計量經濟 HW0303

2.17 The data file *collegetown* contains observations on 500 single-family houses sold in Baton Rouge, Louisiana, during 2009–2013. The data include sale price (in thousands of dollars), *PRICE*, and total interior area of the house in hundreds of square feet, *SQFT*.

a. Plot house price against house size in a scatter diagram.

---

b. Estimate the linear regression model $PRICE = \beta_1 + \beta_2 SQFT + e$. Interpret the estimates. Draw a sketch of the fitted line.

c. Estimate the quadratic regression model $PRICE = \alpha_1 + \alpha_2 SQFT^2 + e$. Compute the marginal effect of an additional 100 square feet of living area in a home with 2000 square feet of living space.

d. Graph the fitted curve for the model in part (c). On the graph, sketch the line that is tangent to the curve for a 2000-square-foot house.

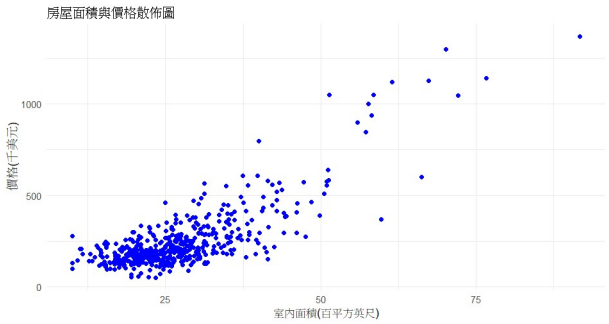
e. For the model in part (c), compute the elasticity of *PRICE* with respect to *SQFT* for a home with 2000 square feet of living space.

f. For the regressions in (b) and (c), compute the least squares residuals and plot them against *SQFT*. Do any of our assumptions appear violated?

g. One basis for choosing between these two specifications is how well the data are fit by the model. Compare the sum of squared residuals (*SSE*) from the models in (b) and (c). Which model has a lower *SSE*? How does having a lower *SSE* indicate a "better-fitting" model?

2.17

(a)



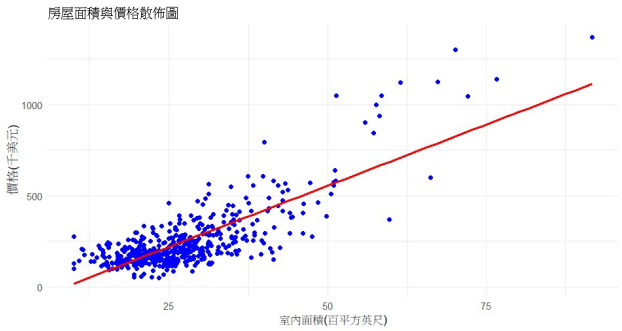
(b)

```
Residuals:
    Min      1Q  Median      3Q     Max
-316.93  -58.90   -3.81   47.94  477.05

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -115.4236    13.0882  -8.819   <2e-16 ***
sqft          13.4029     0.4492  29.840   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 102.8 on 498 degrees of freedom
Multiple R-squared:  0.6413,    Adjusted R-squared:  0.6406
F-statistic: 890.4 on 1 and 498 DF,  p-value: < 2.2e-16
```

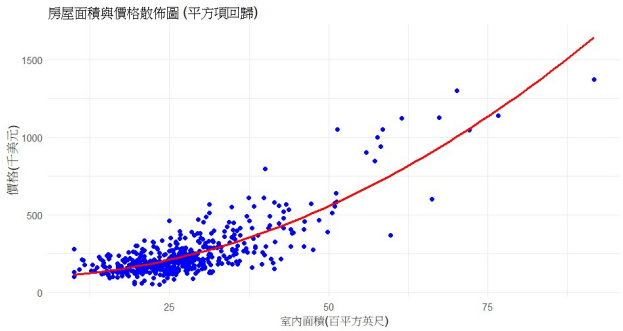
係數為 13.4029，表示室內面積每增加 100 平方英尺，預計的價格會增加 13.4029 千美元，截距表示室內面積為 0 時的價格，在此資料中不具經濟意義。



