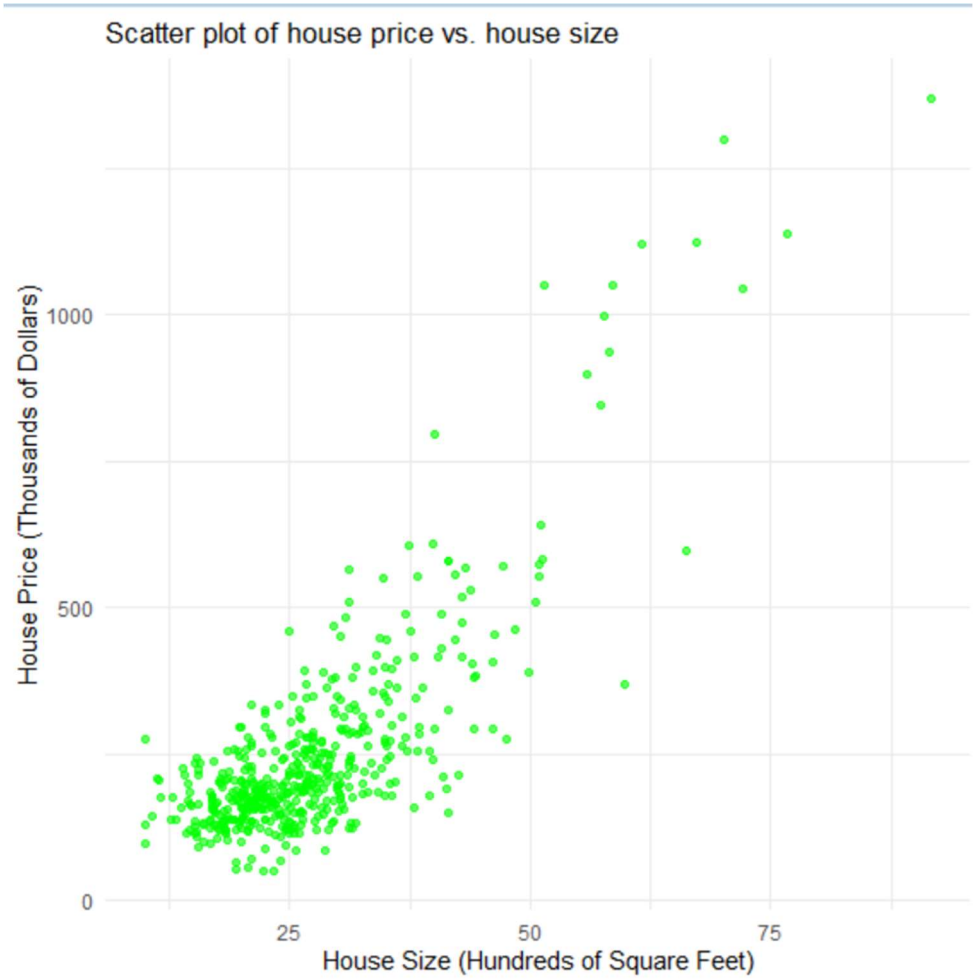
CH02　311707021 丁振朝

Q17

a.

資料型態

```
'data.frame':   500 obs. of  8 variables:
 $ price    : num  96.5 106 114 119.9 122.7 ...
 $ sqft     : num  16.8 17.5 18.4 29.1 14.7 ...
 $ age      : int  10 10 9 9 9 8 10 9 7 10 ...
 $ pool     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ fireplace: int  0 0 0 0 0 0 0 0 0 0 ...
 $ close    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ twostory : int  0 0 0 0 0 0 0 0 0 0 ...
 $ occupied : int  0 0 1 0 1 0 1 1 0 1 ...
```



b.

回歸結果

```
Call:
lm(formula = price ~ sqft, data = collegetown)

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
```

c.

```
Call:
lm(formula = price ~ I(sqft^2), data = collegetown)

Residuals:
    Min      1Q  Median      3Q     Max
-383.67  -48.39   -7.50   38.75  469.70

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 93.565854   6.072226   15.41   <2e-16 ***
I(sqft^2)    0.184519   0.005256   35.11   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 92.08 on 498 degrees of freedom
Multiple R-squared:  0.7122,    Adjusted R-squared:  0.7117
F-statistic:  1233 on 1 and 498 DF,  p-value: < 2.2e-16
```

從上面係數可得模型 ：Price = 93.5659 + 0.1845 SQFT2，邊際效應要進行微分可得模型 = 2*0.1845*SQFT，將 SQTF=20 帶入可得 Price = 7.3808

