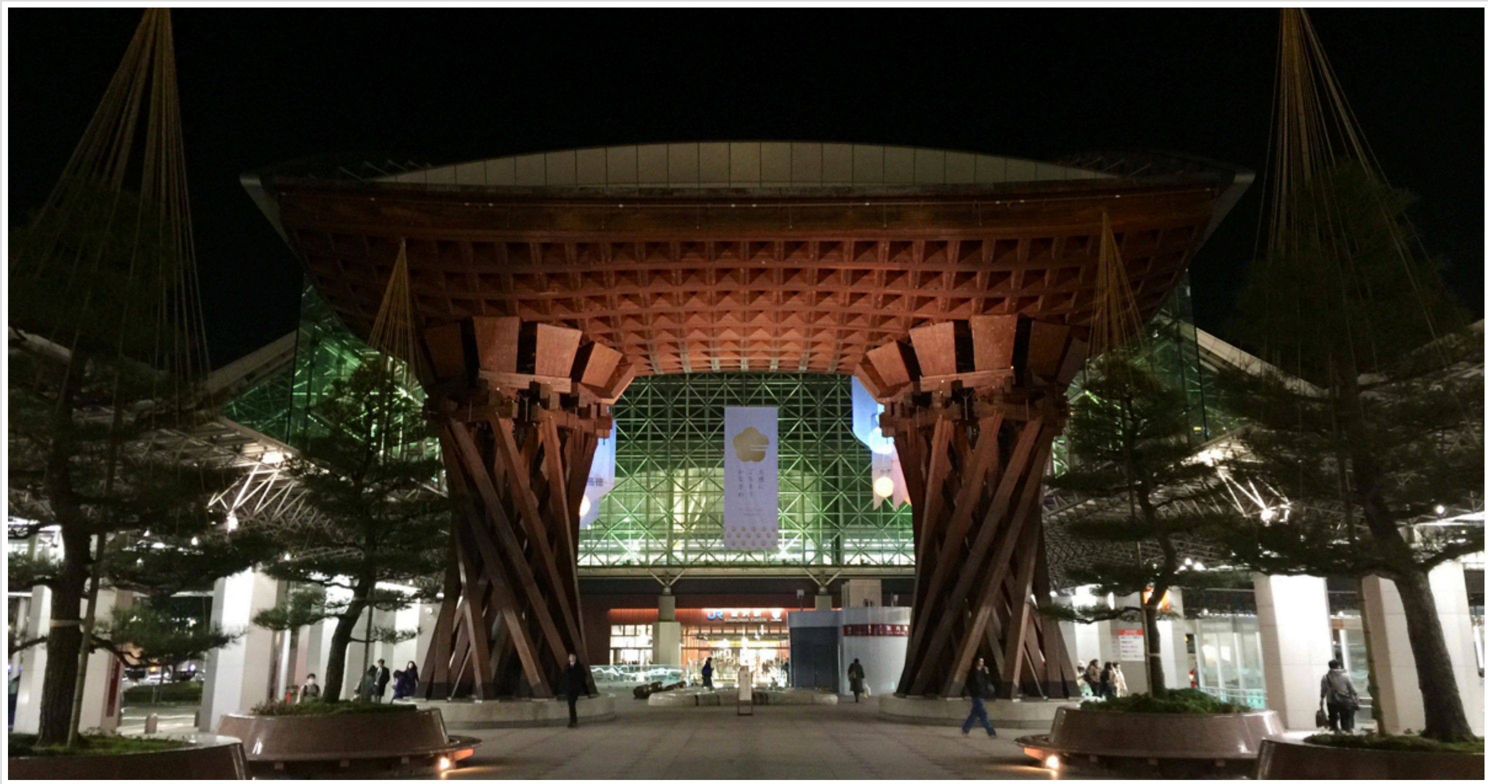
/ Advanced Econometrics I

發表於 2016-12-26 | 分類於 # | 0 Comments | Like 0 Share



找完工作以後，日子依舊忙碌。這個學期 MSBA 有三門課，分別代表 Business Analytics 中三個滿重要的發展方向：分析理論、程式設計、和資料庫。這篇先從偏理論的計量經濟學談起。

秋季學期的安排

和之前上 *Data Visualization* 和 *Statistics* 的夏季學期不同，秋季學期有整整十週（加期末考一週），所以排課就變得滿正常，不過當然也滿考驗時間安排，畢竟不能像夏季學期，可以把全部時間花在單一科目上。這學期 MSBA 學生要修的課有三門：

- *Advanced Econometrics I*（理論、分析）
- *Computational Methods*（實作）
- *Data Management*（資料庫操作、管理）

