

# 統計學 Python 期末報告

資管系 B11705016 黃子維 B11705029 劉芊儀 B10106045 郭菲霓

## 一、研究目的

研究目的在於探討營業毛利率、營業利益率、年度以及每股盈餘這四個財務變數之間的關聯性。本研究將透過回歸分析的方法，分析這些變數在不同年度的財報資料中如何相互影響。特別是，研究將著重於營業毛利率、營業利益率與每股盈餘之間的關係。此外，本研究也將納入時間因素（年度）進行分析這些財務指標的表現。透過對這些關鍵財務指標的深入了解，本研究旨在讓我們了解不同財務數據間的數據間的關係，並且加以探討模型選擇的好壞。

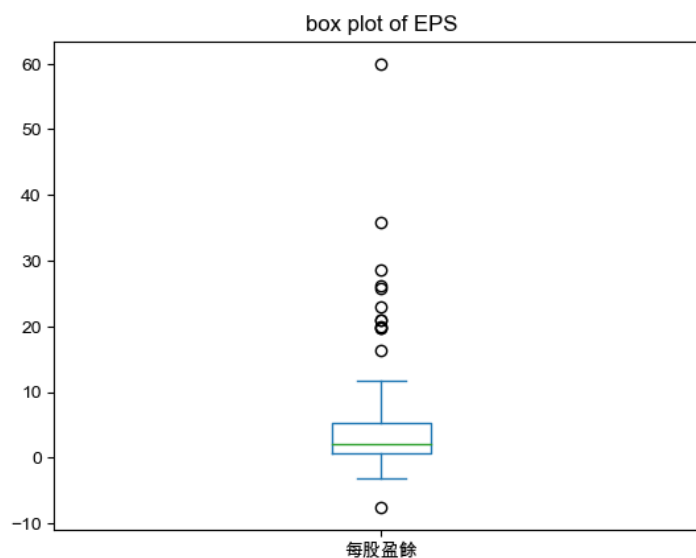
## 二、敘述統計

### 1. 三項目之基本敘述統計：

|           | 營業毛利率     | 營業利益率       | 每股盈餘      |
|-----------|-----------|-------------|-----------|
| 平均數 mean  | 13.203526 | 0.922158    | 4.196000  |
| 標準差 std   | 8.411150  | 6.624067    | 7.196364  |
| 最小值 min   | 3.040000  | - 28.980000 | -7.540000 |
| 第一四分位數 Q1 | 5.915000  | 0.320000    | 0.562500  |
| 第二四分位數 Q2 | 11.640000 | 1.570000    | 2.080000  |
| 第三四分位數 Q3 | 17.645000 | 3.310000    | 5.182500  |
| 最大值 max   | 40.270000 | 15.770000   | 59.980000 |
| 眾數 mode   | 4.800000  | 3.310000    | 1.570000  |

|              |           |           |           |
|--------------|-----------|-----------|-----------|
| 變異數 variance | 70.747452 | 43.878270 | 51.787654 |
| 偏度 skewness  | 1.074630  | -2.264143 | 3.792000  |
| 峰度 kurtosis  | 0.820529  | 8.022236  | 21.483006 |