表示每 100 平方米會增加 738.08 元

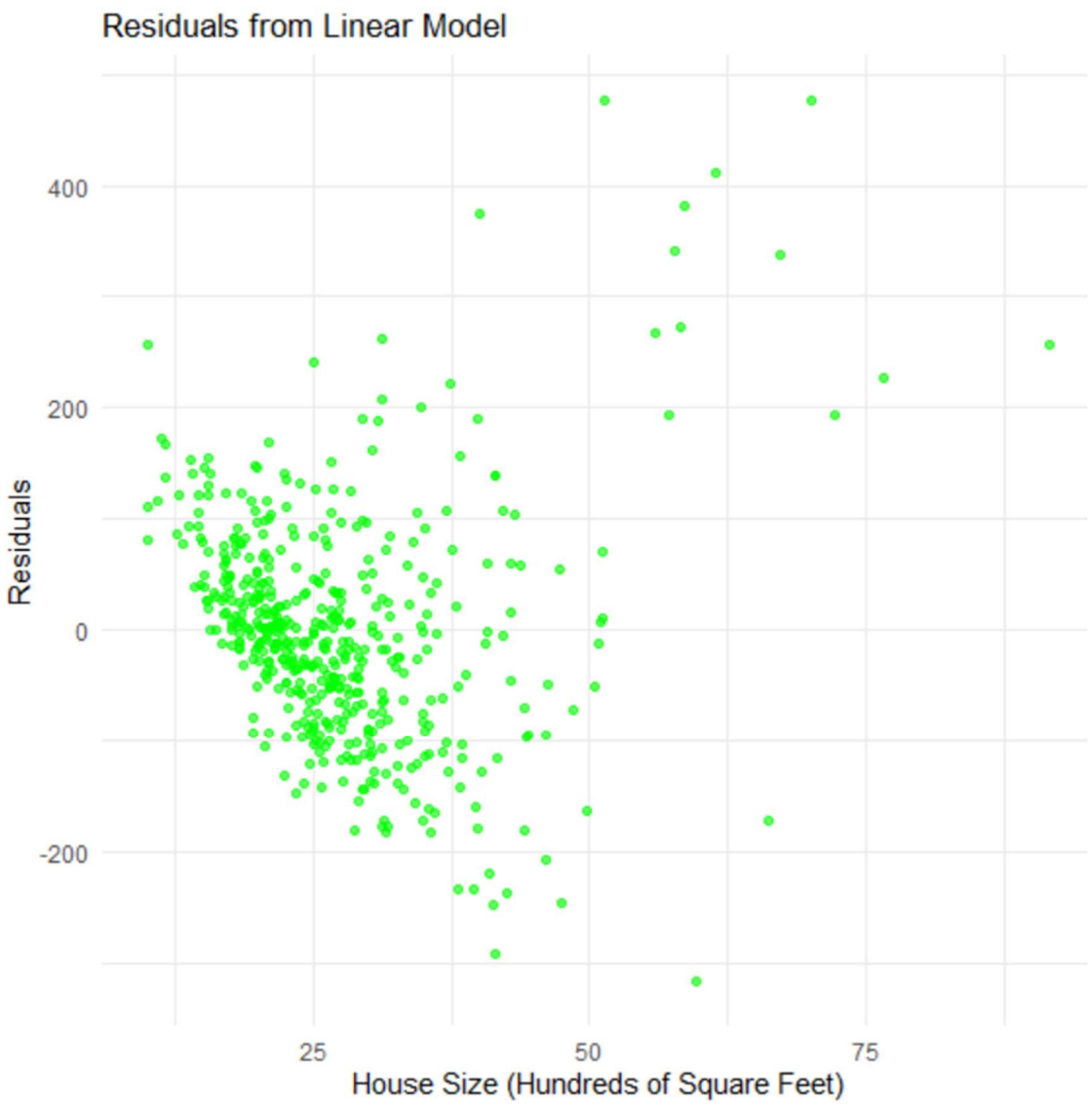
d.

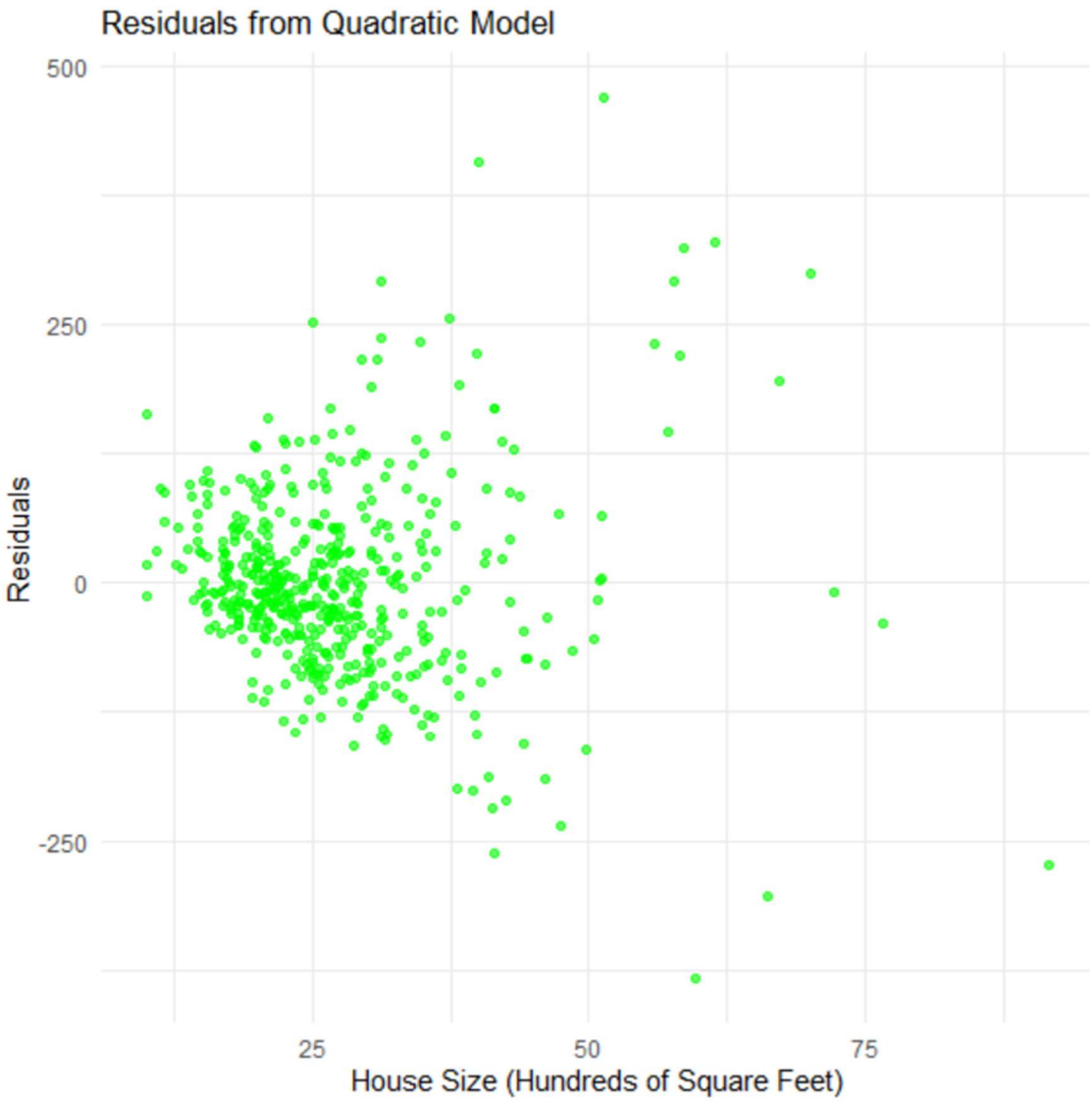
e.