雖然每門課看起來都滿難的（Computational？ Advanced？！），但其實~~也真的滿難~~認真修完一遍以後，真的能體會到學好這三門課是在 Business Analytics 或 Data Science 領域必備的基礎；而且如果有心繼續鑽研，會發現業界前沿的技術和應用，都和這三個發展方向脫離不了關係，所以不管是入門或進階、研究或應用，似乎都得從這三門課開始。

以 *Advanced Econometrics* 這門課為例，雖然一開始會被各式各樣的公式、證明佐矩陣運算，外加數十頁的作業搞得有點暈頭轉向，但只要跨過計算的門檻後，回頭一看，就比較能理解線性迴歸模型（LRM）中各項假設和定理的關係，也才清楚如何正確計算、解讀各項統計數據。所以儘管讀者可能跟我一樣，曾經感覺自己踏入 Data Science，就是比較想走應用，對理論則敬而遠之；但要能正確使用工具與解讀結果，仍需要紮實的理論知識，這就是 *Advanced Economics* 的教學目的。以下我想盡量簡單介紹我們學了些什麼。

> 註：本文公式較多，建議請用電腦閱讀；如果公式無法正常顯示，請試用其它瀏覽器。

老師和教學方法

這門課的老師是 Dr. Jonathan James，他是個熱情洋溢，也很關心學生的好教授，而且平均一門課會講三四個很棒的笑話，活潑的風格常讓我聯想 BoJack Horseman 裡的 Mr. Peanutbutter。教學方法是版書和投影片，作業形式包括手寫證明題，還有用 R 實作數據分析、並解讀結果。順帶一提，Dr. Jonathan 的所有文件都是用 LaTeX 編排，其精美程度讓我忍不住也跳入學習 LaTeX 的深淵了……。

由於 Dr. Jonathan 已經把上課內容都寫進講義裡了，我們整堂課下來也沒用到教科書，不過在 Syllabus 上他指定的教科書是 William H. Greene 的 *Econometric Analysis*。但因為這本書讀起來實在有點生硬（可能是我個人閱讀能力問題），我在朋友推薦下，找了一本 Jeffrey M. Wooldridge 的 *Introductory Econometrics: A Modern Approach*。後者的優點包括：

- **清晰的架構，深入淺出**：在 Simple Regression Model 中就介紹了許多數值和性質，接下來的 Multiple Regression Model 再談 Asymptotics 等等
- **貼心的附錄，照顧數學弱勢**：包括 Linear Function、Matrix、Probability、Statistics 等基本概念，記憶模糊時（which is always）可以快速回顧

Introductory Econometrics 成功救了我好幾次作業和考試，也是我心目中 Econometrics 領域最棒的教科書之一。

課前準備

當然，從附錄的內容可以看出要學好 Econometrics 需要一點線性代數（linear algebra）和統計（statistics）基礎。粗淺一點來說，需要具備的能力如下：

- **矩陣**：會讀矩陣表達式，熟練加減乘除、轉置、逆矩陣等運算

- **統計**：熟練平均、方差（變異數）、標準差等運算，以及母體和樣本的關係
- **機率**：瞭解條件機率、CDF、PMF / PDF 的意義

以上這些算是最低要求，實際開始學以後，還會碰到許多衍生的觀念、證明和公式，入門先具備清晰的基本觀念就好。有任何不清楚的觀念，可以參考 *Statistics* 中提到的相關資源。既然都學到 Econometrics 了，一定要捨得花時間把所有觀念弄懂，畢竟這會直接影響日後的分析能力。例如，如果不熟矩陣運算，就沒辦法很有效率地思考線性迴歸（Linear Regression）中參數的性質，也就很難理解什麼情況下參數會出現偏誤（bias），以及怎樣避免偏誤。如前所述，要能正確使用各類統計工具，必須先瞭解這些工具背後的理論。

線性迴歸

如果上面的運算都沒什麼問題（或是感覺沒什麼問題想先繼續往下讀），就可以來認識 *Advanced Economics* 中主角中的主角——線性迴歸模型（linear regression model，LRM）了：

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon_i \tag{1}$$

公式中的 i 表示第 i 個樣本（sample observation）；y 為因變量（dependent variable），即模型打算解釋的反應變量（response）；x 為自變量（independent variable），即模型中用來解釋 y 的控制變量（control）。y 和 x 的其它名稱可以參考以下 *Introductory Econometrics* 的整理：

Y 的名稱	X 的名稱
Dependent variable	Independent variable
Explained variable	Explanatory variable
Response variable	Control variable
Predicted variable	Predictor variable
Regressand	Regressor