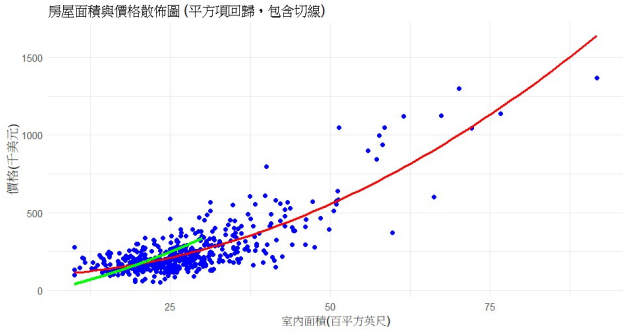
(c)

由 price=β0+β1(sqft)2，取偏微分得到邊際影響 d(sqft)/d(price)=2β2·sqft

因此當 sqft＝20 時，邊際影響為 2 * 13.4029 * 20＝536.116，即當室內面積為 2000 平方英尺時，室內面積每增加 100 平方英尺，價格會上升 536.116 千美元。



(d)



(e)

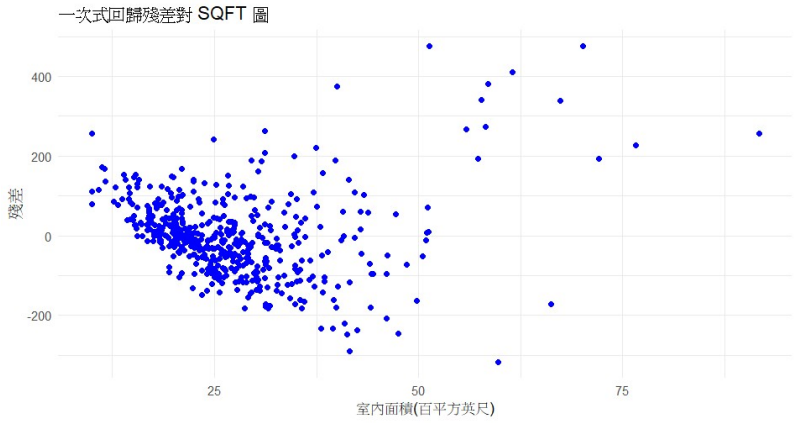
Eprice, sqft = d(sqft)/d(price)×price/sqft

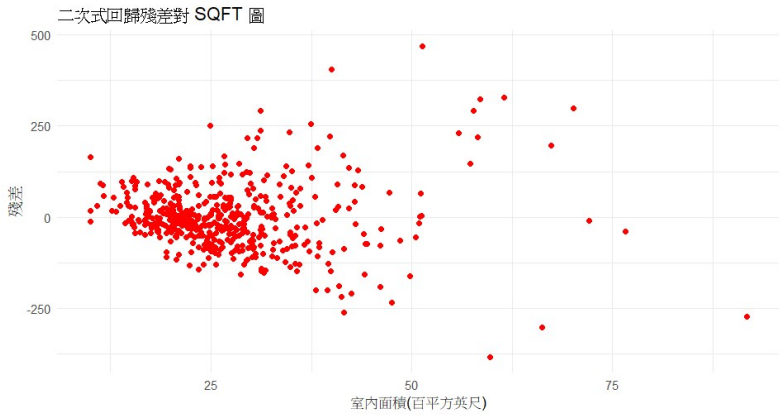
Sqft = 20 時，price = 167.3735，d(sqft)/d(price) = 536.116

因此彈性為 536.116 * 167.3735 / 20 = 4486.5806

(f)

由於殘差的變異數隨著 sqft 增長，表示資料違反了同質性假設。





(g)

線性回歸的 SSE = 5262847

二次式回歸的 SSE = 4222356

由於 SSE 表示的是殘差的平方和，越小的 SSE 表示越接近真實資料，因此二次式的擬合程度較好。

2.25 Consumer expenditure data from 2013 are contained in the file *cex5_small*. [Note: *cex5* is a larger version with more observations and variables.] Data are on three-person households consisting of a husband and wife, plus one other member, with incomes between $1000 per month to $20,000 per month. *FOODAWAY* is past quarter's food away from home expenditure per month per person, in dollars, and *INCOME* is household monthly income during past year, in $100 units.

