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.

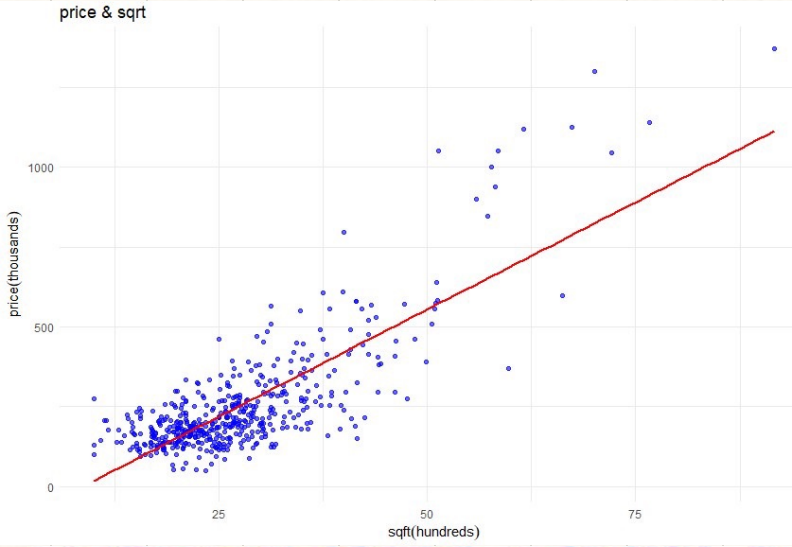
d. Graph the fitted curve for the model in part (c). On the graph, sketch the line that is tangent to the curve for a 2000-square-foot house.

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?

a. b.



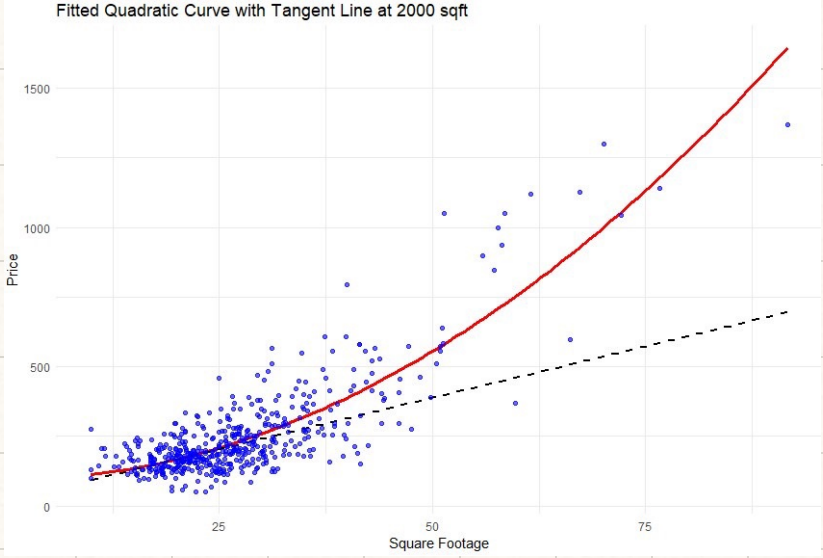

```
> cat("Intercept (Alpha):", coef(tab)[1], "\n")
Intercept (Alpha): -115.4236
> cat("Slope (Beta):", coef(tab)[2], "\n")
Slope (Beta): 13.40294
> print(results$summary)
```

c.

```
> # Output the marginal effect
> cat("The marginal effect of an additional 100 square feet at 2000 square feet is:", marginal_effect_2000, "\n")
The marginal effect of an additional 100 square feet at 2000 square feet is: 7.38076
```

(4)

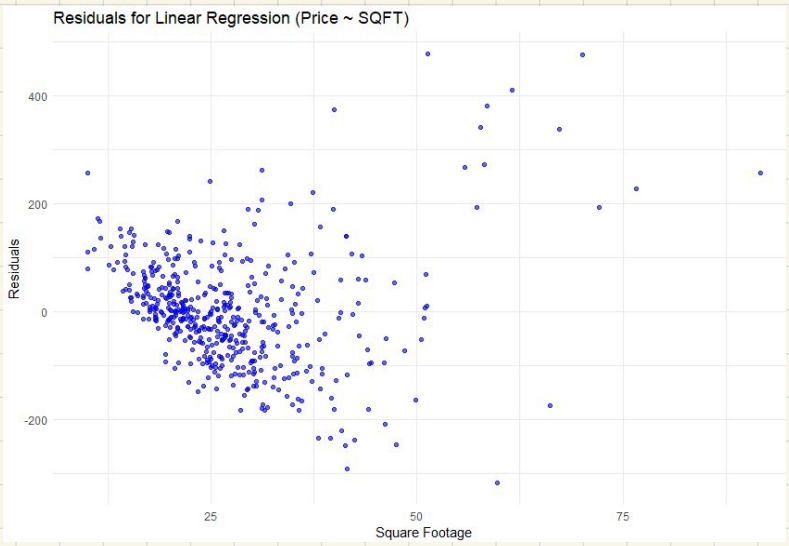
d.