弾性 = 0.88195

```
I(sqft^2)
0.8819511
```

f.


Residuals from Linear Model
400
200
0
-200
Residuals
25
50
75
House Size (Hundreds of Square Feet)



g.

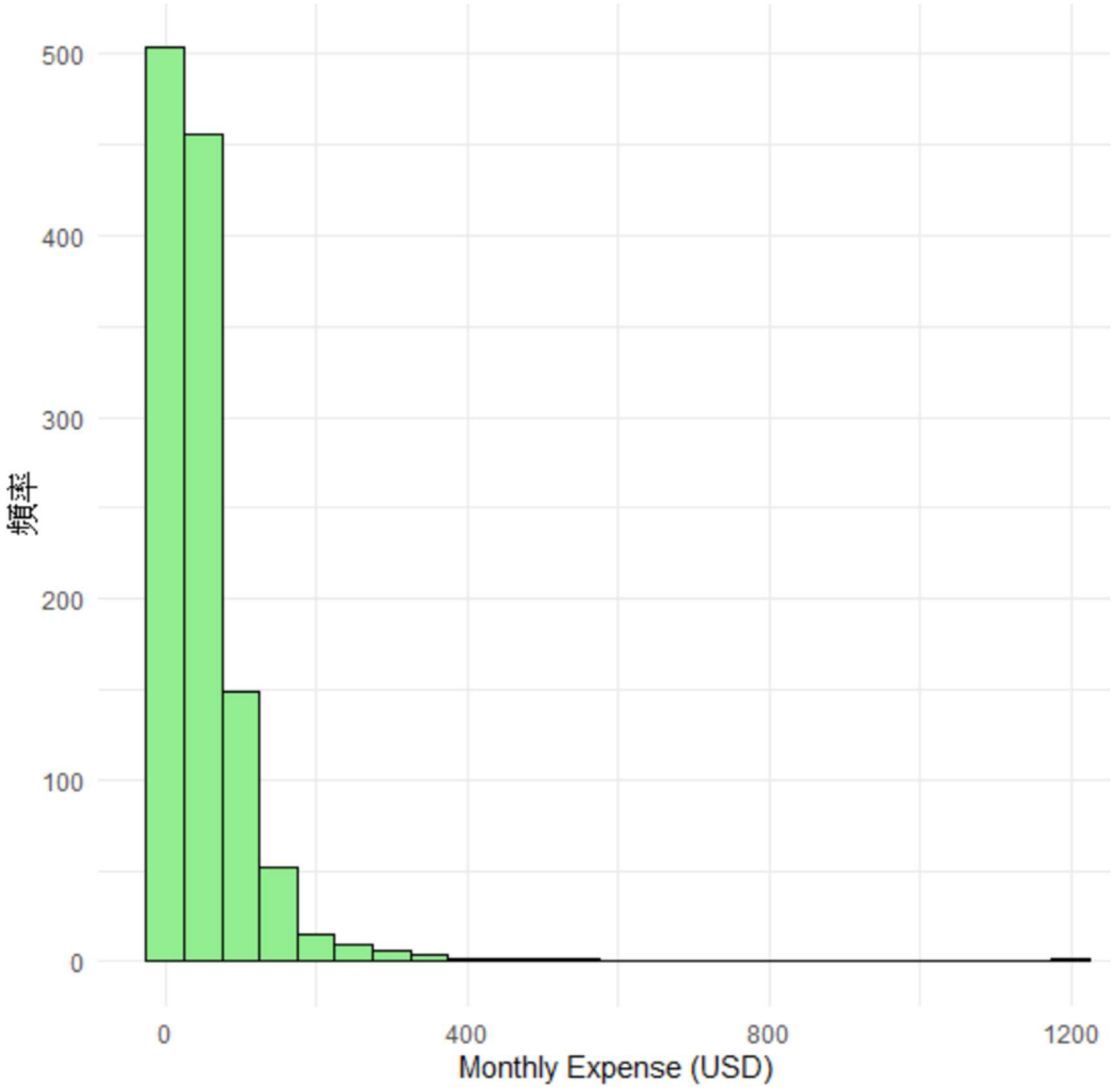
SSE for the linear = 5,262,847.