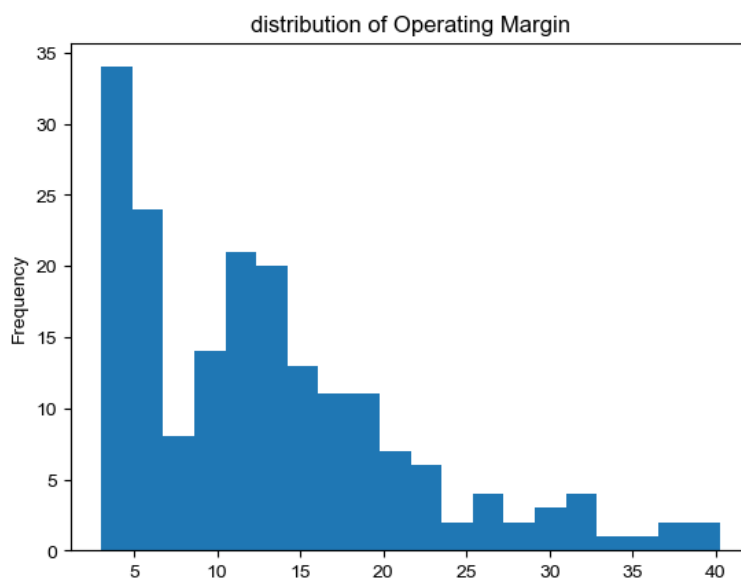
## 2. 每股盈餘盒狀圖：



圖一-每股盈餘和盒狀圖：變異性極大，代表有些年份或公司表現極度優秀

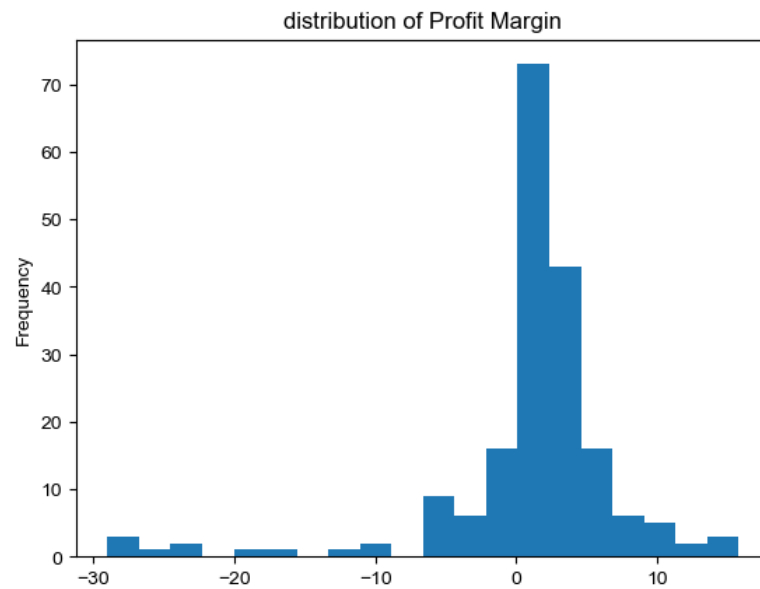
## 3. 三項目之分布情況：

### a. 營業毛利率分布圖



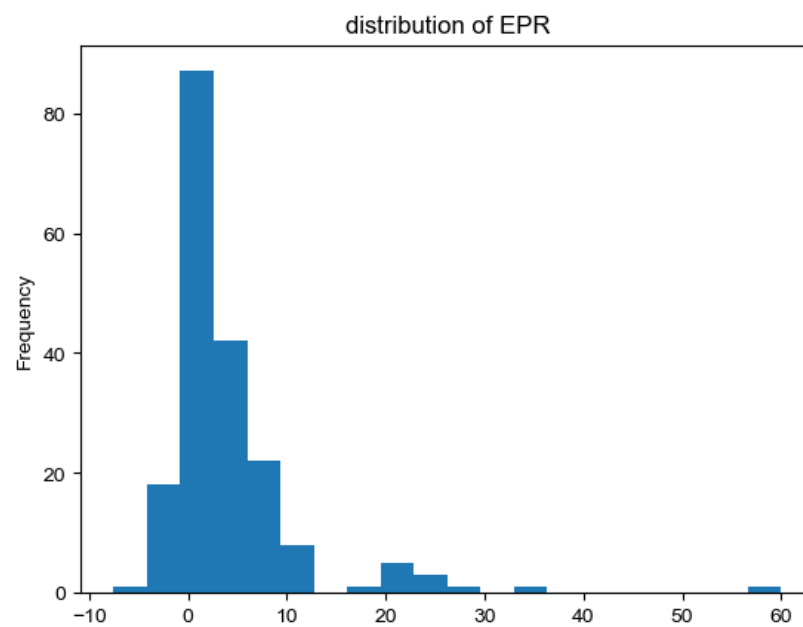
圖二-營業毛利率分布圖：數據成右偏分布，大部分公司營業毛利率低

b. 營業利益率分布



圖三-營業利益率分布圖：左偏分布，大部分公司營業利益率高

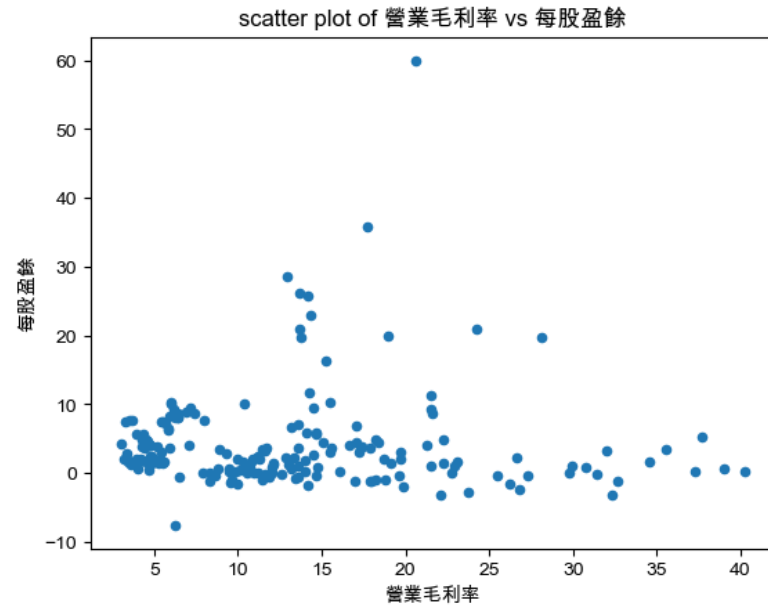
c. 每股盈餘分布



圖四-每股盈餘分不布圖：右偏分布，大部分公司每股盈餘在同一個區間內

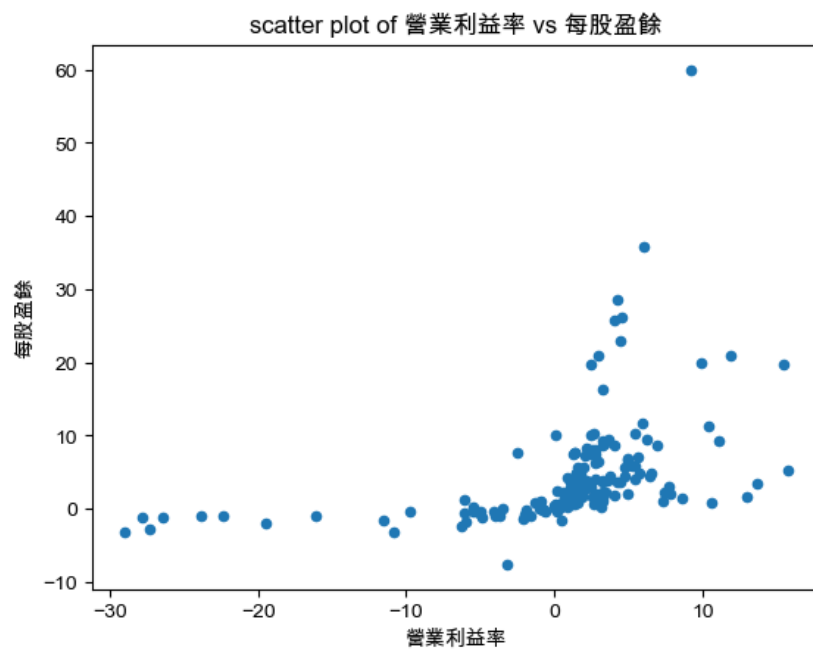
4. 散佈圖：

a. 營業毛利率與每股盈餘之散佈圖



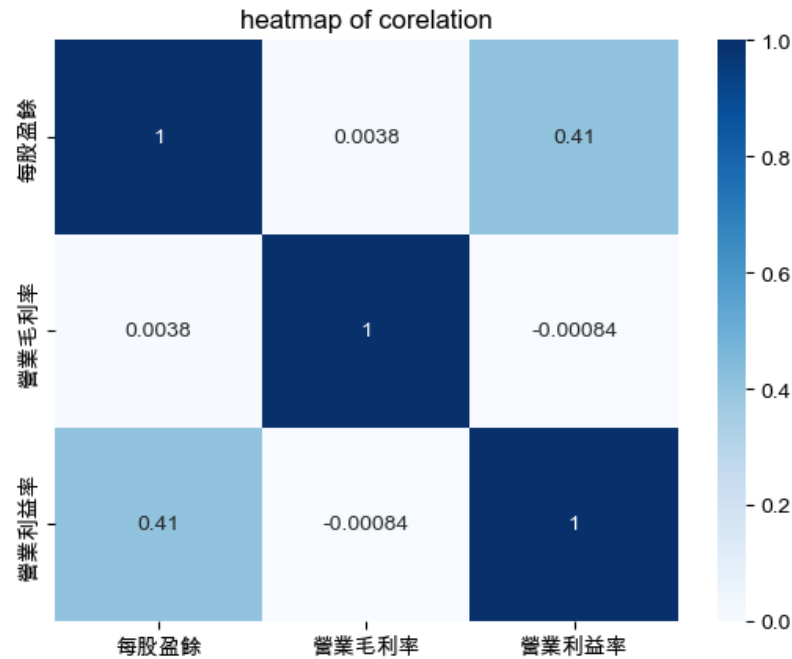
圖五-營業毛利率與每股盈餘之散佈圖：每股盈餘不論營業毛利率為何皆呈現低平穩的情況，但不排除考慮細微分布上的正相關，需要再分析

b. 營業利益率與每股盈餘之散佈圖



圖六-營業利益率與每股盈餘之散佈圖：有明顯集中向上趨勢

5. 相關係數熱力圖：



圖七-相關係數熱力圖：如前兩表，營業利益率和每股盈餘有更高的正相關性

三、迴歸模型配適

1. 簡單迴歸模型：

a. X：營業毛利率 Y：每股盈餘

| OLS regression 每股盈餘 ~ 營業毛利率 |          |                     |        |
|-----------------------------|----------|---------------------|--------|
