ECON10005 Quantitative Methods 1

Tutorial in Week 10

Zheng Fan

The University of Melbourne

Introduction

Zheng Fan

- Ph.D student in Economics at Unimelb. Research interest in Bayesian Econometrics
- Personal website: *zhengfan.site* for some details

Don't be shy if you need help

- Discuss on Ed Discussion Board
- Attend consultation sessions: see Canvas for time and location
- Consult Stop 1, in case of special considerations,
- Contact QM-1@unimelb.edu.au for admin issues
- Send me an email: fan.z@unimelb.edu.au

Housekeeping

- MST Reminder: You should now be able to view your results for the MST
- Final Report Reminders: The final report instructions are now available on LMS
  - Each student should attempt to complete all tasks, discuss them among the group and then write the report. You should arrange a time outside of class to meet.
  - Please remember to discuss the feedback for your draft report with your group and incorporate this feedback in the final report.
  - You should submit both a PDF of your report and an Excel file.
  - Remember to include all graphs and tables in your final PDF report.

Housekeeping

- Final Report Reminders: The final report instructions are now available on LMS
  - There should also be an appendix in your report with all steps of the hypothesis tests and confidence intervals.
  - Do not include any steps or calculations for hypothesis testing in the main report; only use the results from these tests (found in the Appendix) for your discussions.
  - You will be required to individually submit the group contribution survey (as you did with the draft report).
  - Do you have any questions?

Recap: Decisions in Hypothesis Testing

Value of μ in the population	**Do not reject H_0**	**Reject H_0**
$\mu = m$	**Correct**	**Type I Error**
$\mu \neq m$	**Type II Error**	**Correct**

Recap: Decisions in Hypothesis Testing

Value of μ in the population	**Do not reject H_0**	**Reject H_0**
$\mu = m$	**Correct**	**Type I Error**
$\mu \neq m$	**Type II Error**	**Correct**

We have the following

$$\alpha = \text{ Type I error } = \Pr(\text{ reject } H_0 \mid H_0 \text{ is true })$$

$$\beta = \text{ Type II error } = \Pr(\text{do not reject } H_0 \mid H_1 \text{ is true })$$

Recap: Decisions in Hypothesis Testing

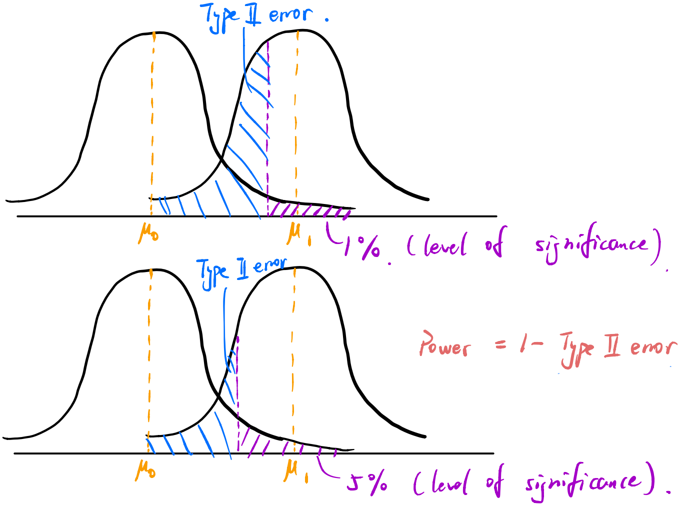
Value of μ in the population	**Do not reject H_0**	**Reject H_0**
$\mu = m$	**Correct**	**Type I Error**
$\mu \neq m$	**Type II Error**	**Correct**

We have the following

$$\alpha = \text{ Type I error } = \Pr(\text{ reject } H_0 \mid H_0 \text{ is true })$$
$$\beta = \text{ Type II error } = \Pr(\text{do not reject } H_0 \mid H_1 \text{ is true })$$