另外，公式中的 β 為參數（parameter or estimate），表示不同 x 對 y 的影響，其中 β_0 為用來調整大小的截距項（intercept）；最後，ε 是 y 中無法以 x 解釋的殘差項（residual）。如果用一句話簡單說明 y、x、β 和 ε，即**試**

著用 β 組合 x，得出最接近 y、即 ε 最小的預估值。在這裡，**線性**（linear）的意義在於組合的方式為**線性組合**（linear combination），即 β 是線性的，不會出現 $\beta_1 x_1 + \beta_1^2 x_2$ 之類的情況。

由此可見，LRM 是試著用 x 基於 β 的線性組合來預測（或說描述）y。為了和實際的反應變量 y 區別，$(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)$ 這部分又被稱為 $\hat{y}$（讀作 *Y-hat*，可稱作 fitted value），所以上述 (1) 式可以表示為：

$$y_i = \hat{y}_i + \varepsilon_i \tag{2}$$

可以理解為「觀測值＝預估值＋雜訊」。到這裡，我已經用第 i 個樣本（一列）解釋完 LRM 的結構，如果用矩陣記錄 n 個樣本（多列）的運算，就可以推廣成下式：

$$Y = X\beta + \varepsilon \tag{3}$$

- Y 是一個 $(n \times 1)$ 的矩陣，因為有 n 個樣本
- X 是一個 $(n \times k)$ 的矩陣，因為根據 (1) 式，在這 n 個樣本中，我們試著用 k 組自變量來解釋 Y
- 最後，ε 是一個 $(n \times 1)$ 的矩陣

如果將 Y、X 和 ε 化為日常中常見的 Excel 表格，可以這樣表示：

樣本編號	**Y（銷量）**	**X1（價格）**	**X2（廣告）**	**…**	**Xk（人流）**	**e（殘差）**
1	20	15	3	…	180	1
2	23	14	3	…	170	1.5
…	…	…	…	…	…	…
n	21	15	4	…	200	1.3

可以看到樣本有 n 個，自變量有 k 組。另外，請回想一下 (1) 式中 β 的表達形式，由於這個模型中有 k 組自變量，相應地應該要有 k 個 β 作為線性組合的比例，所以 (3) 式中的 β 是一個 $(k \times 1)$ 的矩陣。

利用 (3) 式中的矩陣表達，可以很精簡地將上面表格中的數值，以 Y、X、β 和 ε 來表示，它們之間的關係仍和單一樣本中的 y、x、β 和 ε 一致：**試著用 β 組合 X，得出最接近 Y、即 ε 最小的預估值**。

最小平方法

介紹完 LRM 基本架構後的第一個問題，就是「如何算出 β 和 ε？」關鍵在於前面提到的一句話：**得出最接近 Y、即 ε 最小的預估值**。從上面的公式可以得知 ε 的計算方法為：

$$\varepsilon = Y - \hat{Y} = Y - X\beta \tag{4}$$

考慮到 ε 是一個 $(n \times 1)$ 的矩陣，最小值的目標自然是 $\sum_{i=1}^{n} \varepsilon_i$；不過，為了避免 ε_i 正負抵銷，我們實際要計算的是 $\sum_{i=1}^{n} \varepsilon_i^2$。用矩陣來表示，就是 $\varepsilon'\varepsilon$，解釋如下：

$$\varepsilon' = [\varepsilon_1, \varepsilon_2 \ldots, \varepsilon_n], \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \varepsilon'\varepsilon = (\varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2) = \sum_{i=1}^{n} \varepsilon_i^2$$

所以找到**使 $\sum_{i=1}^{n} \varepsilon_i^2$ 達到最小值的 β 的方法**就是最小平方法（ordinary least squares，OLS）。

母體和樣本

在談如何計算 β 和 ε 前，我想先來小小澄清一下（？）母體和樣本的關係。雖然說我們想預測的是 β 和 ε，但隨著資料的不同，所得到的 β 和 ε 也會不同。比方說，在前面的表格中：

- 用從第一個到第 n 個樣本所得到的 β_n 和 ε_n
- 用從第一個到第 $(n/2)$ 個樣本所得到的 $\beta_{n/2}$ 和 $\varepsilon_{n/2}$

兩組數據可能會不一樣，而就算我們選擇採用 β_n 和 ε_n，在我們沒觀測到的地方可能還有：

- 由從第一個到第 $2n$ 個樣本所得到的 β_{2n} 和 ε_{2n}

這是因為我們觀察到的數值，幾乎永遠都是**樣本**（sample）而非象徵全貌的**母體**（population），所以統計值（β_n、ε_n、平均、方差等等）會不斷隨著資料改變而浮動。為了將這些**浮動的數值**和母體中**固定的數值**區分開來，我們會說母體中不變的參數和殘差是 β 和 ε，而我們試著用樣本估計的參數和殘差是 b 和 e。**β 和 ε 是不變的，而 b 和 e 會隨著樣本的變化而改變**。類似的作法也涵蓋了平均和方差：

