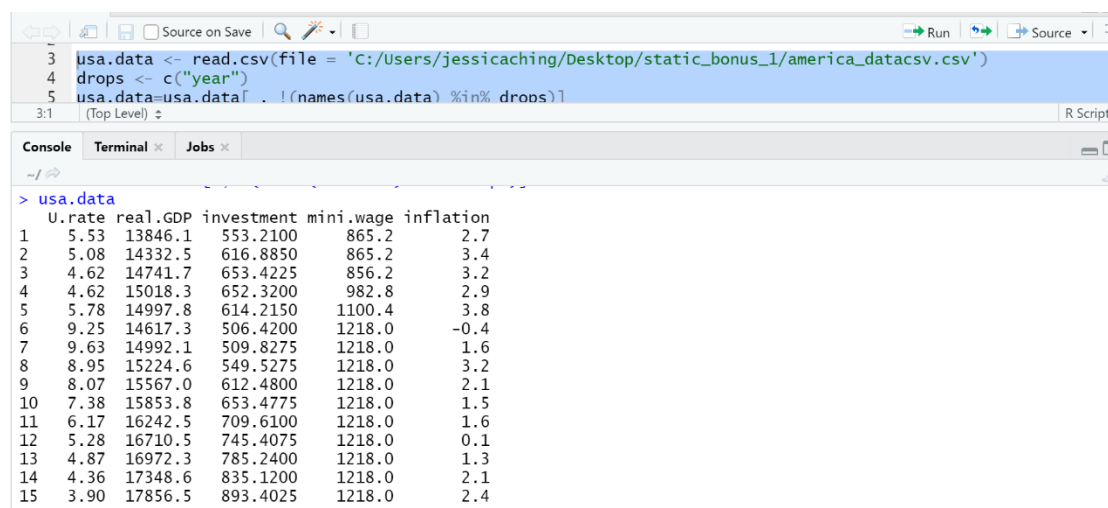


統計二_R 作業

106191127 張伊晴

對於這次的作業，我感興趣的是美國失業率跟其他經濟變因之間的關係。我對模型的規劃是 step 1 的模型想要用 multiple linear regression 估計美國失業率與 real gdp、investment 兩個變因之間的關係。而 step 2 的模型，我想再增加 minimum wage 和 inflation 這兩個變數進去 multiple linear regression，看對於失業率的影響會是什麼。雖然詢問老師後，我設計的這個模型可能會有 unit root (單根) 所造成的 spurious regression，以及 endogeneity 的問題。但我思考後還是想試試看用這個模型做出來的結果會是什麼、是否會因為模型的瑕疵而導致結果跟 business cycle facts 有所出入，而這也是我對這個模型感興趣的原因之一。
<這是我找的資料>



```
3 usa.data <- read.csv(file = 'C:/Users/jessicaching/Desktop/static_bonus_1/america_data.csv')
4 drops <- c("year")
5 usa.data = usa.data[, !(names(usa.data) %in% drops)]

> usa.data
  U.rate real.GDP investment mini.wage inflation
1  5.53  13846.1   553.2100    865.2        2.7
2  5.08  14332.5   616.8850    865.2        3.4
3  4.62  14741.7   653.4225    856.2        3.2
4  4.62  15018.3   652.3200    982.8        2.9
5  5.78  14997.8   614.2150   1100.4        3.8
6  9.25  14617.3   506.4200   1218.0       -0.4
7  9.63  14992.1   509.8275   1218.0        1.6
8  8.95  15224.6   549.5275   1218.0        3.2
9  8.07  15567.0   612.4800   1218.0        2.1
10 7.38  15853.8   653.4775   1218.0        1.5
11 6.17  16242.5   709.6100   1218.0        1.6
12 5.28  16710.5   745.4075   1218.0        0.1
13 4.87  16972.3   785.2400   1218.0        1.3
14 4.36  17348.6   835.1200   1218.0        2.1
15 3.90  17856.5   893.4025   1218.0        2.4
```

首先我先估計第一個 multiple linear regression 的模型。

```
>
> model1 <- lm(U.rate ~ real.GDP + investment, data = usa.data)
> summary(model1)
```

