ECON10005 Quantitative Methods 1

Tutorial in Week 11

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

Group project reminder

Data Analysis Report - Final

- Follow the new instruction.
- The report is due at 10 a.m. on Monday, 26 May

Some clarifications in case you are confused

- The reference to 'house sizes' means dwelling size.
- The term 'block' refers to land size.
- You only need to carry out at a 5% significance level to receive full marks for hypothesis testing and confidence interval. No need to consider 1% or 10%.

Group project reminder

- There is no fixed definition for what constitutes a ‘small’ or ‘large’ block.
- You should refer back to your analysis in Section 3.2 to justify your classification.
- There is no single correct answer—what matters is that your definition is **reasonable and supported by data**.

Example:

- For the question: “Do we have statistical evidence that larger blocks are more expensive than smaller blocks?”
  - Use the variables `large` and `Price` to form two groups based on land size.
  - Compare the mean prices of these groups.
- A similar binary variable can be created for house size when addressing the first dot point question.

Glossary

- Before we discuss this week's tutorial questions, we will briefly review the glossary of simple linear regression provided in the lecture slides for week 10.
- As was the case for one-sample and two-sample statistical inference, we need to distinguish the **population** and the **sample** in the context of simple linear regression.

Linear regression

Linear model: $Y_i = \beta_0 + \beta_1 X_i$

- Y_i : dependent variable (response, outcome etc)
- X_i : explanatory variable (independent, feature, covariates, regressors)
- Intercept: β_0
- Slope coefficient: β_1

Linear regression model: $Y_i = \beta_0 + \beta_1 X_i + U_i$

Population regression function (PRF): $E\left(Y_i \mid X_i\right) = \beta_0 + \beta_1 X_i$

Error term or disturbance term: U_i

Linear Regression: Estimation

Linear regression model: $E\left(Y_i \mid X_i\right) = \beta_0 + \beta_1 X_i$

How to estimate β_0 and β_1 (the line that best describes the relationship)?

Linear Regression: Estimation

Linear regression model: $E\left(Y_i \mid X_i\right) = \beta_0 + \beta_1 X_i$

How to estimate β_0 and β_1 (the line that best describes the relationship)?

Minimise the sum of squared errors

$$\text{minimize} \qquad \text{SSE} = \sum_{i=1}^{n} \hat{U}_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2$$

yields

$$\hat{\beta}_1 = \frac{s_{XY}}{s_X^2} \quad \text{and} \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

where

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$$

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$$

Linear Regression: Statistical Inference

Standard errors of $\hat{\beta}_1$ and $\hat{\beta}_0$

$$\text{s.e.}\left(\hat{\beta}_1\right) = \frac{s_U}{\sqrt{(n-1)s_X^2}} \quad \text{and} \quad \text{s.e.}\left(\hat{\beta}_0\right) = \frac{s_U}{\sqrt{(n-1)s_X^2}} \cdot \sqrt{\frac{\sum_{i=1}^n X_i^2}{n}}$$

where

$$s_U^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{U}_i^2 = \frac{SSE}{n-2} \quad \text{with} \quad df = n-2$$

Linear Regression: Statistical Inference

Standard errors of $\hat{\beta}_1$ and $\hat{\beta}_0$

$$\text{s.e.}\left(\hat{\beta}_1\right) = \frac{s_U}{\sqrt{(n-1)s_X^2}} \quad \text{and} \quad \text{s.e.}\left(\hat{\beta}_0\right) = \frac{s_U}{\sqrt{(n-1)s_X^2}} \cdot \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n}}$$

where

$$s_U^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{U}_i^2 = \frac{SSE}{n-2} \quad \text{with} \quad df = n-2$$

"Goodness of Fit": $R^2 = VAR.S(\hat{Y}_i)/VAR.S(Y_i)$, which explains the proportion of the variation in Y is explained by the regression on X.

Question 1

Recall that you have two samples of data for quantity supplied and quantity demanded given various prices. The two population regression models are:

- We assumed a linear regression for the quantity supplied (QS) by this particular market as a function of price (P). $QS_i = \beta_0 + \beta_1 P_i + U_i$
- We assumed a linear regression for the quantity demanded (QD) by this particular market as a function of price (P). $QD_i = \beta_0 + \beta_1 P_i + U_i$

Question 1

Recall that you have two samples of data for quantity supplied and quantity demanded given various prices. The two population regression models are:

- We assumed a linear regression for the quantity supplied (QS) by this particular market as a function of price (P). $QS_i = \beta_0 + \beta_1 P_i + U_i$
- We assumed a linear regression for the quantity demanded (QD) by this particular market as a function of price (P). $QD_i = \beta_0 + \beta_1 P_i + U_i$

Using your answers for the pre-tutorial 10, question 1, 2, answer the following questions.

1. Write down the estimated regression lines for quantity demanded and quantity supplied.
2. Interpret the two coefficients on P and discuss what these mean in explaining the quantity supplied and the quantity demanded.
3. Comment on the coefficients of determination.