SSE for the quadratic = 4,222,356.

```
> SSE_linear
[1] 5262847
> SSE_quadratic
[1] 4222356
```

Q25

A



mean = 49.2709

median = 32.555

percentiles 25th = 12.0400

percentiles 75th = 67.5025

B.

	Degree	Mean_FOODAWAY	Median_FOODAWAY	Sample
1	Advanced Degree	73.15494	48.15	257
2	College Degree	48.59718	36.11	369
3	No Degree	39.01017	26.02	574

C.

```
> sum(data$foodaway <= 0, na.rm = TRUE)
[1] 178
```

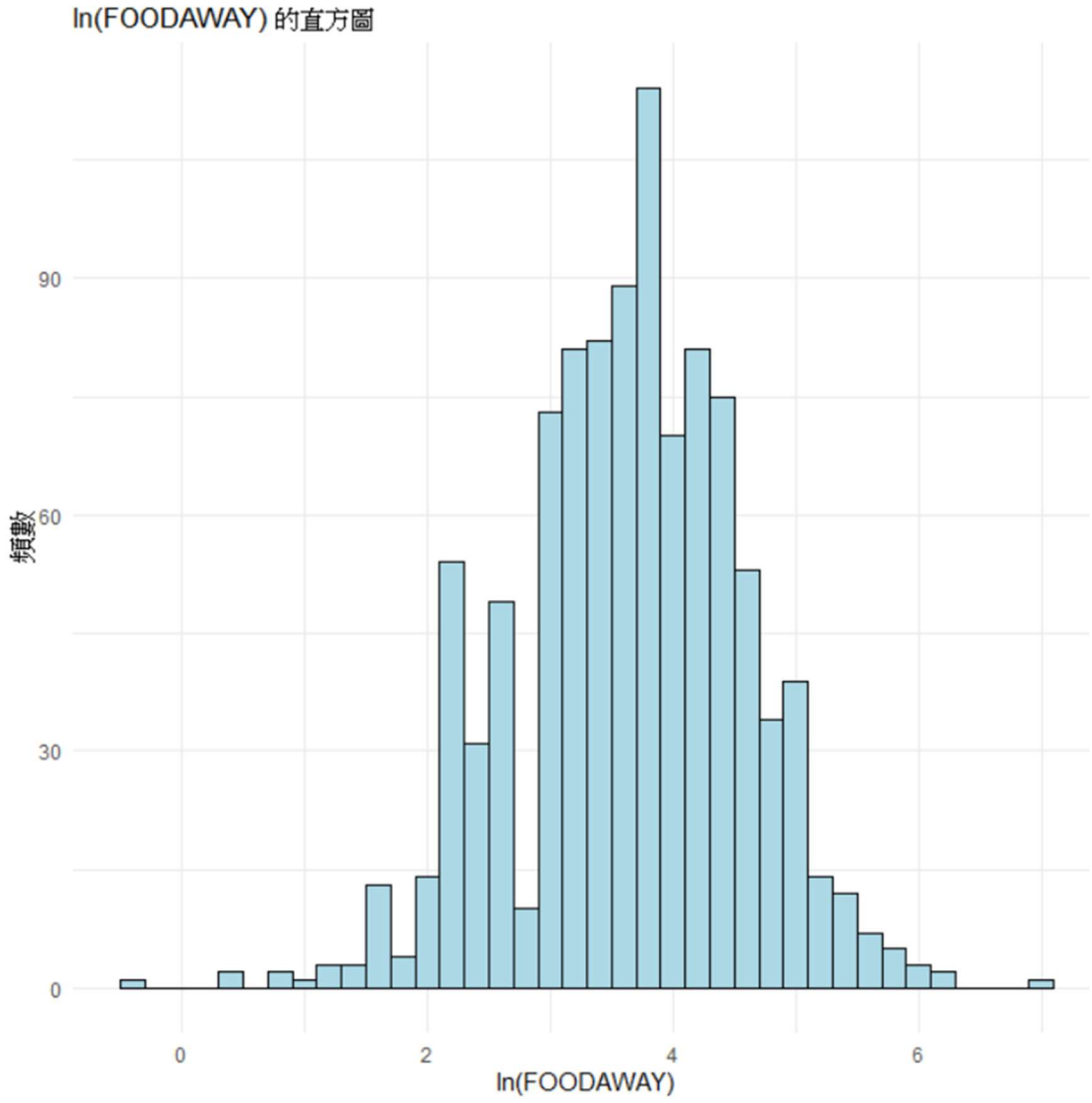
經過 0 或是負值檢定發現有 178 筆資料，把它排除並取 log

mean = 3.6508

median = 3.6845

percentiles 25th = 3.0759

percentiles 75th = 4.2797