| R-squared                   | 0.000    | Adj. R-squared      | -0.005 |
| F-statistic                 | 0.002684 | Prob (F-statistics) | 0.959  |

|       | coef   | std err  | t        | P >  t |
|-------|--------|----------|----------|--------|
| 營業毛利率 | 0.0032 | 0.062    | 0.052    | 0.959  |
|       | JB     | 3888.790 | Prob(JB) | 0.00   |

b. X：營業利益率 Y：每股盈餘

| OLS regression 每股盈餘 ~ 營業利益率 |       |                     |          |
|-----------------------------|-------|---------------------|----------|
| R-squared                   | 0.167 | Adj. R-squared      | 0.162    |
| F-statistic                 | 37.66 | Prob (F-statistics) | 4.89e-09 |

|       | coef   | std err  | t        | P >  t |
|-------|--------|----------|----------|--------|
| 營業利益率 | 0.4438 | 0.072    | 6.136    | 0.000  |
|       | JB     | 4615.710 | Prob(JB) | 0.00   |

由上述的兩張分析結果圖表可以發現，每股盈餘和營業毛利率的R平方值非常

低，低至小數點三位無法顯示，代表其解釋到的變異比例極低，P 值為 0.959，代表期相關性非常低。而每股盈餘和營業利率的 R 平方值則為 0.167，解釋的變異比例比營業利率高出許多，而在 P 值的部分為 0.00，表示其相關性非常顯著。值得注意的是，兩個分析結果的 Prob(JB)皆為 0.00，代表殘差的分佈不成常態分佈，違背原本分析的假設狀態，故此結果可能不準確。

## 2. 多項式迴歸模型：

X：營業利率+營業毛利率 Y：每股盈餘

| OLS regression 每股盈餘 ~ 營業毛利率+營業利率 |       |                     |          |
|----------------------------------|-------|---------------------|----------|
| R-squared                        | 0.167 | Adj. R-squared      | 0.158    |
| F-statistic                      | 18.73 | Prob (F-statistics) | 3.85e-08 |

|       | coef   | std err  | t        | P >  t |
|-------|--------|----------|----------|--------|
| const | 3.7402 | 0.896    | 4.175    | 0.000  |
| 營業毛利率 | 0.0035 | 0.057    | 0.062    | 0.951  |
| 營業利率  | 0.4438 | 0.073    | 6.120    | 0.000  |
|       | JB     | 4597.098 | Prob(JB) | 0.00   |