統計值	**母體（不變）**	**樣本（浮動）**
平均	μ	$\overline{x}$
方差	σ^2	s^2
參數	β	b
殘差	ε	e

所以前面提到的 (3) 式可以再進一步如下表示：

$$Y = X\beta + \varepsilon = Xb + e \tag{5}$$

這邊 b 和 e 的矩陣維度，和 β 和 ε 是相同的。需要注意的是，雖然 ε 和 e 都是殘差（實際上兩者的稱呼會有所區別），但兩者的性質有著微妙的差異：

- ε 是母體中的實際誤差
- e 除了實際誤差外，還包含了受限於樣本數的解釋誤差

比方說，小明是個消費完全理性的人，他平常有記帳的習慣，但有時候記得太快，會按錯個位數，這個誤差就是 ε；小華根據小明的行為，分析他一半的帳目，發現其中有些無法解釋的零頭，這個誤差就是 e。由於樣本只佔母體的一半，e 除了包含小明確實按錯的誤差以外，也包含了樣本資訊不及母體資訊的誤差。這時，根據 e 所計算出的 b 就會和 β 不太一樣，就算拿這個模型去預測小明另一半的帳目，也不會完全準確。

因此從 (5) 式還有上面這個例子可以看出，真正準確的 b，來自一個**非常接近 ε 的** e。除了增加樣本數以外，避免分析過程中可能出現的誤差（bias）也是一個方法，這在後面的假設會提到。

計算結果和延伸

言歸正傳，為了根據樣本求出達成最小 $\sum_{i=1}^{n} e_i^2$ 的 b，可以採取的方法有兩個：

1. Method of Moments：根據 MOM 推導等式
2. Derivative：利用微分後的一階條件得出 b 的最大值

因為計算過程有點複雜，如果都寫在這邊，可能就得轉型成 Math & Note 了，所以上面兩個連結是簡略的證明過程，前面提到的兩本教科書裡也有完整說明。簡而言之，最後可以得出下式：

$$b = (X'X)^{-1} X'Y \tag{6}$$

記得 b 是一個 $(k \times 1)$ 的矩陣。另外 (4) 式中的 $\hat{Y}$ 也可以寫成：

$$\hat{Y} = Xb = X(X'X)^{-1} X'Y \tag{7}$$

雖然這個公式看起來很複雜，但實際推導過一次以後會比較習慣；而有了標準的表達方式以後，也就能推導出 b 的各項性質，也就是在解讀參數時應該注意的細節。例如：

1. **和母體的關係**： $b = (X'X)^{-1} X'Y = (X'X)^{-1} X'(X\beta + \varepsilon) = \beta + (X'X)^{-1} X'\varepsilon$

   由此可看出 b 中包含 ε，以及樣本 X 如何左右 ε 的影響程度。

2. **方差分析**： $SST = \sum_{i=1}^{n} (y_i - \overline{y})^2 = SSR_{egression} + SSE_{rror} = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{n} e_i^2$

當 LRM 的擬合度越好，$SSE=\sum_{i=1}^{n} e_i^2$ 越小，解釋程度 $R^2=\dfrac{SSR}{SST}$ 越高。

3. **期望值**：$E\left(b|X\right)=\beta+\left(X'X\right)^{-1}X'E\left(\varepsilon|X\right)=\beta$

   隨著樣本數增加，b 將逼近 β 的漸進（asymptotic）和一致（consistency）性質。

4. **方差（和協方差矩陣）**：$\widehat{Var\left(b|X\right)}=s^2\left(X'X\right)^{-1},\quad s^2=\dfrac{e'e}{n-k}$

   利用總體方差計算 b 的標準差，確認預估的精準度和有效性（efficiency）。

5. **分布函數**：$b|X\sim N\left(\beta,s^2\left(X'X\right)^{-1}\right),\quad b_k|X\sim N\Big(\beta,s^2\left(X'X\right)^{-1}\Big)_{(k\times k)}$

   對 b_k 進行假設檢驗（hypothesis test）、構建信賴區間（confidence interval）等。

對於讀到這邊感到一頭霧水的讀者，我想說（一）不用擔心，到目前為止提到的內容，都是 `R` 等統計軟體中內建好的功能，所以就算不懂怎麼用矩陣算 b 也沒關係，分析結果中已經包含所有計算；但（二）這些內容有點類似**判斷模型好壞**的基礎，後面會提到這些計算、證明所衍生的性質，它們會直接影響分析過程中選擇什麼工具、如何處理資料和解讀結果。