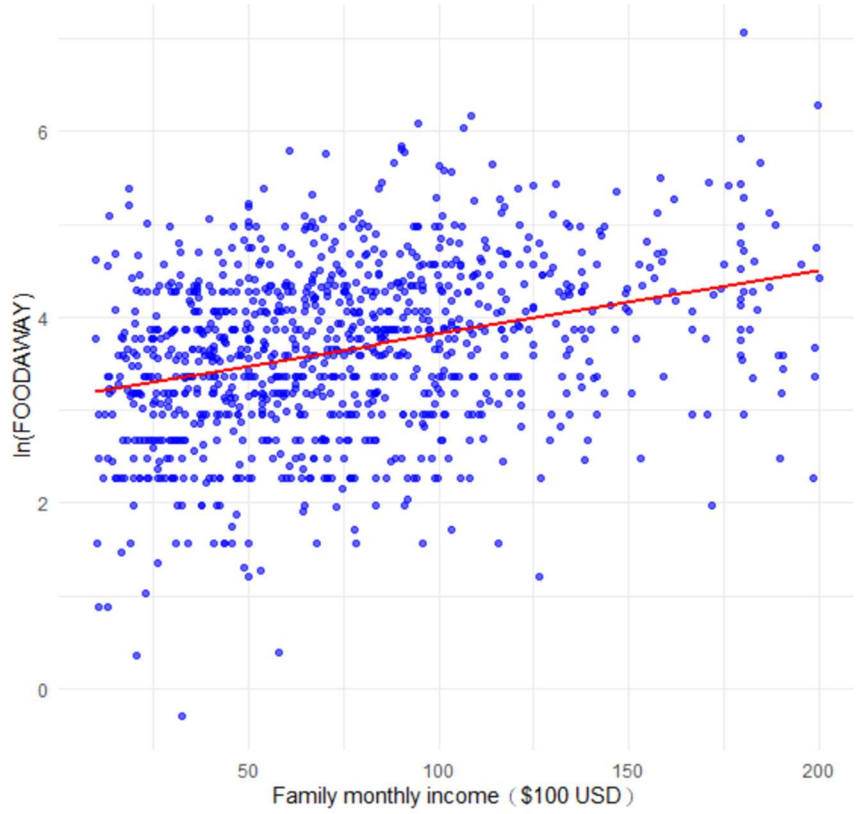


D.

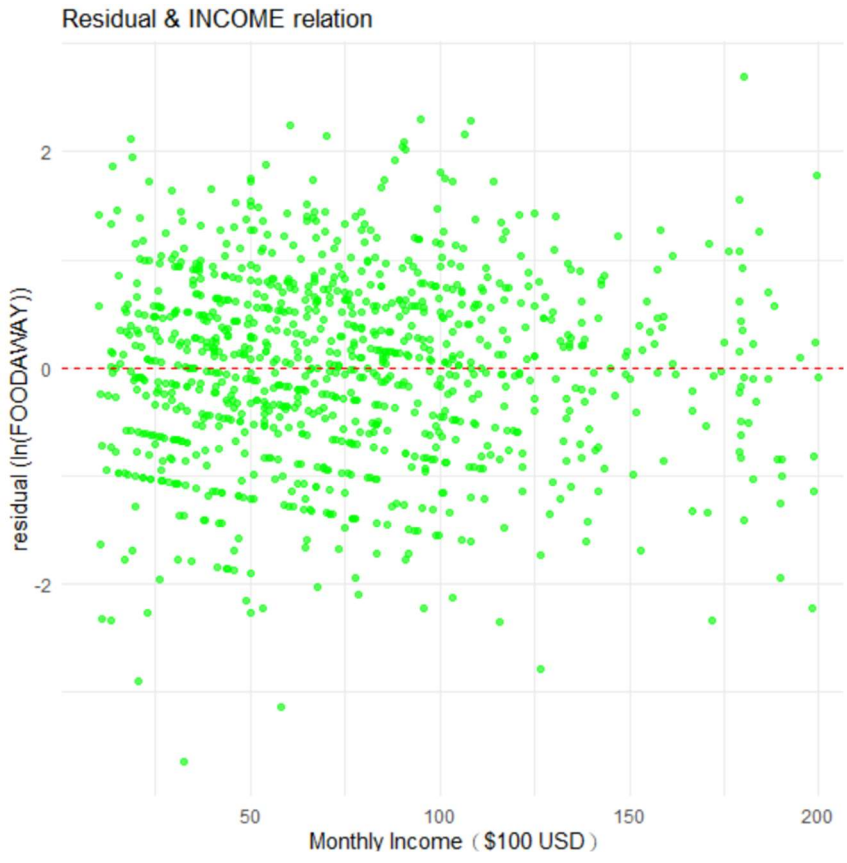
回歸模型如下:

```
Call:
lm(formula = ln_foodaway ~ income, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6547 -0.5777  0.0530  0.5937  2.7000

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.1293004  0.0565503   55.34   <2e-16 ***
income      0.0069017  0.0006546   10.54   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8761 on 1020 degrees of freedom
Multiple R-squared:  0.09826,   Adjusted R-squared:  0.09738
F-statistic: 111.1 on 1 and 1020 DF,  p-value: < 2.2e-16
```