結果類似簡單回歸中每股盈餘對營業利率的結果，P 值為 0.000 可代表統計上的顯著，不過仍染有殘差不成常態分佈的問題，我們可以綜合先前的資料圖表分析來推論有可能是異常值或極端值的影像導致模型會有這樣的問題，有可能會需要使用到更嚴格的方法來設定模型。

## 3. 啞變數迴歸模型：

X：營業利率+營業毛利率+年度 Y：每股盈餘

| OLS regression 每股盈餘 ~ 營業毛利率+營業利率+年度 |       |                     |       |
|-------------------------------------|-------|---------------------|-------|
| R-squared                           | 0.182 | Adj. R-squared      | 0.034 |
| F-statistic                         | 1.227 | Prob (F-statistics) | 0.213 |

|       | coef       | std err  | t        | P >  t |
|-------|------------|----------|----------|--------|
| const | 1.164e+09  | 1.45e+08 | 8.035    | 0.000  |
| 營業毛利率 | -5.089e+07 | 9.42e+06 | -5.401   | 0.000  |
| 營業利率  | 1.499e+07  | 1.78e+07 | 0.842    | 0.401  |
|       | JB         | 1494.059 | Prob(JB) | 0.00   |

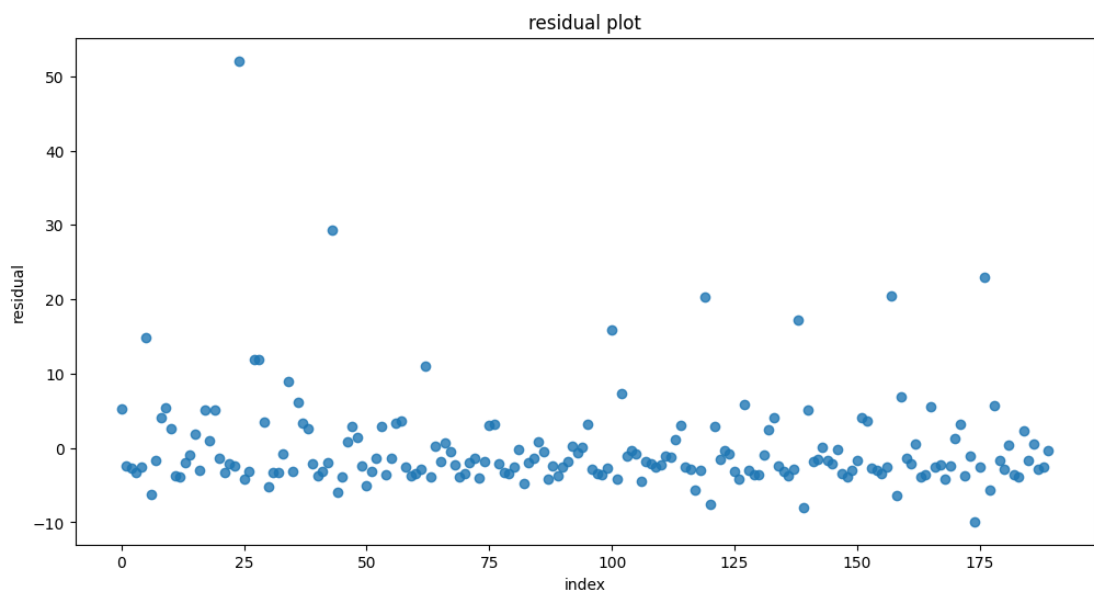
R 平方提升到 0.182，但調整後的 R 平方只有 0.034。這表明雖然整體模型解釋力有所提升，但考慮到自由度的調整，實際上可解釋的變異並不高。這可能是因為「年度」變數增加了模型的複雜度，但並沒有顯著增加模型的解釋力。

