

數量方法（一）

林茂廷老師

2018-04-15

表格

插图

課程大綱

電子書網址: https://bookdown.org/tpemartin/econometric_analysis/

基本資訊

授課老師: 林茂廷

辦公室: 社科大樓 3F01

諮詢時間: 週二下午 2 點 30-4 點 30

電話: 02 86741111 轉 67170

Email: mtlin@gm.ntpu.edu.tw

作業上傳網址: <https://www.dropbox.com/request/wAzlAatjCjvvExQZlzje>

課程討論室: <https://sites.google.com/view/econometric-analysis/home>

分組名單填寫網址: <https://goo.gl/forms/zXymIEOQbgagwRkg2> 分組

結果: 請點此查閱自己的“組號”

課堂程式說明: [Github repo: Econometrics-in-class-practice](#)

主題內容

In this semester, we will focus on Part II and Part VI of Greene's book. The topics that I will cover are:

- [1] The Generalized Regression Model (ch.8)
- [2] Models for Panel Data (ch.9)
- [3] Models for Discrete Choice (ch.23)

[4] Truncation, Censoring, and Sample Selection (ch.24)

[5] Instrumental Variable Estimation (ch.12)

參考書籍

- Greene, William, *Econometric Analysis*, 6th ed, 2007, Prentice Hall.
- [R for Data Science](#).
- [Principles of Econometrics with R](#), Constantin Colonescu, bookdown.org, 2016.
- [Efficient R Programming](#), Colin Gillespie and Robin Lovelace, bookdown.org, 2016.
- [Econometrics with R](#), Franz Mohr, personal blog (Awesome!!! I must say.)

Cheatsheets

- [dplyr](#)
- [R markdown](#)
- [ggplot2](#)

本學期上課前請先依序完成以下項目

1. 安裝: R and R Studio Desktop: <https://www.rstudio.com>
2. 註冊 Github.com: <https://www.github.com>
3. 安裝 Github desktop: <https://desktop.github.com/>
4. 連到<https://gitter.im>用你的 Github 帳號登錄。
5. 註冊 hypothes.is: <https://hypothes.is>

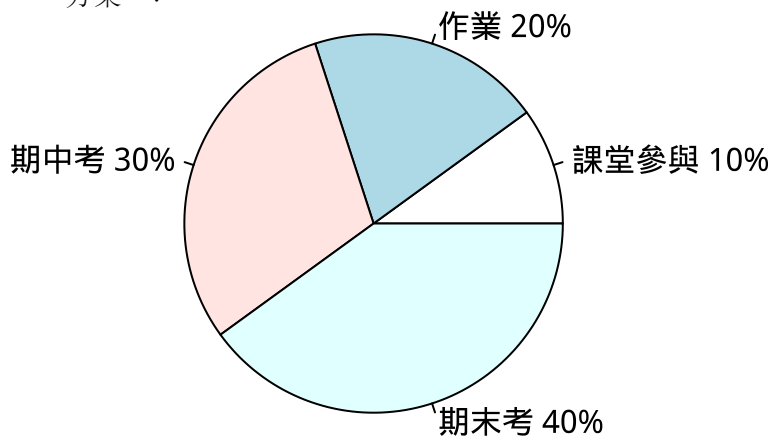
6. 填寫課程資料: <https://goo.gl/hZzv4D>

hypothes.is 電子講義討論群

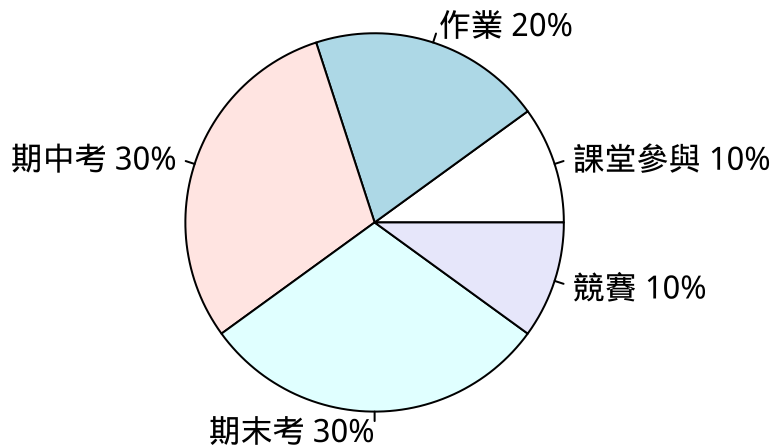
- <https://hypothes.is/groups/xPgdmzMj/shu-fang-106xue-nian>