E


ln(FOODAWAY)
6
4
2
0
50
100
150
200
Family monthly income（$100 USD）

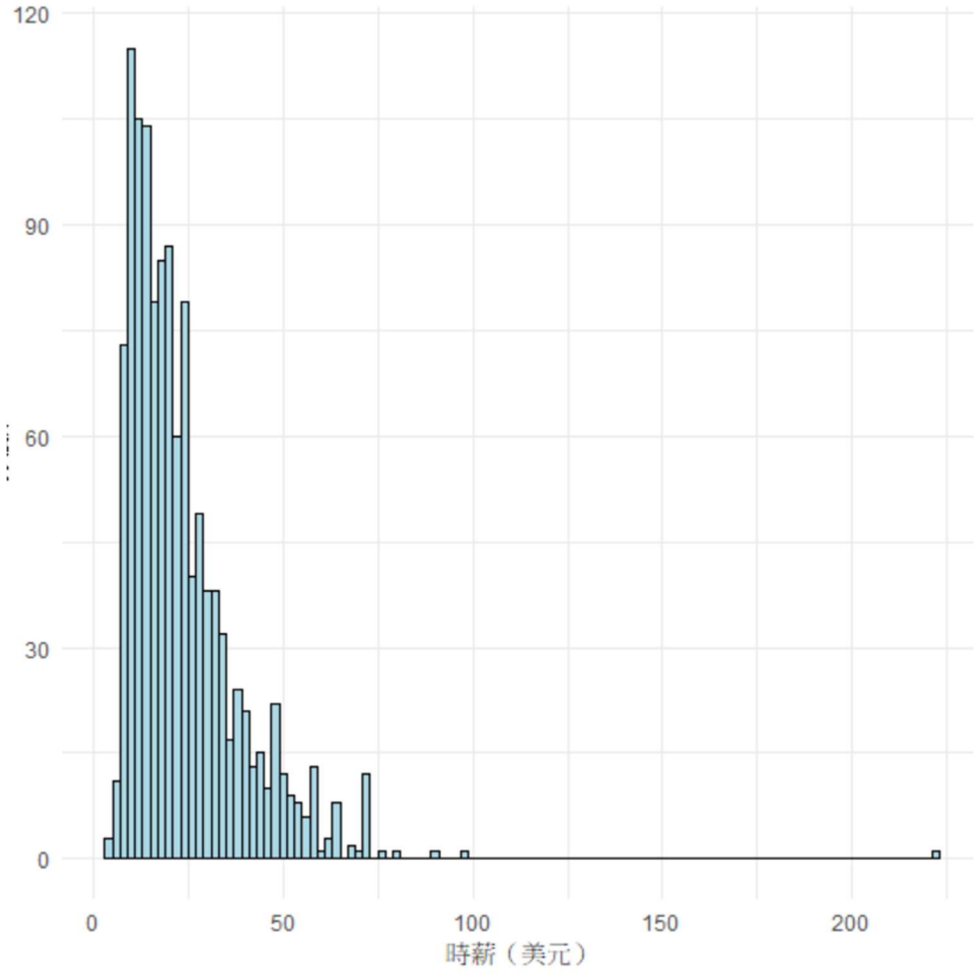
F

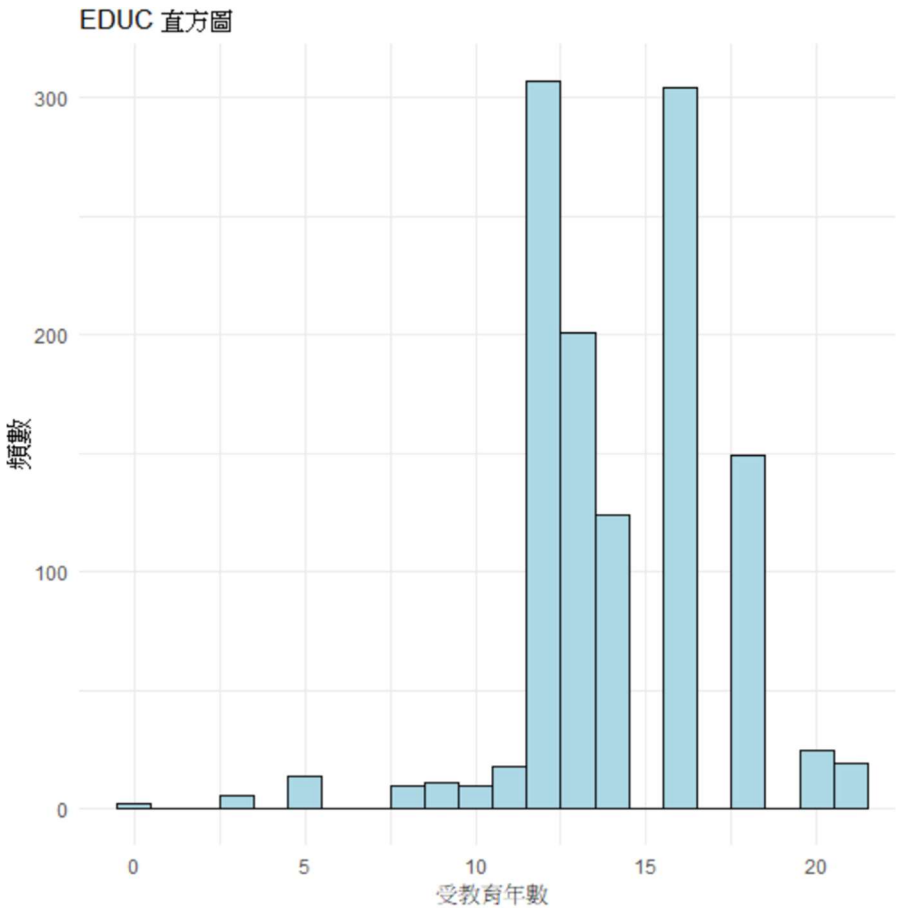

Residual & INCOME relation
residual (ln(FOODAWAY))
2
0
-2
50
100
150
200
Monthly Income（$100 USD）

Q28

A

```
> summary(data$wage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   3.94   13.00   19.30   23.64   29.80  221.10
> summary(data$educ)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0    12.0    14.0    14.2    16.0    21.0
```