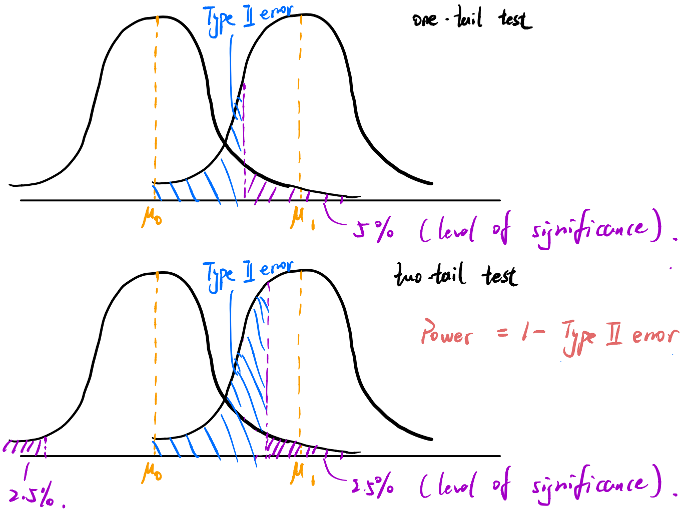
We also have $power = 1 - \beta$

Recap: 1-tail example



Power = 1 − Type II error

Recap: 2-tail example



Question 1

Discuss the questions and answers you have for pre-tutorial 9, question 1.

A researcher is investigating whether a new property valuation model is more accurate than the current model. They take a random sample of 200 recently sold houses in Melbourne and compare the accuracy of the two models using hypothesis testing. Under the null hypothesis, the researcher states that there is no difference in prediction accuracy between the new and current models. The test is conducted at a 5% level of significance.

a. Suppose the data led the researcher to reject the null hypothesis. Describe what a Type I error would mean in this context. What role does the level of significance play in this type of error?

Question 1

a. Suppose the data led the researcher to reject the null hypothesis. Describe what a Type I error would mean in this context. What role does the level of significance play in this type of error?

Question 1

a. Suppose the data led the researcher to reject the null hypothesis. Describe what a Type I error would mean in this context. What role does the level of significance play in this type of error?

- A Type I error occurs when we reject a true null hypothesis. ie

$$\alpha = \text{ Type I error } = \Pr(\text{ reject } H_0 \mid H_0 \text{ is true })$$

Question 1

a. Suppose the data led the researcher to reject the null hypothesis. Describe what a Type I error would mean in this context. What role does the level of significance play in this type of error?

- A Type I error occurs when we reject a true null hypothesis. ie

$$\alpha = \text{ Type I error } = \Pr(\text{ reject } H_0 \mid H_0 \text{ is true })$$

- In this case, since the data led the researcher to reject the null hypothesis, a Type I error would mean they concluded that the new property valuation model is more accurate than the current model, even though, in reality, there MAY BE NO actual difference in accuracy.

Question 1

a. Suppose the data led the researcher to reject the null hypothesis. Describe what a Type I error would mean in this context. What role does the level of significance play in this type of error?

- A Type I error occurs when we reject a true null hypothesis. ie

$$\alpha = \text{ Type I error } = \Pr(\text{ reject } H_0 \mid H_0 \text{ is true })$$

- In this case, since the data led the researcher to reject the null hypothesis, a Type I error would mean they concluded that the new property valuation model is more accurate than the current model, even though, in reality, there MAY BE NO actual difference in accuracy.
- The level of significance (5%) represents the risk the researcher is willing to take (of making this kind of error). This means that there is a 5% chance of rejecting the null hypothesis when it is actually true.

Question 1

Discuss the questions and answers you have for pre-tutorial 9, question 1.

A researcher is investigating whether a new property valuation model is more accurate than the current model. They take a random sample of 200 recently sold houses in Melbourne and compare the accuracy of the two models using hypothesis testing. Under the null hypothesis, the researcher states that there is no difference in prediction accuracy between the new and current models. The test is conducted at a 5% level of significance.

b. Suppose the data led the researcher not to reject the null hypothesis. Describe what a Type II error would mean in this context. What role does the level of significance play in this type of error?

Question 1

b. Suppose the data led the researcher not to reject the null hypothesis. Describe what a Type II error would mean in this context. What role does the level of significance play in this type of error?

Question 1

b. Suppose the data led the researcher not to reject the null hypothesis. Describe what a Type II error would mean in this context. What role does the level of significance play in this type of error?

- A Type II error occurs when we fail to reject a false null hypothesis.

$$\beta = \text{ Type II error } = \Pr(\text{do not reject } H_0 \mid H_1 \text{ is true })$$

Question 1

b. Suppose the data led the researcher not to reject the null hypothesis. Describe what a Type II error would mean in this context. What role does the level of significance play in this type of error?