評分方式

方案一：



方案二：不超過 3 人（限本課程同學）參加[2018 經濟部落客競賽](#)



(PART) Part I: OLS

0.1 OLS

1. “受教育年限越高，是否年薪會越高？”這裡的解釋變數 (X) 與被解釋變數 (Y) 分別是什麼？而 ε 又可能代表那些東西？

0.1.1 因果關連

2. 小明高中畢業，目前年薪 30 萬元。要驗證「若小明有上大學，他可獲得較高的年薪」，請問要做什麼樣的情境比較？

i 代表小明：

被解釋變數	情境
Y_{0i}	只有高中畢業的小明
Y_{1i}	有大學畢業的小明

考慮以下的比較：

$$Y_{0i} - Y_{1i}$$

由於都是小明 (i)，所以上述比較具備有大學文憑帶給小明的薪資效果 (具有因果詮釋力)。

3. 小娟大學畢業，目前年薪 60 萬元。要驗證「若小娟沒有上大學，她可獲得的年薪會變低」，請問要做什麼樣的情境比較？

j 代表小娟：

被解釋變數	情境
Y_{0j}	只有高中畢業的小娟
Y_{1j}	有大學畢業的小娟

在上面的兩個例子，都有一個情境的薪資是觀察不到的。令 Y 代表觀察到的薪資 (可能是大學薪資 Y_1 ，也可能是高中薪資 Y_0 ，視對象實際有無上大學而定)；則

$$Y_k = Y_{0k} + (Y_{1k} - Y_{0k})T_k$$

這裡 k 泛指任何人。上式表示：

每一個有大學文憑的人（即 $T=1$ ），他的實際薪資是他的「高中文憑薪資」再加上「大學學歷效果」所造成。

令 $\delta_k \equiv Y_{1k} - Y_{0k}$ 代表每個人的大學文憑效果。

假設每個人的文憑效果相同 $\delta_k = \delta$ ，則：

$$Y_k = Y_{0k} + \delta T_k$$

這裡的 δ 即為我們所要估計的效應值。

效應值是邏輯論述下的描述，並不是所謂的迴歸模型，也還不涉及任何統計操作。

0.1.2 效應評估

給定效應模型：

$$Y_k = Y_{0k} + \delta T_k (\#eq : effectmodel) \quad (1)$$

我們想進一步評估效應參數 δ 。

- 考慮用 $Y_{1j} - Y_{0i}$ 代表上大學（對上只有高中畢業）給小明和小娟帶來的薪資效果，你需要什麼假設？

Read more...

- 小娟若沒有上大學，她現在的薪資和小明一樣。（ $Y_{0i} = Y_{0j}$ ）

任選兩個人他們的高中文憑薪資要相同的可能性很低，你會怎麼做呢？

Read more...

隨機抽不同學歷的兩「群」人（最好大樣本），比較他們的薪資差異。考慮以下平均薪資差異：

$$\bar{Y}_{college} - \bar{Y}_{highschool}$$

針對 @ref(eq:effectmodel) 式，上式是在估計什麼？

Read more...

$$\mathbb{E}(Y|T = 1) - \mathbb{E}(Y|T = 0)$$

0.1.3 選擇偏誤

$E(Y|T = 1) - E(Y|T = 0)$ 會等於 δ 嗎?

Read more...

$E(Y_0|T = 1) - E(Y_0|T = 0) + \delta$

隨機抽樣下，大學畢業生的平均「高中文憑薪資」與高中畢業生的平均「高中文憑薪資」有很大的機會是不同的。

你可以想到造成不同的理由嗎?

經濟直覺：

立足點高的人（即「高中文憑薪資」高的）比較傾向上大学（即選擇接受 $T = 1$ 的試驗），所以單純比較實際大學平均薪資與實際高中平均薪資，其中的差異除了反應大學文憑效果外，也反應了立足點差異程度。

若個體立足點高低會影響其選擇接受試驗與否，此時造成的效應評估偏誤稱之為「選擇偏誤（selection bias）」。

延伸重點...

