://docs.google.com/spreadsheets/d/1q8ZytYAcd4HoX6kKQrrQ7_Odp7JdcbJu7e_Tk8nBWmo/edit?usp=sharing

Who has more shoes - males or females?

$H_0: \mu_F = \mu_M \quad (\mu_F - \mu_M = 0)$

$H_A: \mu_F < \mu_M \quad (\mu_F - \mu_M < 0)$

$\mu_F > \mu_M \quad (\mu_F - \mu_M > 0)$

$\boxed{\mu_F \neq \mu_M \quad (\mu_F - \mu_M \neq 0)}$

We're going to do a t-test to see if men or women have more shoes.

Part 1

Who has more shoes - males or females?

$H_0: \mu_F = \mu_M \quad (\mu_F - \mu_M = 0)$

$H_A: \mu_F \neq \mu_M \quad (\mu_F - \mu_M \neq 0)$

$\bar{x}_F =$ $\bar{x}_M =$

$S_F =$ $S_M =$

Part 2

Who owns more shoes?

$H_0: \mu_F - \mu_M = 0$ $\bar{X}_F = 33.14$ $\bar{X}_M = 18$

$H_A: \mu_F - \mu_M \neq 0$ $S_F = 31.36$ $S_M = 34.27$

$n_F = 7$ $n_M = 11$

$$\text{Standard error} = \sqrt{\frac{S_F^2}{n_F} + \frac{S_M^2}{n_M}} =$$

Part 3

Who owns more shoes?

$H_0: \mu_F - \mu_M = 0$ $\bar{X}_F = 33.14$ $\bar{X}_M = 18$

$H_A: \mu_F - \mu_M \neq 0$ $S_F = 31.36$ $S_M = 34.27$

$n_F = 7$ $n_M = 11$

$$\text{Standard error} = \sqrt{\frac{S_F^2}{n_F} + \frac{S_M^2}{n_M}} = 15.72 \qquad t\text{-statistic} = \frac{\bar{X}_F - \bar{X}_M}{SE} =$$

Part 4

Who owns more shoes?

$H_0: \mu_F - \mu_M = 0$ $\bar{X}_F = 33.14$ $\bar{X}_M = 18$

$H_A: \mu_F - \mu_M \neq 0$ $S_F = 31.36$ $S_M = 34.27$

$n_F = 7$ $n_M = 11$

$$\text{Standard error} = \sqrt{\frac{S_F^2}{n_F} + \frac{S_M^2}{n_M}} = 15.72 \qquad t\text{-statistic} = \frac{\bar{X}_F - \bar{X}_M}{SE} = 0.96$$

t-critical value

- Accept H_0
- Reject H_0

Part 5. Calculate a 95% confidence interval for the true difference between pairs of shoes owned by males and females.

5a)

Who owns more shoes?

$H_0: \mu_F - \mu_M = 0$ $\qquad \bar{x}_F = 33.14$ $\qquad \bar{x}_M = 18$

$H_A: \mu_F - \mu_M \neq 0$ $\qquad S_F = 31.36$ $\qquad S_M = 34.27$

$n_F = 7$ $\qquad n_M = 11$

$$\text{Standard error} = \sqrt{\frac{S_F^2}{n_F} + \frac{S_M^2}{n_M}} = 15.72 \qquad t\text{-statistic} = \frac{\bar{x}_F - \bar{x}_M}{SE} = 0.96$$

t-critical value $(\alpha = 0.05) = 2.12$

$$CI: \; \bar{x} \pm t \cdot SE$$

$\bar{x}$	t	SE
☐ 33.14	☐ 2.12	☐ 31.36
☐ 18	☐ -2.12	☐ 34.27
☐ 15.14	☐ 0.96	☐ 15.72
☐ -15.14	☐ -0.96	☐ -15.72

5b)

Who owns more shoes?

$H_0: \mu_F - \mu_M = 0$ $\qquad \bar{x}_F = 33.14$ $\qquad \bar{x}_M = 18$

$H_A: \mu_F - \mu_M \neq 0$ $\qquad S_F = 31.36$ $\qquad S_M = 34.27$

$n_F = 7$ $\qquad n_M = 11$

$$\text{Standard error} = \sqrt{\frac{S_F^2}{n_F} + \frac{S_M^2}{n_M}} = 15.72 \qquad t\text{-statistic} = \frac{\bar{x}_F - \bar{x}_M}{SE} = 0.96$$

t-critical value $(\alpha = 0.05) = 2.12$

$$CI: \; \bar{x} \pm t \cdot SE$$

$\bar{x}$	t	SE
☐ 33.14	☑ 2.12	☐ 31.36
☐ 18	☑ -2.12	☐ 34.27
☑ 15.14	☐ 0.96	☑ 15.72
☑ -15.14	☐ -0.96	☐ -15.72

95% CI:

(__________ , __________)

lower bound, upper bound

Part 6

Who owns more shoes?