以 `R` 來比喻的話，（一）的情況如下：

```
Call:
lm(formula = mpg ~ cyl + disp + hp + drat + wt, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-3.7014 -1.6850 -0.4226  1.1681  5.7263

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
                  #1         #4      #5       #5
(Intercept) 36.00836    7.57144   4.756  6.4e-05 ***
cyl         -1.10749    0.71588  -1.547  0.13394
disp         0.01236    0.01190   1.039  0.30845
hp          -0.02402    0.01328  -1.809  0.08208 .
drat         0.95221    1.39085   0.685  0.49964
wt          -3.67329    1.05900  -3.469  0.00184 **
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.538 on 26 degrees of freedom
                              #2
Multiple R-squared:  0.8513,    Adjusted R-squared:  0.8227
                     #2                                #2
F-statistic: 29.77 on 5 and 26 DF,  p-value: 5.618e-10 #5
             #5                                #5
```

> 註：上述註記中， `#1` 為由第一個性質（和母體的關係）推導而得，以此類推。

（二）所涵蓋的問題則包括：

- 一開始 `lm()` 中的資料選取是否妥當？
- `Estimate` 、 `Std. Error` 是否有偏差？如何解決？
- 如何正確解讀 `(Adjusted) R-squared` 、 `F-statistic` 和 `t value`？
- 分析結果是否有改進空間？如何取捨 bias 和 variance？

所以，這邊的分水嶺大概是「我知道怎麼操作 `lm()` 、 `glm()` 甚至 `rpart()` 等函數，但我不清楚這些統計結果是否正確，也不知道應該按什麼步驟處理資料、或微調（tweak）函數來改善預測結果」。*Econometrics* 可以幫助你瞭解背後的原理，但要學好 *Econometrics*，還是得從上面這些內容開始。

假設與性質

學好上面這些基礎的重要理由之一，在於弄清楚 LRM 內部是怎麼一回事，還有這些分析是建立在哪些假設上。說到這裡，我不得不說 Dr. ~~Peanutbutter~~ Jonathan 把課程規劃得滿好。他先介紹了 LRM 的五個很重要的假設，它們分別是：

1. **Linearity**：b 的關係一定要是線性（X 則不在此限）
2. **Full Rank**：X 矩陣必須為滿秩，可以理解為不能有兩組以上完全相關的 X
3. **Mean Independence**：$E(\varepsilon|X)=0$，瞭解 X 無助於瞭解 ε
4. **Homoskedasticity & Non-autocorrelation**：$Var(\varepsilon|X)=\sigma^2$ 和 $Cov(\varepsilon_i,\varepsilon_j|X)=0, \forall i \neq j$，即 ε 的分布獨立於 X，且 ε 之間不相關
5. **Normality**：$\varepsilon \sim N\left(0,\sigma^2\right)$，$\varepsilon$ 的分布符合正態分布 N

> 註：這邊提到的五個假設，和 Wikipedia - Linear regression 與 *Introductory Econometrics* 上所提到的有些微差異，但在原理上應該是相通的，端看講者側重說明 LRM 的哪些面向。

介紹完這五個基本假設後，我們就開始探討「如果實際資料不（完全）符合這些假設，該如何避免潛在的偏差」。於是，基於這五項假設，我們另外學了以下內容，它們的證明和推導，都是來自上面所提到的計算結果和延伸。

假設	**期中考前的延伸**	**期末考前的延伸**
Linearity	利用多項式（polynomial）改善 R^2	增設交互項（interaction term）以解讀 X 之間的交互關係
Full Rank	留意虛擬變量（dummy variable）的設定	當 $k > n$時，可採取 Ridge 或 Lasso 等 Penalized Regression
Mean Independence	檢討 b 在不同樣本情況下可能出現的偏差（bias），並判斷偏差方向	利用工具變量（instrumental variable，IV）消弭 b 偏差
Homoske & Non-Auto	瞭解 OLS 估計的 *MVLUE* 性質，以及影響 $Var(b_k)$ 的三個因素	利用多種方法（註）修正 $\widehat{SE}(b_k)$ 偏差，和用 Bootstrapping 取樣
Normality	對 b_k 或 $\hat{y}_0$ 進行假設檢驗、計算信賴區間、t 值、F 值和 p 值	基於 MLE 構建 $L(\theta)$，並用 Newton-Raphson 等方法估計 b

> 註：針對 Heteroskedasticity 的解決方案，包括 Weighted Least Squares（WLS）、Huber-White Standard Errors、Breusch-Pagan Test、 ~~DoubleJ Test~~ 和 Clustered Standard Error 等。