Call:

```
lm(formula = U.rate ~ real.GDP + investment, data = usa.data)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.58970	-0.21438	0.01547	0.27757	0.41494

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-7.7543150	1.9064589	-4.067	0.00156 **
real.GDP	0.0023050	0.0001885	12.228	3.92e-08 ***
investment	-0.0333952	0.0019183	-17.408	7.00e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3744 on 12 degrees of freedom

Multiple R-squared: 0.9674, Adjusted R-squared: 0.962

F-statistic: 178.3 on 2 and 12 DF, p-value: 1.192e-09

雖然我的模型大概只有 Estimate 會是正確無誤的，但我想我還是可以試著解釋看看這個結果。假設模型設定正確下，在控制其他變數的影響後，real GDP 上升 1 個單位，失業率上升 0.002305 個單位；控制其他變數的影響後，investment 上升 1 個單位，失業率下降 0.0333952 個單位。我認為投資跟失業率的反向變動情形是符合 business cycle facts，但這個模型裡似乎沒有反應出 real GDP 跟失業率的常態關係。不過，我覺得它們之間的變動關係十分的小，除了模型本身的因素外，我覺得也有可能是會影響失業率的原因本來就很多，而 real GDP 只是其中之一。

在假定模型正確下，這個模型 F 統計量所對應的 p-value 都非常的小，因此我們可以拒絕 $H_0: b_1=b_2=b_3=0$ 的虛無假設。而模型中的 R-squared 某些方面衡量了這個模型配飾的好不好，如果 R-squared 越高，代表給定一樣的 TSS，它的 RSS 越好、ESS 越低。R-squared 越高，模型可以解釋得越多，被解釋的變數 y 的變動中，有越多的比例可以被解釋變數 x 所解釋。而這個模型的 R-squared 很高，有將近 97% 可以被解釋變數所解釋。不過我們可以看到 Adjusted R-squared 適度地調降了 R-squared 的功用，這是因為 Adjusted R-squared 裡面加了懲罰項，可以幫助我們修正 R-squared 的功能。

再來，是我第二的模型，我在原本的模型中多加了幾個解釋變數。

```
Source
Console Terminal x Jobs x
~/
> summary(model2)

Call:
lm(formula = U.rate ~ real.GDP + investment + mini.wage + inflation,
    data = usa.data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.57373 -0.20979  0.01335  0.28502  0.40702

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.0694303   6.6581824   -1.062   0.31331
real.GDP      0.0021894   0.0010321    2.121   0.05990 .
investment   -0.0324944   0.0082024   -3.962   0.00268 **
mini.wage     0.0004582   0.0038499    0.119   0.90763
inflation     0.0055069   0.1143391    0.048   0.96254
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4099 on 10 degrees of freedom
Multiple R-squared:  0.9675,    Adjusted R-squared:  0.9545
F-statistic: 74.39 on 4 and 10 DF, p-value: 2.122e-07
```

跟原本的模型相比，我發現 residuals 的 median、max、min 的值都很像，可能表示兩個模型的誤差原因、所造成的因素都很相近。而多加的兩個變數都跟失業率成同向關係，像是在控制其他變數不變的影響後，Mini wage 上升 1 個單位，失業率上升 0.0004582 個單位。不過我發現最低薪資的 estimate 超級小，我認為這可能是因為政府不會常常改變法定最低薪資，但失業率是會因種種經濟

環境因素而時常變化的，所以最低薪資的 estimate 才會跟其他解釋變數比起來，它的值小非常的多。

而當我們再考慮兩個模型，其中一個變數比另一個多時，比較 R-squared 就沒意義了。所以可以用比較 adjusted R-squared 來看兩個模型。模型 2 的 R-squared 雖然比模型 1 高，但模型 2 的 adjusted R-squared 卻是比模型 1 還要小。所以可以看出模型 2 的變數雖然比模型 1 多，但在估計和解釋上卻不如模型 1 來的好。由此可知，model2 的 R-squared 增加，但不見得 model2 就比較好。

此外，我覺得模型跑出來的結果跟 business cycle facts 還是有些落差存在的。像是變數中只有投資跟失業率呈現反向關係，real GDP 和 inflation 居然都和失業率成同向變動，雖然影響的數字小，但還是有點不符合現實。

作業反思：

這次的作業使我試著用已學的統計能力，解釋在經濟學中所學到的知識和經濟現象。雖然模型有誤差和瑕疵存在，但我覺得這次作業是滿有趣的，因為我終於不是在只有學習知識，而是試著運用知識來處理想知道的問題。

從 R 程式跑的結果看來，失業率和這些變數的關係不太顯著，但也有可能是因為這個實證問題，本來就不太適合用 lm() 的模型做。不過失業率跟有些經濟變因的變動方向，還是可以看出來的，如 Investment 上升，失業率下降，兩者呈現反向關係。所以我想之後如果還要繼續做這個實證問題的研究，應該要嘗試看看新的模型或方法，以符合總體經濟變數的性質，這些變數之間所呈現出的相關性可能也會比較正確。

參考資料：

https://www.theglobaleconomy.com/USA/GDP_constant_dollars/

<https://datatofish.com/multiple-linear-regression-in-r/>