從統計數據跟薪資來看，可以發現到薪資的分佈呈現右偏，有少數極端高薪者影響整體平均值。而教育年數則顯示出典型的學制影響，12 年（高中）與 16 年（大學）是主要的群體。進一步分析可以探討教育年數與薪資之間的關聯，可能會發現受教育程度較高者的薪資水準也較高。

Q28

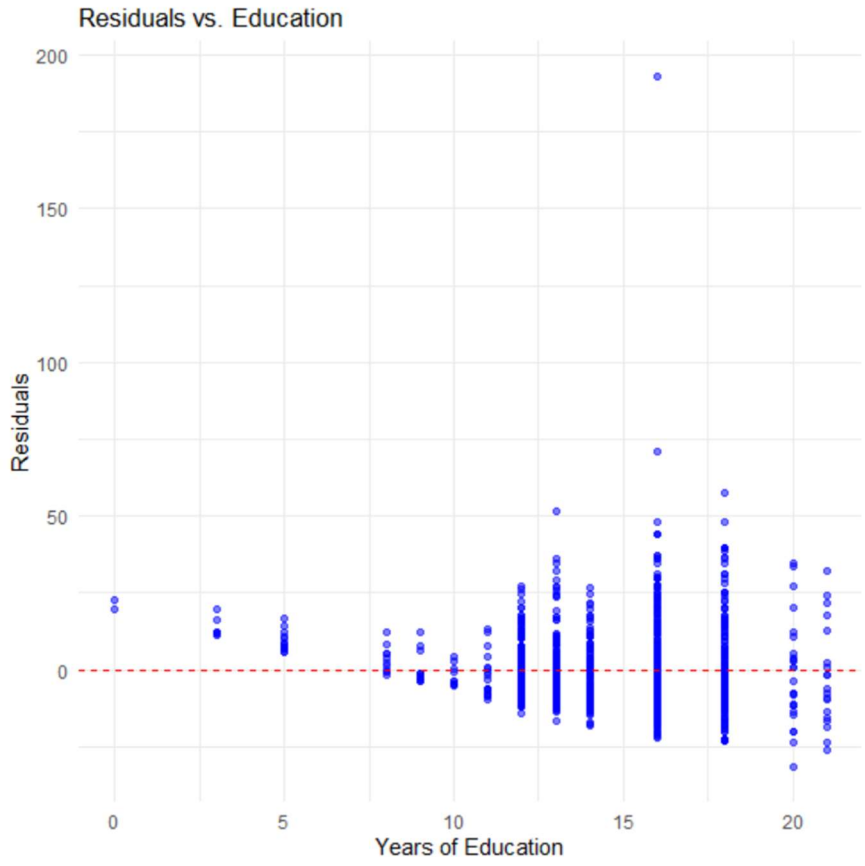
B

```
Call:
lm(formula = wage ~ educ, data = cps5_small)

Residuals:
    Min      1Q  Median      3Q     Max
-31.785  -8.381  -3.166   5.708 193.152

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.4000     1.9624    -5.3 1.38e-07 ***
educ          2.3968     0.1354    17.7  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.55 on 1198 degrees of freedom
Multiple R-squared:  0.2073,    Adjusted R-squared:  0.2067
F-statistic: 313.3 on 1 and 1198 DF,  p-value: < 2.2e-16
```

線性回歸模型顯示教育年限對工資有顯著的正向影響（每增加 1 年教育，工資增加約 2.4 單位），但模型的解釋力有限（R 平方僅 0.2073），預測誤差較大（殘差標準誤 13.55），且可能存在異常值或非線性關係

Q28

C



結果顯示存在方差不均（Heteroscedasticity），可以看到教育年限低於 10 年時，殘差的變異性較小，點分佈較集中。教育年限在 10-16 年時，殘差的變異性顯著增加，點分佈更分散。教育年限高於 16 年時，殘差變異性似乎略有減小，但仍比低教育年限時大。 這表明方差並非恆定，違反了 SR4（方差齊性假設）。

Q28

D

男性

```
Call:
lm(formula = wage ~ educ, data = cps5_small, subset = (female ==
    0))

Residuals:
    Min      1Q  Median      3Q     Max
-27.643  -9.279  -2.957   5.663 191.329

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -8.2849     2.6738  -3.099  0.00203 **
educ          2.3785     0.1881  12.648  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.71 on 670 degrees of freedom
Multiple R-squared:  0.1927,    Adjusted R-squared:  0.1915
F-statistic:   160 on 1 and 670 DF,  p-value: < 2.2e-16
```

女性

```
Call:
lm(formula = wage ~ educ, data = cps5_small, subset = (female ==
    1))

Residuals:
    Min      1Q  Median      3Q     Max
-30.837  -6.971  -2.811   5.102  49.502

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -16.6028     2.7837  -5.964 4.51e-09 ***
educ          2.6595     0.1876  14.174  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.5 on 526 degrees of freedom
Multiple R-squared:  0.2764,    Adjusted R-squared:  0.275
F-statistic: 200.9 on 1 and 526 DF,  p-value: < 2.2e-16
```

黑人

```
Call:
lm(formula = wage ~ educ, data = cps5_small, subset = (black ==
    1))

Residuals:
    Min      1Q  Median      3Q     Max
-15.673  -6.719  -2.673   4.321  40.381

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -6.2541     5.5539  -1.126    0.263
educ          1.9233     0.3983   4.829 4.79e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.51 on 103 degrees of freedom
Multiple R-squared:  0.1846,    Adjusted R-squared:  0.1767
F-statistic: 23.32 on 1 and 103 DF,  p-value: 4.788e-06
```