解決潛在問題

讀者如果沒學過這些性質，可能會覺得一頭霧水，所以我想沿用前面的例子，說明瞭解這些性質有多重要，如果讀者有一些數據分析經驗，應該能比較好理解。

首先，前面我用來說明 (3) 式表達法 $Y = X\beta + \varepsilon$ 的例子如下：

樣本編號	Y（銷量）	X1（價格）	X2（廣告）	…	Xk（人流）	e（殘差）
1	20	15	3	…	180	1
2	23	14	3	…	170	1.5
…	…	…	…	…	…	…
n	21	15	4	…	200	1.3

Mean Independence

乍看之下這份資料很正常，但如果我們試著質疑它，就會引出一些潛在的問題。例如，如果我將第 k 項資料（人

流）删除：

樣本編號	Y（銷量）	X1（價格）	X2（廣告）	…	Xk-1（車流）	e（殘差）
1	20	15	3	…	40	2
2	23	14	3	…	50	2.5
…	…	…	…	…	…	…
n	21	15	4	…	45	0.3

表格一：少了第 k 項資料（人流）

請問：

- 跟原始資料相比，除了少了最後一項以外，b（即分析結果中的 `Estimate`）會出現怎樣的變化？
- 如果「人流」確實和「銷量」有關，這樣的變化是偏差還是修正？

如果將和「銷量」幾乎毫不相關的「極光強度」加入模型中：

樣本編號	Y（銷量）	X1（價格）	X2（廣告）	…	Xk（極光強度）	e（殘差）
1	20	15	3	…	100%	0.9
2	23	14	3	…	70%	1.7
…	…	…	…	…	…	…
n	21	15	4	…	15%	3.2

表格二：將第 k 項資料換成「極光強度」

請問：

- 跟表格一相比，除了多最後一項以外，b 會出現怎樣的變化？

如果我們知道「體驗」，即顧客對消費過程中的感受，會對銷量會造成影響，但實際上「體驗」卻難以量化：

樣本編號	Y（銷量）	X1（價格）	X2（廣告）	…	Xk（體驗）	e（殘差）
1	20	15	3	…	?	2
2	23	14	3	…	?	2.5
…	…	…	…	…	…	…
n	21	15	4	…	?	0.3

表格三：將第 k 項資料換成「體驗」

> 註：因「體驗」無法量化，此處殘差沿用自表格一。

請問：

- 少了「體驗」這項參數的 b，可能存在的偏差為何？

以上三個狀況，都是在 Mean Independence 這項假設背後可能存在的問題，而且延續母體和樣本的思考，就連**最初的表格**也值得我們思考「b 是否存在偏差」以及「這樣的資料選取是否正確」。由於算出正確的 `Estimate` 幾乎是 LRM 基本中的基本，從性質和假設出發，培養對偏差的判斷能力非常重要。

Homoskedasticity & Non-autocorrelation

除了 Mean Independence 以外，假設四是否成立會影響分析結果中的 `Std. Error`，即 `Estimate` 的分布。由於 Homoskedasticity（*homo* 單一，*skedasticity* 方差性）假設殘差 ε 的方差（variance）不隨 X 變化，且 Non-autocorrelation 假設各 X 間的 ε 不相關，因此，理論上的方差矩陣 $Var\left(b|X\right)$ 長得像這樣：

$$Var\left(b|X\right)=\sigma^2(X'X)^{-1}=\begin{bmatrix} Var(\beta_1) & 0 & \dots & 0 \\ 0 & Var(\beta_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Var(\beta_k) \end{bmatrix}$$

將矩陣對角線上所有 $Var(\beta_k)$ 開根號，就會得到分析結果中的 `Std. Error` 。然而，實際上我們預估的 $\widehat{Var\left(b|X\right)}$ 卻可能面臨（一）方差隨 X 變化，且（二）ε 彼此相關的狀況：

$$\widehat{Var\left(b|X\right)}=\Sigma=\begin{bmatrix} Var(b_1) & Cov(\varepsilon_2,\varepsilon_1) & \dots & Cov(\varepsilon_k,\varepsilon_1) \\ Cov(\varepsilon_1,\varepsilon_2) & Var(b_2) & \dots & Cov(\varepsilon_k,\varepsilon_2) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\varepsilon_1,\varepsilon_k) & Cov(\varepsilon_2,\varepsilon_k) & \dots & Var(b_k) \end{bmatrix}$$

