ECON10005 Quantitative Methods 1

Tutorial in Week 10

Zheng Fan

The University of Melbourne

Introduction

18th May Thursday. in week.

$\boxed{\text{2pm}}$

Zheng Fan

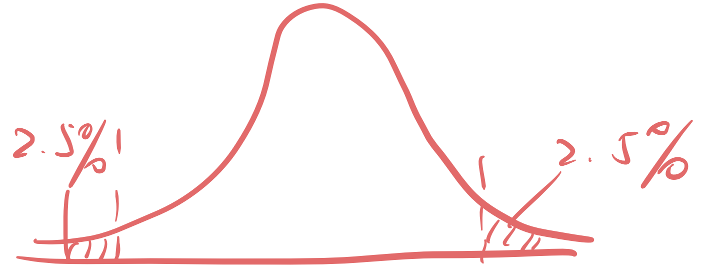
- Ph.D student in Economics at Unimelb
- Email me with subject code titled: fan.z@unimelb.edu.au

Don't be shy if you need help

- Visit Ed Discussion Board (read others' questions first)
- Lecturer's consultation sessions: see Canvas
- In case of special considerations, consult Stop 1
- For admin issues contact Chin via qm1-economics@unimelb.edu.au

But before asking any questions, make sure you have read the Ed discussion board, subject guide, announcements and etc on Canvas!!!

A. Confidence Interval



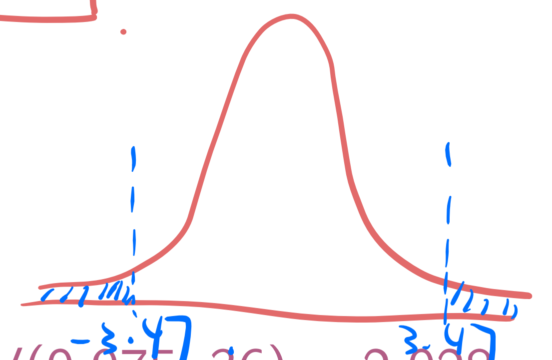
Given in this questions $n = 37$, level of significance $\alpha = 5\%$, $\boxed{s_D = 13.69}$

	Baseline	18 Months
Mean	58.81	51.00
Standard Deviation	~~15.53~~	~~15.35~~

You have done a hypothesis testing for the $\boxed{\text{matched pairs}}$ in the pre-quiz.

$\mu_{18} - \mu_{Base} = 0$.

$$H_0 : \mu_D = 0 \qquad H_1 : \mu_D \neq 0$$



Test statistics is $t = (\bar{D} - 0)/(s_D/\sqrt{n}) = -3.470$;

Critical value is $t_{0.025,36} = \text{T.INV.2T}(0.05, 36) = \text{T.INV}(\underline{0.975}, 36) = 2.028$;

df = n - 1.

p-value $= \text{T.DIST.2T}(3.470, 36) = 0.001 < 0.05$;

$= 2 \times \text{T.DIST}(-3.47, 36)$

Reject the null. $\rightarrow$ Will you make the same decision with CI? Discuss.

$[\bar{x} \pm t_{c,v}\, se(\bar{x})]$

A. Confidence Interval

$\mu \in CI \longrightarrow$ cannot reject

$\mu \notin CI \longrightarrow$ reject

Let's verify the results using confidence interval.

The 95% confidence interval is

$$
\begin{aligned}
& \left[\bar{D} \pm t_{\alpha/2,n-1} \times \text{s. e.}(\bar{D})\right] \\
= & \left[\bar{D} \pm t_{0.025,36} \times \frac{s_D}{\sqrt{n}}\right] \\
= & \left[-7.81 \pm 2.028 \times \frac{13.69}{\sqrt{37}}\right] \\
= & [-12.374, -3.246]
\end{aligned}
$$

Does not contain $\mu_D = 0$, rejecting the null at 5% significance level for a two-tail test.

$H_0: \mu_D = 0 \notin CI \longrightarrow$ reject.

B. Hypothesis Testing

Given $n_E = 158$, $n_L = 93$, level of significance $\alpha = 5\%$, mean and sd below

	Machiavellianism	Narcissism	Psychopathy
Economics/ Business	3.00(0.63)	2.98(0.50)	2.22(0.61)
Law	2.75(0.73)	2.84(0.49)	2.05(0.59)

You should have done hypothesis testings for the independent samples in the pre-quiz. Let's revise these results.