F-統計量降低到 1.227，且 P 值從  $3.85e-08$  變為 0.213，表示加入年度變數後模型的整體顯著性降低，可能意味著「年度」並不是一個影響「母公司盈餘」的重要因素。

#### 4. 殘差分析：

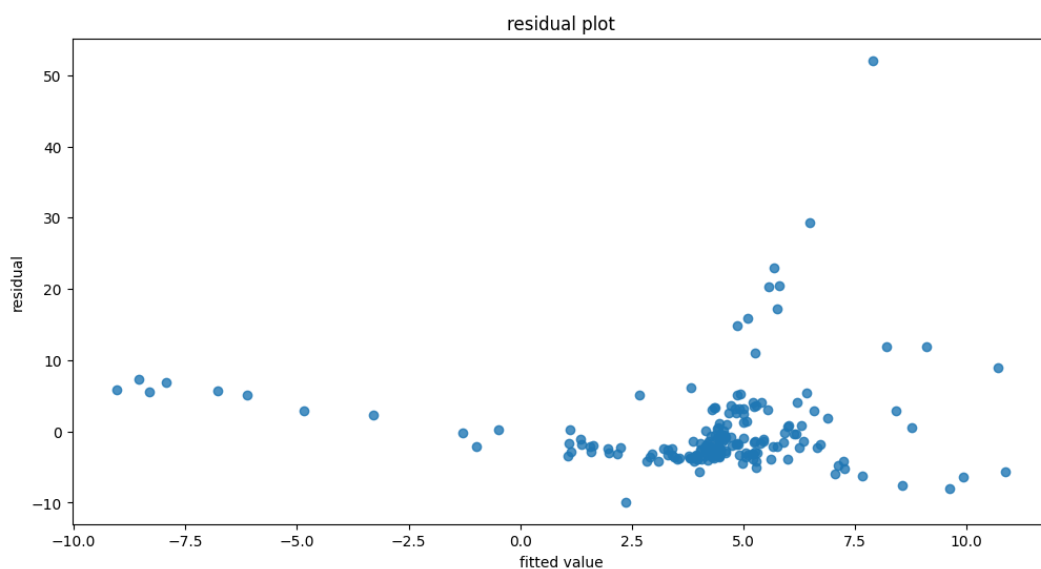
接下來我們以多項式回歸模型作為殘差分析的目標模型進行分析：

##### a. 獨立性（Independence）：



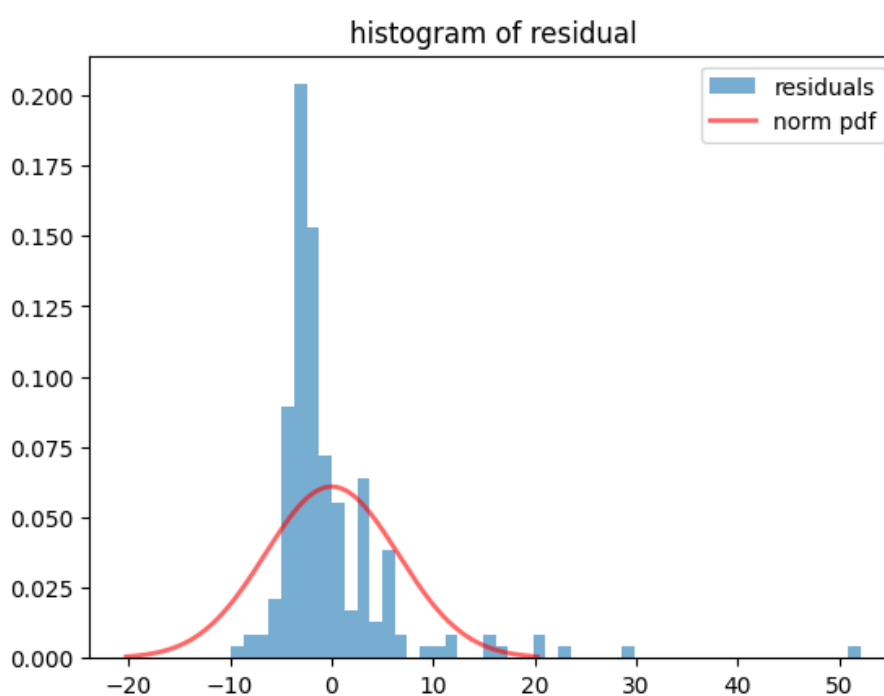
殘差圖中的點大多集中在 0 上下，但有許多極端值，特別是一些殘差達到 10 以上，甚至高達 50。這些大的殘差值可能表明模型對這些觀察值的擬合不佳。考慮對極端值進行處理，例如使用剔除、轉換或透過更健的迴歸技術來降低它們的影響。