其中 Σ 無法以 $s^2(X'X)^{-1}$ 形式表達。

什麼情況下，分析會違反這兩個假設呢？首先是資料本身的性質。比方說在上面的原始資料裡，「銷量」和「價格」之間的關係就有可能違反 Homoskedasticity：便宜商品的銷量分布（很多到很少），和昂貴商品的銷量分布（少到很少）不同，因此分析結果中的 `Std. Error` 就可能出現偏差，後續的分析，如 `t value` 、 `Pr(>|t|)` 、 `F-statistic` 等等也就不可靠。（不過倒是不會影響回 `Estimate` ，因為 b 偏差與否和 Mean Independence 有關。）

另一種情況，則和資料的分群（cluster）有關。例如，如果原始資料和「地區」有關：

樣本來源	Y（銷量）	X1（價格）	X2（廣告）	…	Xk（人流）	e（殘差）
台北	20	15	3	…	180	1
台北	23	14	3	…	170	1.5
…	…	…	…	…	…	…
台中	21	15	4	…	200	1.3

表格四：將「樣本編號」換成「樣本來源」

這時可以看出因為某些樣本來自相同地區，殘差也較有可能彼此相關，而出現某種特徵（pattern），如果用一般的方式計算方差會得到錯誤的 `Std. Error` 。對以上兩個問題的處理方法有興趣的讀者，可以參考這份 University of Maryland 的 *Introduction to Robust and Clustered Standard Errors (PDF)*； `R` 的處理方法可以參考這篇 *Easy Clustered Standard Errors in R*。

Normality

最後的 Normality 雖然看起來很簡單：就算是在漸進（asymptotic，即樣本數足夠多）的情況下，只要接受這項假設，就能利用不同的分布計算信心區間等；但另一方面，一旦接受了這個假設，我們也能用之前提過的最大似然估計法（MLE）估計 b。簡略的步驟如下：

1. 利用假設的分布密度函數（pdf）得出特定參數 θ（包含 b 和 s^2）的機率函數： $p\left(y|X=x;\theta\right)$
2. 連乘機率函數，得 likelihood function： $L(\theta)=\prod_{i=1}^{n}p\left(y_i|x_i;\theta\right)$
3. 對似然函數取自然對數（註），得 log-likelihood function： $logL(\theta)=\sum_{i=1}^{n}logp\left(y_i|x_i;\theta\right)$
4. 在 $logL(\hat{\theta})$ 最大的情況下， $\hat{\theta}$ 即估計結果。

> 註：這個步驟的理由通常是連加 $\sum$ 比連乘 $\prod$ 更便於計算，詳細的說明可以參考這份 CMU 的 *The Method of Maximum Likelihood for Simple Linear Regression (PDF)*。

最後的 $\hat{\theta}$ 會包含 b 和 s^2，理論上和 OLS 所得出的結果會一樣。雖然在單一模型的情況下，用以上四個步驟代替 `lm()` 有點多此一舉，但在聯立模型（simultaneous equations models）的情況下，需要用 MLE 方法才能考慮到殘差之間的相關性。

一樣用前面的資料說明，不過這次除了「銷量」以外，我們還得考慮另一個「利潤」的表格：

樣本來源	Y（利潤）	X1（價格）	X2（廣告）	…	Xk（人流）	e（殘差）
台北	1.8	15	3	…	180	1.1
台北	2.5	14	3	…	170	0.9
…	…	…	…	…	…	…
台中	1.7	15	4	…	200	1.4

表格五：將「銷量」換成「利潤」

這時，如果我們想估計「利潤」、「銷量」這兩個 Y 和 X 的關係，不能只對這兩個表格分別用 `lm()`，因為兩組殘差 e 之間的關係是不確定的；如果我們也要為兩組 e 建立一個類似 $Var\left(b|X\right)$ 的 covariance matrix，就得用 MLE 一次估計兩組模型，這時的 θ 會包含兩組 b、兩個 s^2、和一個 e 間的相關係數 ρ（註）。

> 註：為了確保估計出來的 s^2 恆正，這邊其實要用到一些特殊方法；一般來說也可以用 e^x。

本質上，MLE 估計方法有點像在玩拉霸：不斷投入一組 θ，看得出來的 $logL(\theta)$ 多大，直到找到最大的 $logL(\hat{\theta})$。這也的確就是網格式參數搜尋（grid-search）的原理，不過我們也可以用一階微分、Newton-Raphson 等方法求出最大值；在 `R` 裡面，有 `optim()`、`nlminb()` 等函數可以用來求極值，不過要注意的是大部分的函數**只支援求最小值**，所以可能需要先將 log-likelihood function 加上負號，求 $-logL(\theta)$ 的最小值。