B. Hypothesis Testing

$var(aX+bY) = a^2 var(X) + \cdots$

$se(\bar{D})^2 = var(\bar{D}) = var(\bar{X}_E - \bar{X}_L) = var(\bar{X}_E) + var(\bar{X}_L)$

$= se(\bar{X}_E)^2 + se(\bar{X}_L)^2$

Given $n_E = 158$, $n_L = 93$, level of significance $\alpha = 5\%$, mean and sd below

	Machiavellianism	Narcissism	Psychopathy
Economics/ Business	3.00(0.63)	2.98(0.50)	2.22(0.61)
Law	2.75(0.73)	2.84(0.49)	2.05(0.59)

$var(\bar{x}) = \frac{\sigma^2}{n} = var\left(\frac{1}{n}\sum x_i\right) = \frac{1}{n^2} \cdot n\, var(x_i)$

$\bar{D} = \bar{X}_E - \bar{X}_L = 0.25;$

$se(\bar{D}) = \sqrt{se(\bar{X}_E)^2 + se(\bar{X}_L)^2} = \sqrt{s_X^2/n_E + s_L^2/n_L} = 0.0908$

The test statistics is $t = 0.25/0.0908 = 2.7537$

$t = \frac{\bar{D} - \mu_D}{se(\bar{D})}$

p-value = T.DIST.2T(2.7537, 158 + 93 - 2) = 0.0063 < 0.05

$= 2 \times$ T.DIST$(-2.7537, df)$

Reject H_0.

$df = n_1 + n_2 - 2$

B. Hypothesis Testing

Given $n_E = 158$, $n_L = 93$, level of significance $\alpha = 5\%$, mean and sd below

	Machiavellianism	Narcissism	Psychopathy
Economics/ Business	3.00(0.63)	2.98(0.50)	2.22(0.61)
Law	2.75(0.73)	2.84(0.49)	2.05(0.59)

$\bar{D} = \bar{X}_E - \bar{X}_L = 0.14;$

$se(\bar{D}) = \sqrt{se(\bar{X}_E)^2 + se(\bar{X}_L)^2} = \sqrt{s_X^2/n_E + s_L^2/n_L} = 0.0645$

The test statistics is $t = 0.25/0.0908 = 2.1696$

p-value = T.DIST.2T(2.1696, 158 + 93 - 2) $= 0.0310 < 0.05$

Reject H_0.

B. Hypothesis Testing

Given $n_E = 158$, $n_L = 93$, level of significance $\alpha = 5\%$, mean and sd below

	Machiavellianism	Narcissism	Psychopathy
Economics/ Business	3.00(0.63)	2.98(0.50)	2.22(0.61)
Law	2.75(0.73)	2.84(0.49)	2.05(0.59)

$\bar{D} = \bar{X}_E - \bar{X}_L = 0.17;$

$se(\bar{D}) = \sqrt{se(\bar{X}_E)^2 + se(\bar{X}_L)^2} = \sqrt{s_X^2/n_E + s_L^2/n_L} = 0.0781$

The test statistics is $t = 0.25/0.0908 = 2.1770$

p-value = T.DIST.2T(2.1770, 158 + 93 - 2) = 0.0304 < 0.05

Reject H_0.

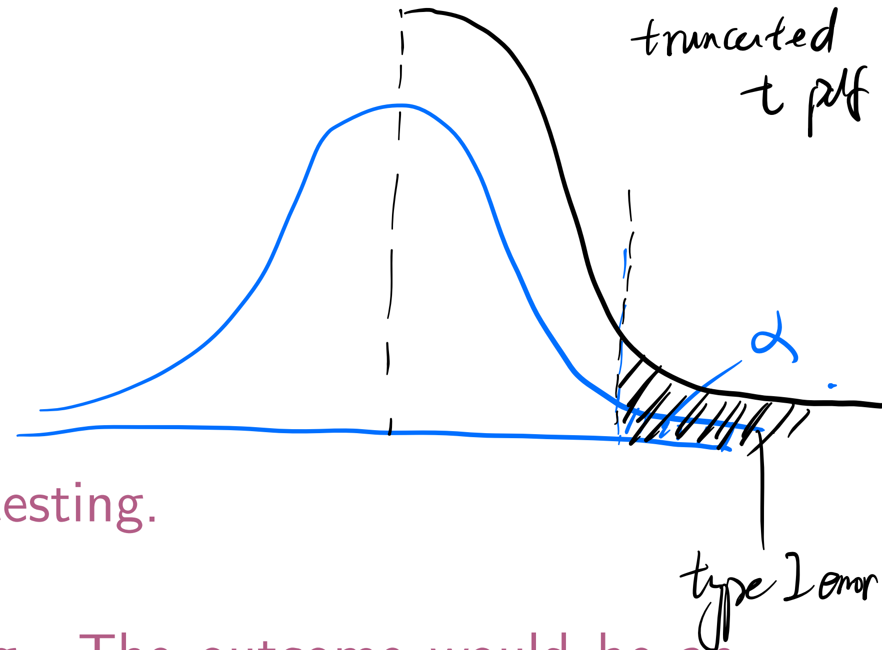
B. Hypothesis Testing

Given $n_E = 158$, $n_L = 93$, level of significance $\alpha = 5\%$, mean and sd below