白人

```
Call:
lm(formula = wage ~ educ, data = cps5_small, subset = (black ==
    0))

Residuals:
    Min      1Q  Median      3Q     Max
-32.131  -8.539  -3.119   5.960 192.890

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -10.475      2.081  -5.034  5.6e-07 ***
educ           2.418      0.143  16.902  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.79 on 1093 degrees of freedom
Multiple R-squared:  0.2072,    Adjusted R-squared:  0.2065
F-statistic: 285.7 on 1 and 1093 DF,  p-value: < 2.2e-16
```

Q28

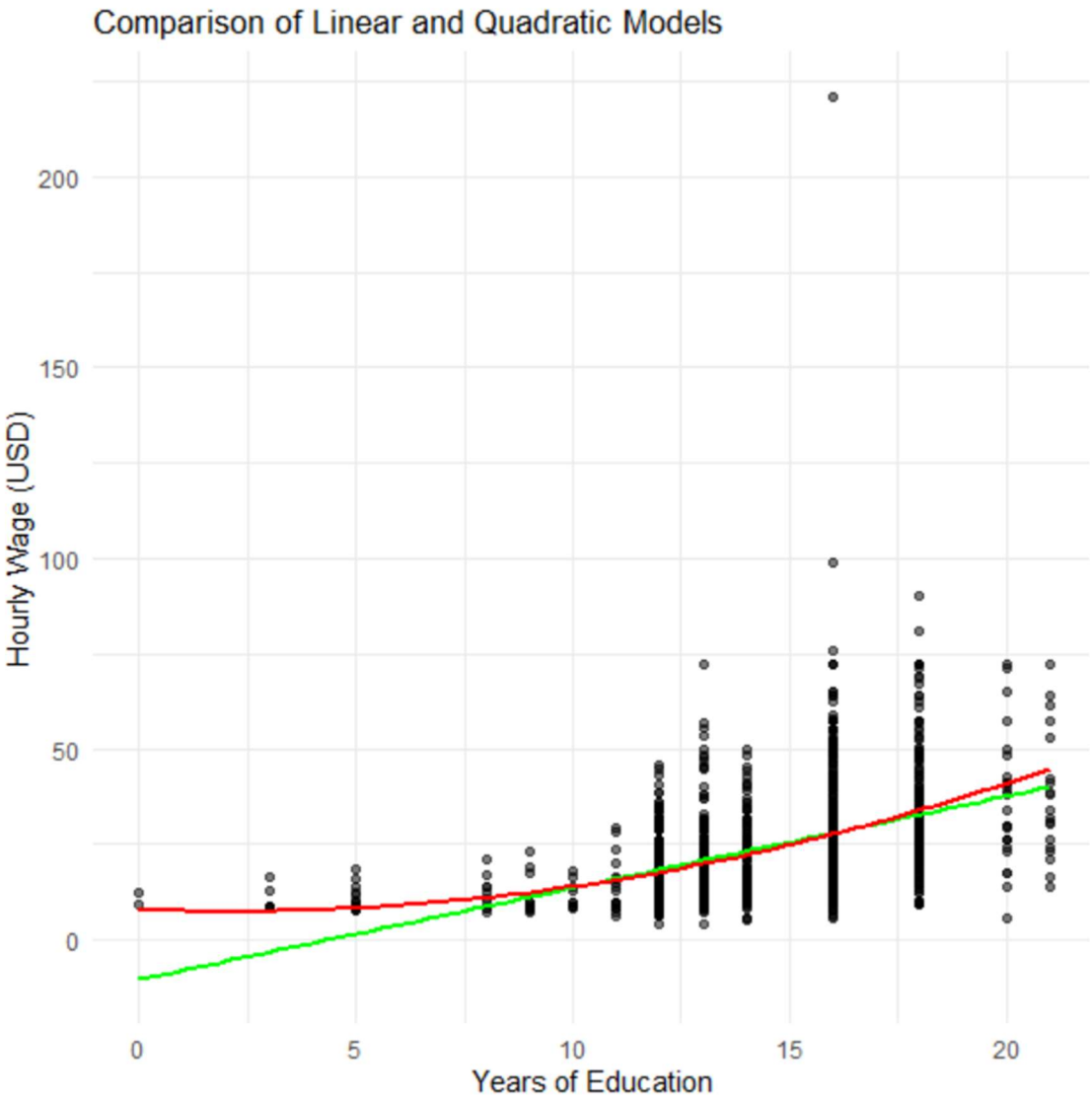
E

```
lm(formula = wage ~ I(educ^2), data = cps5_small)

Residuals:
    Min      1Q  Median      3Q     Max
-34.820  -8.117  -2.752   5.248 193.365

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.916477   1.091864   4.503 7.36e-06 ***
I(educ^2)   0.089134   0.004858  18.347  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.45 on 1198 degrees of freedom
Multiple R-squared:  0.2194,    Adjusted R-squared:  0.2187
F-statistic: 336.6 on 1 and 1198 DF,  p-value: < 2.2e-16
```

根據結果，受教育年資為 12 年，每增加一年，薪資就會增加約 2.1392 美元/小時。受教育年資為 16 年時，每年受教育年限的薪資增幅升至約 2.8523 美元/小時。與線性迴歸（預測每小時增加 2.40 美元）相比，二次模型表明，教育的好處會根據已經受教育的年數而變化。

Q28

F



二次模型比線性方程式更能捕捉數據變化，特別是在教育程度較低的情況下。