a. Construct a histogram of *FOODAWAY* and its summary statistics. What are the mean and median values? What are the 25th and 75th percentiles?

b. What are the mean and median values of *FOODAWAY* for households including a member with an advanced degree? With a college degree member? With no advanced or college degree member?

c. Construct a histogram of ln(*FOODAWAY*) and its summary statistics. Explain why *FOODAWAY* and ln(*FOODAWAY*) have different numbers of observations.

d. Estimate the linear regression $\ln(FOODAWAY) = \beta_1 + \beta_2 INCOME + e$. Interpret the estimated slope.

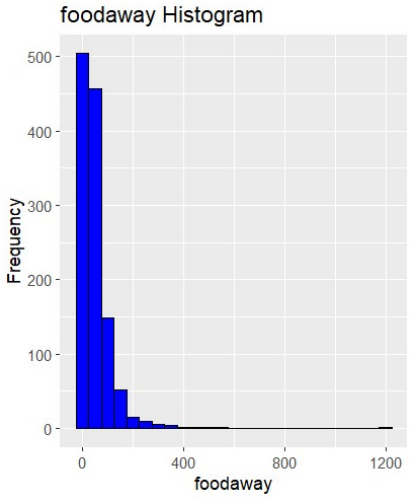
e. Plot ln(*FOODAWAY*) against *INCOME*, and include the fitted line from part (d).

f. Calculate the least squares residuals from the estimation in part (d). Plot them vs. *INCOME*. Do you find any unusual patterns, or do they seem completely random?

2.25

(a)

Mean = 49.27，Median = 32.55，25^{th} = 12.04，75^{th} = 67.50



(b)

對 advanced :

Mean = 73.1549，Median = 48.15

對 college :

Mean = 48.5972，Median = 36.11

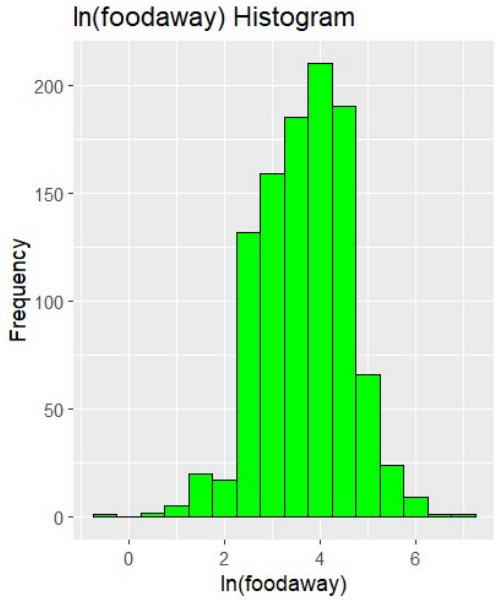
對 no degree :

Mean = 39.0102，Median = 26.02

(c)

Mean = 3.6508，Median = 3.6865

由於對 foodaway 取了自然對數，因此若家庭沒有外出用餐的紀錄，foodaway 會被記錄成 0，再取 ln 之後會變得無意義，會被視為缺失值，才導致樣本數變少。

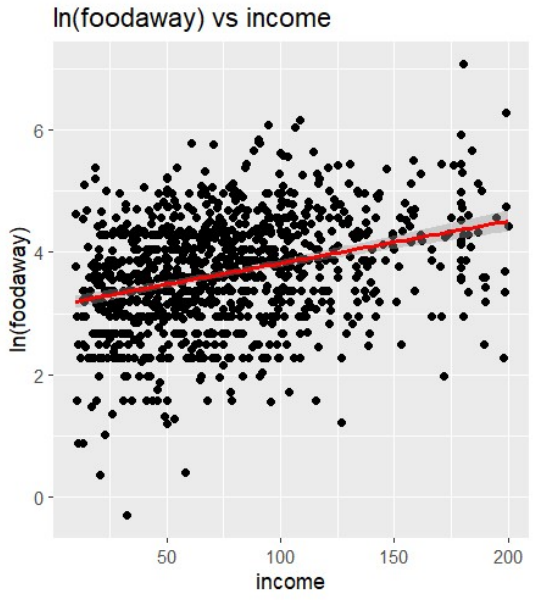


(d)