MLE 方法裡最核心也最困難的步驟，應該是在前幾步將模型化為 likelihood function，不過，也可以說只要能寫出 likelihood function，MLE 方法可以用來預估各式各樣的模型。所以不只是一般的 LRM，效用函數（utility function）、強化學習（reinforcement learning）等問題也能用 MLE 方法解決。比較可惜的是我們這學期幾乎沒什麼機會自己建構 likelihood function……或許得涉獵一些期刊才能見識 MLE 方法應用之廣。~~然而為了寫這篇文章我最近是沒什麼空。~~

延伸學習

學完這麼多東西以後，我的直接感受是用 `lm()` 、 `glm()` 等函數要注意的事情真多，不能像以前一樣，一看到資料，以為排個 $Y \sim X_1 + X_2 + ... + X_k$ 下去分析就好；解讀結果的時候，除了瞭解 causal effect，也要注意是否有偏差。當然，如果要精進分析結果，也不能只是盲目嘗試，唯有清楚原理中影響結果的因素，才有改善的空間。~~總之就是面倒くさいなあ。~~

具體來說，我覺得學完這門課程之後可以精進的方向有三：

1. **熟練前面的內容**
   從課前準備開始提到的矩陣運算、統計和機率，到最後的 MLE 都很重要，但老實說我學了一遍、外加參考不少資料，也不太確定自己到底掌握了幾分。考慮到 *Econometrics* 還是偏基礎的科目，想接觸其它分支前，還是得先把上面提到的內容學好。

2. **處理更多樣的資料**
   前面舉利用的幾個表格是橫截面數據（cross-sectional data），即在特定時間點下的數據，但 *Econometrics* 會面對的資料型態，還包括考慮到時間變化的的時間序列（time series data）和縱橫 / 面板數據（panel data）。*Introductory Econometrics* 的中後部分就是在講解後兩者，相信這也是我們下學期 *Advanced Econometrics II* 的重心。

3. **使用更多分析方法**
   除了 `lm()` 裡包含的基本統計數據以外，依照不同的數據型態，還有不同的 bias 和 variance 取捨所衍生出的分析方法也不同；前者包括 ANOVA、時間序列分析等等，後者包括 Ridge、Lasso 等等，也和後續的降維（dimension reduction）、非線性方法有關，之前推薦過的 *An Introduction to Statistical Learning* 裡有詳細說明這些進階方法；姊妹書 *The Elements of Statistical Learning* 更包含背後的數理推導。

結語和勘誤表

落落長的文章到此也差不多該告一個段落。雖然我剛開始寫 McK & Note 的時候，從沒想過自己會寫這樣一篇充滿數學公式的文章，不過在蒐集資料的過程中，我也意識到網路上把這些假設統整起來、寫清楚的文章比較少，所以決定把自己還記得的上課內容都寫一寫，也當作是把 *Econometrics* 複習了一遍。希望這篇文章能為中文讀者填補一些資訊落差，也希望我的講解能讓讀者對 *Econometrics* 產生一些興趣，至少別把它當成艱澀、難懂的知識。

不過，我也清楚這篇文章是個大工程，充滿了公式和標注，即使寫完了也不敢說完全沒有錯誤，所以我在這篇文章最後預留了勘誤表。我還有一段時間要跟 *Econometrics* 打交道，想必偶爾還是會回來讀一讀這篇文章，如果發現有誤，會將修正記錄在這裡；如果讀者發現有任何不清楚的地方，也請不吝指正。

日期	位置	修改前	修改後	原因
‘16 Dec 28	五個表格	ε（殘差）	e（殘差）	樣本估計應為 e

攝影地點：JR 金沢駅、飛驒高山、穴水町、JR 新宿駅

Cal Poly # MSBA # Econometrics

/ (How) I got an offer

0 Comments　McK & Note　1 Login

Recommend　Share　從最好的優先排列

Start the discussion…

Be the first to comment.

ALSO ON **MCK & NOTE**

讀 / 數位全球化

1 comment • 9個月前•

Jimmy Lin — 補充：關於無國界世界的哲學討論，可以參考哈佛教授 Michael Sandel 在 BBC Radio 4 …

讀 / BCG《你的策略，需要策略》

2 comments • 6個月前•

Jimmy Lin — 向賴董看齊 <(_ _)>

/ CFA IRC 準備過程

4 comments • 8個月前•

錢昭豪 — Hi 我是今年政大隊的成員！我們是不是有在芝加哥碰過面呢？我在找BCG的書找到這個blog :) All the best.

/ CFA IRC 東南亞神秘力量

2 comments • 8個月前•

Jimmy Lin — 真假，我來看看！

Subscribe　加入 Disqus 到你的網站 Add Disqus Add　隱私