Let's spend 5-10 minutes discussing this in groups.

Question 1

1. Write down the estimated regression lines for quantity demanded and quantity supplied.

The estimated regression lines are:

$$\widehat{QS}_i = 6.898 + 1.037P_i$$
$$\widehat{QD}_i = 2.644 - 0.120P_i$$

Question 1

2. Interpret the two coefficients on P and discuss what these mean in explaining the quantity supplied and the quantity demanded.

Supply Function:

- The coefficient on P is 1.037.

How do we interpret this coefficient?

Question 1

2. Interpret the two coefficients on P and discuss what these mean in explaining the quantity supplied and the quantity demanded.

Supply Function:

- The coefficient on P is 1.037.

How do we interpret this coefficient?

A positive coefficient indicates that when the price increases, the quantity supplied will also increase.

This estimate suggests that with every $1 increase in price, the quantity supplied increases by 1.037 tonnes on average.

Question 1

Demand Function:

- The coefficient on P is -0.120.

Question 1

Demand Function:

- The coefficient on P is -0.120.

A negative coefficient indicates that when the price increases, the quantity demanded will decrease.

This estimate suggests that with every $1 increase in price, the quantity demanded will decrease by 0.120 tonnes on average.

(It is very important to include “on average” in this interpretation.)

These results align with economic theory: there is a positive relationship between price and quantity supplied, and an inverse relationship between price and quantity demanded.

Question 1

3. Comment on the coefficients of determination. **What does this mean?**

Demand function: The coefficient of determination is 0.922.

Question 1

3. Comment on the coefficients of determination. What does this mean?

Demand function: The coefficient of determination is 0.922.

- This means that 92.2% of the variation in the quantity demanded is explained by the price (or the model).
- This indicates a strong fit, as a large proportion of the variation in demand is explained by the model.

Supply function: the coefficient of determination is 0.171.

Question 1

3. Comment on the coefficients of determination. **What does this mean?**

Demand function: The coefficient of determination is 0.922.

- This means that 92.2% of the variation in the quantity demanded is explained by the price (or the model).
- This indicates a strong fit, as a large proportion of the variation in demand is explained by the model.

Supply function: the coefficient of determination is 0.171.

- This means that 17.1% of the variation in the quantity supplied is explained by the price (or the model). This is a weak fit, as the model explains only a small portion of the variation in quantity supplied.
- It suggests that other factors, aside from price, also play a significant role in determining the supply of potatoes.

Question 2

Refer to the pre-tutorial 10, question 3.

Recall that the spreadsheet Funds.xlsx contains data on the performance of 157 investment funds, with their annual Return (Y) and Risk (X), reported in percentages. Finance theory suggests that Risk and Return are positively related, meaning riskier funds should offer higher average returns.

Question 2

Refer to the pre-tutorial 10, question 3.

Recall that the spreadsheet Funds.xlsx contains data on the performance of 157 investment funds, with their annual Return (Y) and Risk (X), reported in percentages. Finance theory suggests that Risk and Return are positively related, meaning riskier funds should offer higher average returns.

1. Write down the estimated regression line.
2. Interpret the coefficient on X and discuss what it means in explaining the annual return. Does this agree with the finance theory?
3. Comment on the coefficient of determination.

Let's spend 5 minutes discussing this in groups.

Question 2

1. Write down the estimated regression line.

The estimated regression line is:

$$\widehat{Return}_i = -0.611 + 0.957\,\mathrm{Risk}_i$$

Question 2

2. Interpret the coefficient on X and discuss what it means in explaining the annual return. Does this agree with the finance theory?

Question 2

2. Interpret the coefficient on X and discuss what it means in explaining the annual return. Does this agree with the finance theory?

- The coefficient on Risk is 0.957.
- A positive coefficient indicates that there is a positive relationship between the risk and returns.
- This estimate suggests that every one per cent increase in risk is associated with an increase of return by 0.957 percent on average.

Question 2

$$\widehat{Return}_i = -0.611 + 0.957\,\text{Risk}_i$$

The coefficient on Risk is 0.957 .

Do these results align with the finance theory?

Question 2

$$\widehat{Return}_i = -0.611 + 0.957\,\text{Risk}_i$$

The coefficient on Risk is 0.957 .

Do these results align with the finance theory?

Yes, there is a positive relationship between risk and the returns on an investment.

Question 2

3. Comment on the coefficients of determination.

The coefficient of determination is 0.275.

What does this mean?

Question 2

3. Comment on the coefficients of determination.

The coefficient of determination is 0.275.

What does this mean?

- The coefficient of determination is 0.275. This means that 27.5% of the variation in the annual return is explained by the variation in risk.
- While this shows some explanatory power, it also means that 72.5% of the variation is unexplained by the model.
- This suggests that other factors besides risk could play a significant role in determining the annual return, which is common in financial data where many factors influence returns.

Any final questions?

Thanks for your attention! 😄

Let me know if you have any questions.