迴歸係數 ＝0.0069，截距 ＝3.1293

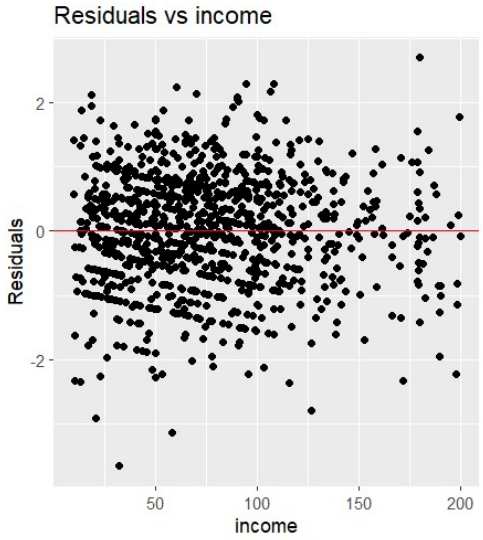
斜率代表收入每增加 100 美元，ln(foodaway)預期增加 0.0069 美元

(e)



(f)

從殘差圖來看，我不認為此資料違反同質性假設或資料非隨機，僅有的問題只是樣本收入多集中於 0~150 之區間而已。



2.28 How much does education affect wage rates? The data file *cps5_small* contains 1200 observations on hourly wage rates, education, and other variables from the 2013 Current Population Survey (CPS). [Note: *cps5* is a larger version.]

- **a.** Obtain the summary statistics and histograms for the variables *WAGE* and *EDUC*. Discuss the data characteristics.
- **b.** Estimate the linear regression $WAGE = \beta_1 + \beta_2 EDUC + e$ and discuss the results.
- **c.** Calculate the least squares residuals and plot them against *EDUC*. Are any patterns evident? If assumptions SR1–SR5 hold, should any patterns be evident in the least squares residuals?
- **d.** Estimate separate regressions for males, females, blacks, and whites. Compare the results.
- **e.** Estimate the quadratic regression $WAGE = \alpha_1 + \alpha_2 EDUC^2 + e$ and discuss the results. Estimate the marginal effect of another year of education on wage for a person with 12 years of education and for a person with 16 years of education. Compare these values to the estimated marginal effect of education from the linear regression in part (b).
- **f.** Plot the fitted linear model from part (b) and the fitted values from the quadratic model from part (e) in the same graph with the data on *WAGE* and *EDUC*. Which model appears to fit the data better?

2.28

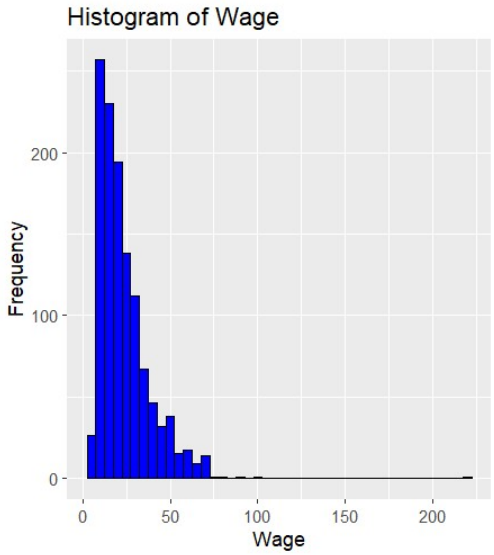
(a)