要能對因果關係進行合理評估，除了邏輯上兩者要存在可能因果外，還需要：

- 個體會不會被 treated（即最後有沒有上大学， T ），必需與個體還沒決定（或被決定）接受 treated「前」的狀態（即只拿高中文憑的薪資成就， Y_0 ）無關。

另外，

- 隨機抽樣（random sampling）並不一定產生隨機試驗（random assignment of treatments）的效果。

0.1.4 條件式獨立

有些時候「立足點」與「受試選擇」會有關連（以至於產生選擇偏誤），主要是透過某些影響立足點的變數造成。

考慮如下情境：

1. 「高中文憑薪資」與「上大学」的關連主要是透過「家庭所得」：

* 「家庭所得」越高的家庭，小孩就算只有高中文憑，家庭資源也會使得他們的薪資表現不錯。

* 「家庭所得」越高，個人越可能選擇「上大学」。

2. 其他有可能影響「高中文憑薪資」的因素，假設都不會影響個人「上大學」決策。

比較來自相同「家庭所得」的一大群人，若當中有大學文憑的平均薪資與只有高中文憑的平均薪資有差異，這差異會是「高中文憑薪資」立足點差異造成的嗎？

在固定「家庭所得 (HIncome)」條件下，「高中文憑薪資」(立足點)與「上大學」彼此無關連，此稱為條件獨立 (conditional independence)，數學表示為：

$$Y_{0i} \perp T_i | HIncome_i (\#eq : cia) \quad (2)$$

它也意含

$$\mathbb{E}(Y_{0i} | HIncome_i, T_i) = \mathbb{E}(Y_{0i} | HIncome_i)$$

0.1.5 複迴歸模型

比較母體中同家庭所得層次樣本，此時有無「上大學」兩群人的平均薪資差異：

$$\mathbb{E}(Y | T = 1, HIncome) - \mathbb{E}(Y | T = 0, HIncome)$$

會得到大學文憑效應值嗎？

Read more...

$$\begin{aligned} & \mathbb{E}(Y | T = 1, HIncome) - \mathbb{E}(Y | T = 0, HIncome) \\ &= \mathbb{E}(Y_0 | T = 1, HIncome) - \mathbb{E}(Y_0 | T = 0, HIncome) + \delta \quad (\#eq : ch1 - 174) \\ &= \mathbb{E}(Y_0 | HIncome) - \mathbb{E}(Y_0 | HIncome) + \delta \\ &= \delta \end{aligned} \quad (3)$$

隨機變數訊息拆解

任何隨機變數一定可以被拆解成「可被解釋變數捕捉」的部份和「無法由被解釋變數捕捉」的部份，例如：

$$Y = \mathbb{E}(Y | T, HIncome) + (Y - \mathbb{E}(Y | T, HIncome))$$

令 $\varepsilon = Y - \mathbb{E}(Y|T, HIncome)$ ，則：

$$Y = \mathbb{E}(Y|T, HIncome) + \varepsilon.$$

其中

$$\mathbb{E}(Y|T, HIncome) = \mathbb{E}(Y_0|T, HIncome) + \delta T = \mathbb{E}(Y_0|HIncome) + \delta T \quad (4)$$

若 $\mathbb{E}(Y_0|HIncome) = \beta_0 + \beta_1 HIncome$ 為線性關連，則

$$Y = \beta_0 + \beta_1 HIncome + \delta T + \varepsilon \quad (\#eq : ch1 - 190) \quad (5)$$

表示用迴歸模型 @ref(eq:ch1-190) 所得到的母體迴歸線可以用來代表 $\mathbb{E}(Y|T, HIncome)$ 。

幾個重要觀念

1. 迴歸模型可以幫我們得到 $\mathbb{E}(Y|T, HIncome)$ 值。
2. 但 $\mathbb{E}(Y|T, HIncome)$ 是否有助於提煉出所要的效應值得回到效應模型來看。

由本節的問題討論，我們得知：

若「高中文憑薪資高低」（立足點）與「上大學與否」在相同「家庭所得」條件下會獨立，即 @ref(eq:cia) 條件獨立成立，則上述複迴歸式的 OLS 估計式可以得到大學文憑效應的合理估計。

當迴歸分析的旨在衡量「因果效應」時，我們使用複迴歸的目的在於：

- 確保「試驗變數」與「立足點」在相同「控制背景」下，具有《條件性獨立》的特質。