	Machiavellianism	Narcissism	Psychopathy
Economics/ Business	3.00(0.63)	2.98(0.50)	2.22(0.61)
Law	2.75(0.73)	2.84(0.49)	2.05(0.59)

After carrying out the Machiavellianism test, does the result suggest it might be more appropriate to next carry out a one tail test for Narcissism? That is, do the higher scores for the Economics / Business students suggest we should carry an upper tailed test? Discuss.

B. Hypothesis Testing



Short answer: Never use your data twice in testing.

This is not valid logic for hypothesis testing. The outcome would be an increased rate of Type I errors relative to the significance level, i.e. t-pdf becomes a truncated t-pdf (beyond QM1).

Think about an analogy: good archer is able to shoot a target that he pointed at first. Not to point at the thing where his arrow landed later after his random shot.

C. Hypothesis Testing

A psychologist theorises that more male than female drivers will continue to drive when lost, hoping to find the location they seek rather than ask for directions. To test this theory, she took a random sample of 350 male drivers and 280 female drivers and asked each what they did when lost. Of these, 293 of the males and 202 of the females said they continue driving. Does this support the psychologist's theory? Use the 1% level of significance.

Define the notation

p_1: population proportion of male who will continue to drive when lost
p_2: population proportion of female who will continue to drive when lost

The hypotheses are

$$H_0 : p_1 = p_2 \qquad H_1 : p_1 > p_2$$

C. Hypothesis Testing

$H_0: p_1 = 50\%$

A psychologist theorises that more male than female drivers will continue to drive when lost, hoping to find the location they seek rather than ask for directions. To test this theory, she took a random sample of 350 male drivers and 280 female drivers and asked each what they did when lost. Of these, 293 of the males and 202 of the females said they continue driving. Does this support the psychologist's theory? Use the 1% level of significance.

Define the notation

p_1: population proportion of male who will continue to drive when lost
p_2: population proportion of female who will continue to drive when lost

The hypotheses are

$$H_0: p_1 = p_2 \quad H_1: p_1 > p_2 \longrightarrow \boxed{\text{upper tail test.}}$$

This is an independent samples situation.

C. Hypothesis Testing

$\bar{X} \leftarrow \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$, $\bar{X} = \hat{p}$.

$se(\hat{p}_1)^2 = var(\hat{p}_1) = var(\bar{X}_1) = var(\frac{1}{n} \sum X_i) = \frac{1}{n^2} \cdot n\, var(X_i) = \frac{p(1-p)}{n}$.

X_i	0	1
Pr	$1-p$	p

$Ber(p)$: $Var(X_i) = p(1-p)$

$Bino(p, n)$. $Var(X) = np(1-p)$

$X = \sum_{i=1}^{n} X_i$.

The sample statistics are

Males: $n_1 = 350$, $\hat{p}_1 = 293/350 = 0.837$

Females: $n_2 = 280$, $\hat{p}_2 = 202/280 = 0.721$

Standard errors:

$se(\hat{p}_1) = \sqrt{\hat{p}_1(1 - \hat{p}_1)/n_1} = 0.020$, $se(\hat{p}_2) = \sqrt{\hat{p}_2(1 - \hat{p}_2)/n_2} = 0.027$

$se(\hat{p}_1 - \hat{p}_2) = \sqrt{se(\hat{p}_1)^2 + se(\hat{p}_2)^2} = 0.033$

Test statistics:

$$t = \frac{\hat{p}_1 - \hat{p}_2}{se(\hat{p}_1 - \hat{p}_2)} = 3.477$$

p-value = T.DIST.RT(3.477, 628) = 0.0003 < 0.01.

= T.DIST(-3.477, 628)

Reject H_0, there is evidence at the 1% level of significance that a higher proportion of male drivers than female drivers continue to drive when lost.

The end

Thanks for your attention! 😃

Feel free to leave and see you next week!