Wage 的敘述統計量 ：

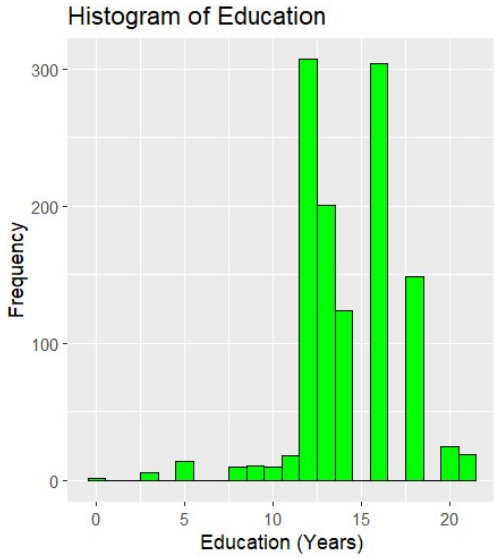
```
   Min.   1st Qu.  Median    Mean    3rd Qu.   Max.
   3.94   13.00    19.30     23.64   29.80     221.10
```

Educ 的敘述統計量 ：

```
   Min. 1st Qu.  Median    Mean   3rd Qu.  Max.
    0.0    12.0    14.0    14.2    16.0    21.0
```



Wage 呈現右長尾分布。



Educ 則呈現左長尾分布，並且集中在 12 及 16 年。

(b)

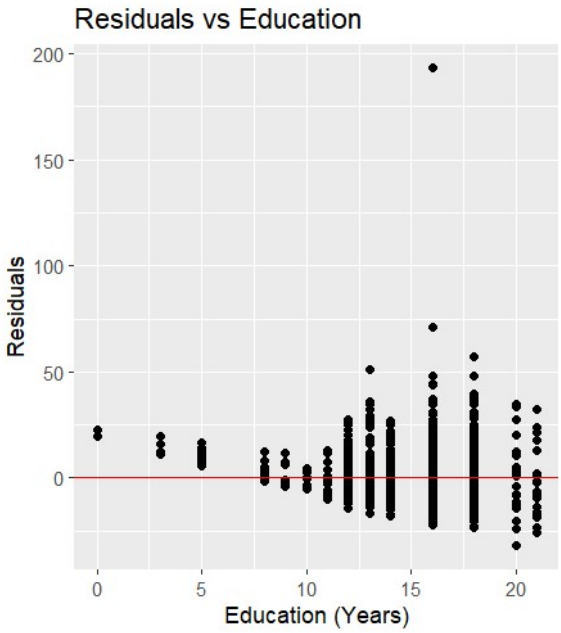
迴歸模型斜率為 2.3968，截距為-10.4000

截距在此迴歸中不具經濟意義，因為當教育年數為 0 時也不可能出現負薪資，而斜率表示教育年數每增加一年，薪資預期將增加 2.3968 單位美元。

(c)

由這張殘差圖，看起來殘差對解釋變數有一定的負斜率關係，可能違反了解是變數於殘差無相關的假設，並且隨著教育年數增加

殘差的變藝術也在增加，這表示資料違反了同質性假設，如果 SR1-SR5 都成立的話，不應該觀察到任何趨勢或模式，殘差應該是服從常態分配且完全隨機的。



(d)

由於截距在此資料中不具經濟意義，因此我著重討論斜率，以下是個性別及人種教育程度對薪資迴歸的斜率項：

男性：2.3785，女性：2.6595，黑人：1.9233，白人：2.418

以結論來說，受教育對薪資的影響，女性比男性更大，白人又比黑人更大。

(e)

二次式迴歸：

Wage = 4.9165 + 0.08913 * educ ^2

邊際效應為 2 * 0.08913 * educ，因此：

教育年份為 12 年時的邊際效應：2.1392

教育年份為 16 年時的邊際效應：2.8523

在(b)中，任何教育年份的邊際影響都是 2.3968，可知教育年份為 12 年時邊際影響小於簡單迴歸之邊際影響，教育年份為 16 年時大於簡單迴歸之邊際影響。

(f)

單純以圖形觀察，二次項迴歸較符合真實的資料型態。