- A Type II error occurs when we fail to reject a false null hypothesis.

$$\beta = \text{ Type II error } = \Pr(\text{do not reject } H_0 \mid H_1 \text{ is true })$$

- In this context, it means the researcher concluded that there is no difference in accuracy between the new and current property valuation models, when in fact, the new model may be more accurate. A Type II error is possible since the data did not lead to rejecting the null hypothesis.

Question 1

b. Suppose the data led the researcher not to reject the null hypothesis. Describe what a Type II error would mean in this context. What role does the level of significance play in this type of error?

- A Type II error occurs when we fail to reject a false null hypothesis.

$$\beta = \text{ Type II error } = \Pr(\text{do not reject } H_0 \mid H_1 \text{ is true })$$

- In this context, it means the researcher concluded that there is no difference in accuracy between the new and current property valuation models, when in fact, the new model may be more accurate. A Type II error is possible since the data did not lead to rejecting the null hypothesis.

- While the significance level (commonly set at 5%) controls the probability of making a Type I error, it does not directly control the probability of a Type II error. However, there is a trade-off: lowering the significance level reduces the chance of a Type I error but can increase the possibility of a Type II error if other factors (like sample size) are not adjusted.

Question 1

Discuss the questions and answers you have for pre-tutorial 9, question 1.

A researcher is investigating whether a new property valuation model is more accurate than the current model. They take a random sample of 200 recently sold houses in Melbourne and compare the accuracy of the two models using hypothesis testing. Under the null hypothesis, the researcher states that there is no difference in prediction accuracy between the new and current models. The test is conducted at a 5% level of significance.

c. Suppose the data led the researcher to reject the null hypothesis. Describe what a Type II error would mean in this context. Is a Type II error possible in this situation? Explain

Question 1

c. Suppose the data led the researcher to reject the null hypothesis. Describe what a Type II error would mean in this context. Is a Type II error possible in this situation? Explain

Value of μ in the population	**Do not reject H_0**	**Reject H_0**
$\mu = m$	**Correct**	**Type I Error**
$\mu \neq m$	**Type II Error**	**Correct**

Question 1

c. Suppose the data led the researcher to reject the null hypothesis. Describe what a Type II error would mean in this context. Is a Type II error possible in this situation? Explain

Value of μ in the population	**Do not reject H_0**	**Reject H_0**
$\mu = m$	**Correct**	**Type I Error**
$\mu \neq m$	**Type II Error**	**Correct**

- A Type II error occurs when we fail to reject a false null hypothesis. However, since the researcher has rejected the null hypothesis, making a Type II error is not relevant to this situation.
- Type II errors can only occur when the decision is not to reject the null hypothesis, even though it is false.

Question 2

Consider question 2 of the pre-tutorial 9. Interpret the interval estimates you constructed for parts "a" and "c".

Consider the two hypothesis tests performed in Lecture 8 using independent samples, where we tested whether women have lower financial literacy than men, and whether the Australian-born population has higher financial literacy than the overseas-born population (Data available in financial lit scores-tutorial 9). You will now construct and interpret 90%, 95%, and 99% confidence intervals related to these comparisons. For each of the following scenarios, construct the confidence intervals using appropriate methods:

a. A confidence interval for the average financial literacy for Australian women. (90%)

c. Confidence interval for the mean difference in financial literacy scores between native-born individuals and immigrants (i.e., for $\mu_1 - \mu_2$). (95%)

Question 2 a

We are considering the population mean for a single population.

The random variable X represents the scores women received for the financial literacy test.

μ: Average financial score for the women in the population.

To calculate this 90% confidence interval, we need:

$$\left[\bar{X} \pm t_{\alpha/2,n-1}\frac{s_X}{\sqrt{n}}\right]$$

Question 2 a

Substituting all the statistics into formula

$$\bar{X} = 5.767, \quad \text{s.e. } (\bar{X}) = \frac{s_X}{\sqrt{n}} = 0.166, \quad t_{\alpha/2,n-1} = t_{0.1/2,344-1} = 1.6493$$

$$\alpha = 0.1 \text{ and df} : n - 1 = 344 - 1 = 343$$

The 90% confidence interval is:

$$\begin{aligned}\left[\bar{X} \pm t_{\alpha/2,n-1}\frac{s_X}{\sqrt{n}}\right] &= [5.767 \pm 1.6493 \times 0.166] \\ &= [5.4937, 6.0412]\end{aligned}$$

Question 2 a

At a 90% confidence level, this interval contains the true population mean scores for the women.