##### b. 變異同質性（Homogeneity）：



在這個圖表中，當擬合值增加時，殘差似乎有向上擴散的趨勢，這可能表明存在異方差性（變異數不一致性）或模型未能完整某些方面。

c. 常態性（normality）：



與理想的長常態分布曲線相比，實際的殘差分布顯示出明顯的偏態和峰度。殘差的分布顯示中心峰值高於正態分布的峰值，且尾部較厚，這樣可能會導致回歸估計有誤，以及信賴區間和假設檢定的有效性。

5. 模型預估結果：

a. 每股盈餘 ~ 營業毛利率：

OLS regression 每股盈餘 ~ 營業毛利率



|             |          |                     |        |
|-------------|----------|---------------------|--------|
| R-squared   | 0.000    | Adj. R-squared      | -0.006 |
| F-statistic | 0.002786 | Prob (F-statistics) | 0.868  |

|       |        |          |          |        |
|-------|--------|----------|----------|--------|
|       | coef   | std err  | t        | P >  t |
| 營業毛利率 | 0.0122 | 0.073    | 0.167    | 0.868  |
|       | JB     | 2947.246 | Prob(JB) | 0.00   |

|      |                    |
|------|--------------------|
| MSE  | 16.563398340891435 |
| MAD  | 4.0441897290531115 |
| MAPE | 16.499172130942757 |

整體模型的解釋能力非常弱，統計模型缺乏統計顯著性

b. 每股盈餘 ~ 營業利益率：

| OLS regression 每股盈餘 ~ 營業毛利率 |       |                     |          |
|-----------------------------|-------|---------------------|----------|
| R-squared                   | 0.156 | Adj. R-squared      | 0.15     |
| F-statistic                 | 27.74 | Prob (F-statistics) | 4.73e-07 |

|       |        |          |          |        |
|-------|--------|----------|----------|--------|
|       | coef   | std err  | t        | P >  t |
| 營業利益率 | 0.4505 | 0.086    | 5.267    | 0.000  |
|       | JB     | 3373.658 | Prob(JB) | 0.00   |

|      |                    |
|------|--------------------|
| MSE  | 14.848703941441556 |
| MAD  | 3.5973331960773427 |
| MAPE | 14.802002418941882 |

模型具高度解釋能力，殘差的非常態分佈可能需要進一步的檢查和調整，以確保模型的假設條件得到滿足。雖然模型顯示一定的預測能力，但仍有改進的空間，尤其是在減少誤差指標方面。

c. 每股盈餘 ~ 營業毛利率 + 營業利益率：

| OLS regression 每股盈餘 ~ 營業毛利率 |       |                     |          |
|-----------------------------|-------|---------------------|----------|
| R-squared                   | 0.156 | Adj. R-squared      | 0.145    |
| F-statistic                 | 13.78 | Prob (F-statistics) | 3.23e-06 |

|       |        |         |       |        |
|-------|--------|---------|-------|--------|
|       | coef   | std err | t     | P >  t |
| 營業利益率 | 0.4505 | 0.086   | 5.247 | 0.000  |

|       |        |          |          |       |
|-------|--------|----------|----------|-------|
| 營業毛利率 | 0.0041 | 0.067    | 0.06     | 0.952 |
|       | JB     | 3359.445 | Prob(JB) | 0.00  |

|      |                    |
|------|--------------------|
| MSE  | 14.84763854728292  |
| MAD  | 3.5969220860745335 |
| MAPE | 14.901542959134046 |