$H_0: \mu_F - \mu_M = 0$ $\quad \bar{x}_F = 33.14 \quad \bar{x}_M = 18$

$H_A: \mu_F - \mu_M \neq 0$ $\quad S_F = 31.36 \quad S_M = 34.27$

$n_F = 7 \quad n_M = 11$

$$\text{Standard error} = \sqrt{\frac{S_F^2}{n_F} + \frac{S_M^2}{n_M}} = 15.72 \qquad t\text{-statistic} = \frac{\bar{x}_F - \bar{x}_M}{SE} = 0.96$$

t-critical value $(\alpha = 0.05) = 2.12$

What proportion of the difference in pairs of shoes owned can be attributed to gender? $\quad \bar{x}_F - \bar{x}_M = 15$

$$r^2 = \frac{t^2}{t^2 + df} = \underline{\underline{0.054}}$$

Pooled Variance

$$\text{Standard error} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \qquad df = n_1 + n_2 - 2$$

Assumes samples are approximately the same size.

One sample

$$S^2 = \frac{SS}{df} = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$

$$\text{Pooled variance} = S_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{(n_1-1)\,Var1 + (n_2-1)\,Var2}{df_1 + df_2}$$

$$= \frac{(n_1-1)\,s_1^2 + (n_2-1)\,s_2^2}{(n_1+n_2-2)}$$

Q5 Part 1

$$\text{Pooled variance} = S_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$

X	$X-\bar{X}$	$(X-\bar{X})^2$	Y	$Y-\bar{Y}$	$(Y-\bar{Y})^2$
5	3	9	3	-3	9
6	4	16	7	1	1
1	-1	1	8	2	4
-4	-6	36			

$SS_X = \Sigma(x_i - \bar{x})^2 = \boxed{62}$ $\qquad SS_Y = \Sigma(Y_i - \bar{Y})^2 = \boxed{14.}$

$$S_p^2 = \frac{62+14}{3+2} = \frac{76}{5} = 15.2$$

Q5 Part 2

$$\text{Pooled variance} = S_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} =$$

X	$X-\bar{X}$	$(X-\bar{X})^2$	Y	$Y-\bar{Y}$	$(Y-\bar{Y})^2$
5	3	9	3	−3	9
6	4	16	7	1	1
1	−1	1	8	2	4
−4	−6	36			

$SS_x = \Sigma(x_i - \bar{x})^2 = \boxed{62}$ $\qquad$ $SS_Y = \Sigma(Y_i - \bar{Y})^2 = \boxed{14}$

$\bar{x} = 2$ $\qquad$ $\bar{Y} = 6$

Part 3

$$\text{Pooled variance} = S_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{62 + 14}{5} = \underline{\underline{15.2}}$$

X	Y
5	3
6	7
1	8
−4	

$\bar{x} = 2$ $\qquad$ $\bar{Y} = 6$

$SS_x = \Sigma(x_i - \bar{x})^2 = 62$

$SS_Y = \Sigma(Y_i - \bar{Y})^2 = 14$

$$\text{Standard error} = \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}} =$$

$$= \sqrt{\frac{15.2}{4} + \frac{15.2}{3}} = \underline{\underline{2.98}}$$

Part 4

$$\text{Pooled variance} = S_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{62+14}{5} = 15.2$$

X	Y
5	3
6	7
1	8
-4	
$\bar{X} = 2$	$\bar{Y} = 6$

$SS_x = \Sigma(x_i - \bar{x})^2 = 62$

$SS_y = \Sigma(Y_i - \bar{Y})^2 = 14$

$$\text{Standard error } (S_{\bar{x}-\bar{y}}) = \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}} = \sqrt{\frac{15.2}{4} + \frac{15.2}{3}} = 2.98$$

$$t\text{-statistic} = \frac{(\bar{X}-\bar{Y}) - (\cancel{\mu_x - \mu_Y})^{0}}{S_{\bar{x}-\bar{y}}} = \frac{-4}{2.98 \simeq 3} = -1.33.$$

Part 5

$$\text{Pooled variance} = S_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{62+14}{5} = 15.2$$

X	Y
5	3
6	7
1	8
-4	
$\bar{X} = 2$	$\bar{Y} = 6$

$SS_x = \Sigma(x_i - \bar{x})^2 = 62$

$SS_y = \Sigma(Y_i - \bar{Y})^2 = 14$

$$\text{Standard error } (S_{\bar{x}-\bar{y}}) = \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}} = \sqrt{\frac{15.2}{4} + \frac{15.2}{3}} = 2.98$$

t-statistic $= -1.34$

t-critical value $= \pm \text{tinv}(0.05, 5) = \pm 2.571.$ $(\alpha = 0.05)$

$df = n_x + n_y - 2 = df_x + df_y$

o ~~Accept~~ the null $(H_0: \mu_x = \mu_Y)$

o Reject the null $(H_A: \mu_x \neq \mu_Y)$