You can see this uncertainty more in 95% and 99% confidence intervals (as the confidence level increases, the length of the confidence interval increases as well).

Question 2 c

Here, we have two populations that are involved with two independent samples.

The two random variables are:

X_1 : Financial literacy scores of natives

X_2 : Financial literacy scores of migrants

Define the population parameters:

Mean of X_1 : μ_1 the average scores for natives

Mean of X_2 : μ_2 the average scores for migrants

Sample Statistics

$\bar{X}_1$: sample mean of scores for natives

$\bar{X}_2$: sample mean of scores for immigrants

Question 2 c

To calculate this 95% confidence interval, we need:

$$\left[(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2, n_1+n_2-2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right]$$

Question 2 c

Here are the statistics we had:

	Native	Immigrant
Mean	6.8	6.6
Standard Deviation	3.0	3.2

Sample mean difference is: $\left(\bar{X}_1 - \bar{X}_2\right) = 6.8 - 6.6 = 0.2$

s.e.$\left(\bar{X}_1 - \bar{X}_2\right) = \sqrt{\frac{3^2}{219} + \frac{3.2^2}{468}} = 0.2510, \quad t_{0.025,685} = 1.9634$

The 95% confidence interval is: $[0.2 \pm 1.9634 \times 0.2510] = [-0.326, 0.659]$

Question 2 c

At a 95% confidence level, this interval contains the true population mean difference in financial literacy rates between natives and immigrants.

- The confidence interval is narrow and includes 0.
- 0 indicates that the difference between the mean scores for the two groups is zero.

Question 2 c

At a 95% confidence level, this interval contains the true population mean difference in financial literacy rates between natives and immigrants.

- The confidence interval is narrow and includes 0.
- 0 indicates that the difference between the mean scores for the two groups is zero.

If you conduct a two-tail hypothesis test at a 5% significance level, you will not be able to reject the null hypothesis ($\mu_1 - \mu_2 = 0$), meaning that we don't find evidence that the mean financial literacy scores for these two groups are different.

$\rightarrow$ suggest CI can lead to the same conclusion as a two-tail hypothesis test (cautious).

Question 3

Now consider the hypothesis test from Lecture 8, where we used a matched pairs sample to evaluate whether the training provided to employees at a multinational company was successful. Using the same data, construct a 95% and 99% confidence interval for the mean difference in sales (after training before training).

Mean (DIFF)	147.600
Variance (DIFF)	76228.531
n	50
df	49

- Discuss how the length of the confidence intervals increases when the significance level is reduced.

Question 3

X_1: Sales before training.

X_2: Sales after the training.

Define another random variable D: the difference in sales after and before the training (match pair differences).

$$D = X_2 - X_1$$

μ_D is the mean of the differences in the sales (after training – before training).

Recall that the degrees of freedom for matched pairs experiment is $n - 1$, and there are $n = 50$ employees.

Question 3

Let's spend 5-10 minutes calculating the 95% confidence interval by hand using the whiteboards. The relevant statistics are displayed below.

For practice, use the relevant table (go to the Week 7 Lectures module) to find the critical value.

Here is the formula you would use to calculate the confidence interval.

$$\left[\bar{X}_D \pm t_{\alpha/2,n-1}\sqrt{\frac{s_D^2}{n}}\right]$$

Mean (DIFF)	147.600
Variance (DIFF)	76228.531
n	50
df	49

Question 3

Using the critical value of 2.009 , the 95% confidence interval is:

$$\begin{aligned}
&\left[(\bar{X}_D) \pm t_{\alpha/2,n-1}\sqrt{\frac{S_D^2}{n}}\right] \\
&= \left[147.600 \pm 2.009\sqrt{\frac{76228.531}{50}}\right] \\
&= [69.157, 226.043]
\end{aligned}$$

If we use Excel, the critical value would be 2.010 rather than 2.009. The answer is close to what we calculated before i.e. [69.135, 226.065].

Question 3

Similarly, the 99% confidence interval is:

$$[147.6 \pm 2.680 \times 39.046]$$
$$[42.959, 252.241]$$

The length of the confidence interval is larger for the 99% confidence level compared to the 95% confidence level.

Any final questions?

Thanks for your attention! 😄

Let me know if you have any questions.