和上一個模型幾乎差不多，不過解釋能力稍有下降，可以看出營業毛利率對每股盈餘的預估是沒什麼幫助的。

d. 每股盈餘 ~ 營業毛利率 + 營業利益率 + 年度：

| OLS regression 每股盈餘 ~ 營業毛利率 |       |                     |          |
|-----------------------------|-------|---------------------|----------|
| R-squared                   | 0.156 | Adj. R-squared      | 0.145    |
| F-statistic                 | 13.78 | Prob (F-statistics) | 3.23e-06 |

|       | coef   | std err  | t        | P >  t |
|-------|--------|----------|----------|--------|
| 營業利益率 | 0.4505 | 0.086    | 5.247    | 0.000  |
| 營業毛利率 | 0.0041 | 0.067    | 0.06     | 0.952  |
|       | JB     | 3359.445 | Prob(JB) | 0.00   |

|      |                    |
|------|--------------------|
| MSE  | 14.84763854728292  |
| MAD  | 3.5969220860745335 |
| MAPE | 14.901542959134046 |

結果和上一個模型是一模一樣的。

#### 四、結論

四個模型中，使用簡單回歸中的營業利益率和多項式回歸模型的表現是最好的，從一開始在敘述統計分析階段，就可以發現其實營業毛利率對每股盈餘不是一個很好的估計變數，除了散步圖可以看出來相關性不大之外，不管在一開始的 ols 分析還是後來的模型預估中也可以看出他的表現可以說是極差。其中在所有分析中都有一個問題，可以看出殘差的分佈不是常態分布，這樣可能會違背原本迴歸先前的假設，所以可能在資料處理時需要剔除異常值和極端值。不過總體來說，對於每股盈餘這個財務數據來說，可能需要更精密的分析方法，或是考慮到更多不同的變數，才可以得到更精確的模型。

#### 五、附錄

1. 分工表：

黃子維：程式碼、書面報告

劉芊儀：程式碼、書面報告

郭菲霓：程式碼、書面報告

2. Python 學習心得：

- a. 黃子維：這次在學期末用一個大型的 python 報告來實際操作財務報表的迴歸分析模型，除了讓我真的自己從頭到尾操作過一次這些統計工具和 python，我覺得最有趣的部分還是自己去解讀程式跑出來的數據，ols regression 的結果裡面有超級多數據，而且不同的 regression 結果間的變化也需要一步一步的分析，裡面還有一些比較不熟悉的内容相像 Durbin Watson 和 Jarque Bera 也是我上網搜尋查資料才熟悉的内容，我也體會到統計軟體真的很強大，並且很多商業上的數據我們都可以用更科學的方法去解讀他，或許我們這次的迴歸模型不是 100 分，但在過程中去了解變數分析出來的結果也有非常大的收穫，期待日後在不同的專題和報告中都能活用這學期所學的統計學理論和程式語言的應用，謝謝陳老師和助教的悉心教導。
- b. 劉芊儀：我覺得大家一起運用 Python 完成這份期末報告，更讓我有一種真正在運用所學知識的體驗。比起是在紙上計算練習題，我更喜歡利用資料來實際動手操作，自己去跑資料，去分析計算結果，才更讓我有一種真正了解統計學這門科目的概念，其實跑出迴歸分析資料後要去解讀不同統計數據也很需要時間，在不同變數做回歸後的結果之間的比較是最具挑戰也最令人印象深刻的部分。這次期末報告真的讓我獲益良多，我變得更加熟悉 Python 這個語言，以及更加了解關於利用 Python 來進行統計方面的應用。這些收穫甚至得以讓我應用在其他課堂上的報告，協助我能夠對於數據資料有更全面且專業的分析，真的很謝謝教授和助教們指導，才可以讓我從中學到這麼多東西。
- c. 郭菲霓：透過這學期的 python 課程，我學到怎麼利用程式跑以及呈現統計課上學到的知識，讓我們不用花太多的時間在手動計算許多複雜的公式及數據，而是將時間花在解讀每個統計圖表的意義。在大數據時代，擁有快速分析數據特性及解讀統計圖表意義的能力更尤為重要，因此很謝謝助教每週帶我們了解及練習不同模型及實作各式各樣的題目。