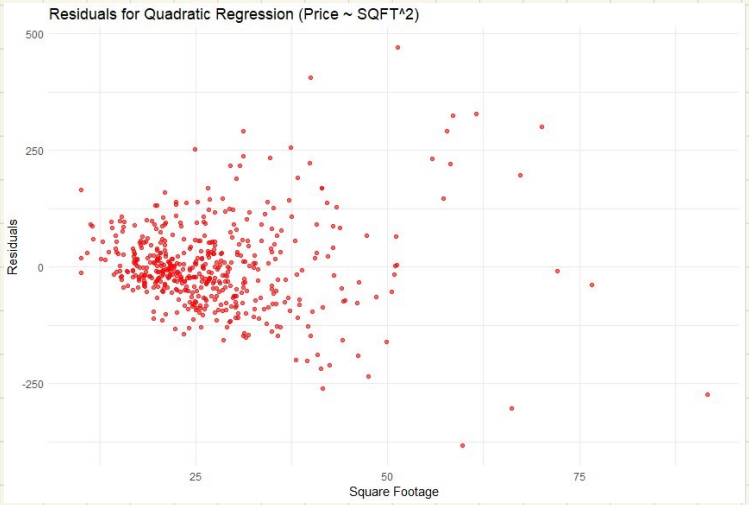


e.

```
> # Output the result
> cat("The elasticity of PRICE with respect to SQFT at 2000 square feet is:", elasticity_at_2000, "\n")
The elasticity of PRICE with respect to SQFT at 2000 square feet is: 0.8819511
```

f.





由圖可知 sqft ↑, 殘差↑,
不符合常數古差假設

g.

```
> # 4. 顯示結果
> cat("Model 1 (Linear) Residual Sum of Squares (RSS):", rss1, "\n")
Model 1 (Linear) Residual Sum of Squares (RSS): 5262847
> cat("Model 2 (Quadratic) Residual Sum of Squares (RSS):", rss2, "\n")
Model 2 (Quadratic) Residual Sum of Squares (RSS): 4222356
```

model in part (C) 較佳

$SSE(c) < SSE(b)$

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?

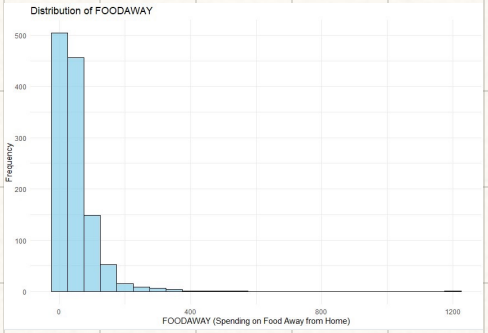
c. Construct a histogram of ln(*FOODAWAY*) and its summary statistics. Explain why *FOODAWAY* and ln(*FOODAWAY*) have different numbers of observations.

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?

a.




```
Mean (均值): 49.27085
Median (中位數): 32.555
25th Percentile: 12.04
75th Percentile: 67.5025
```

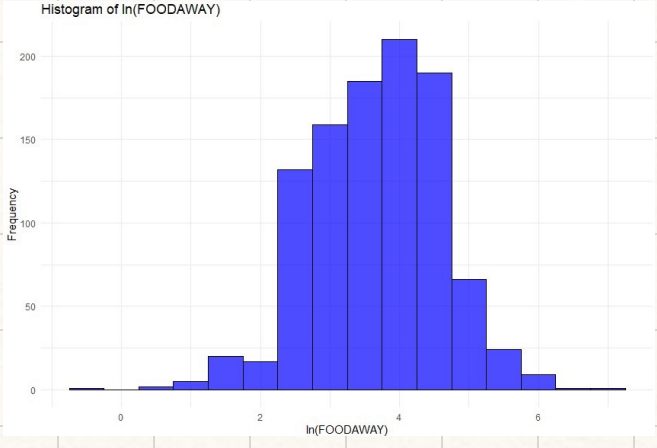
b.

	Category	Mean	Median
1	Advanced Degree	73.15494	48.15
2	College Degree	48.59718	36.11
3	No Degree	39.01017	26.02

c.

	Statistic	Value
1	Mean	3.650804
2	Median	3.686499
3	25th Percentile	3.075929
4	75th Percentile	4.279717

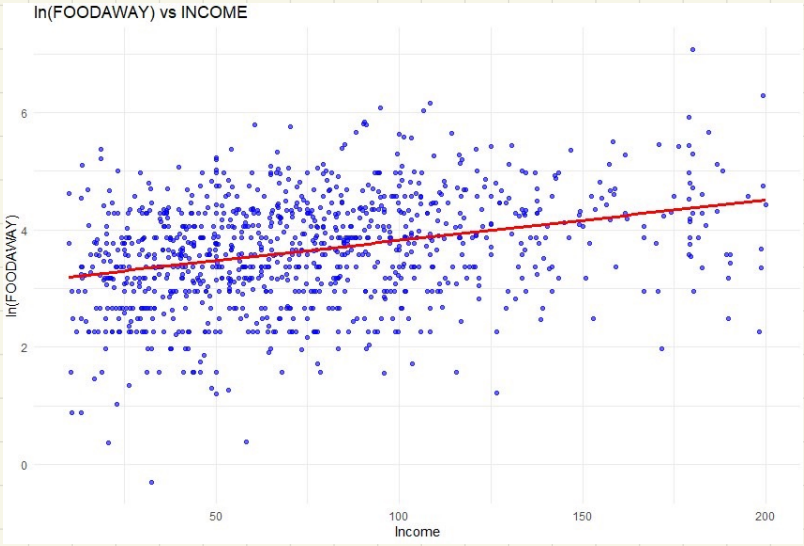
foodaway 有些為0
而ln0無定義，為缺失值



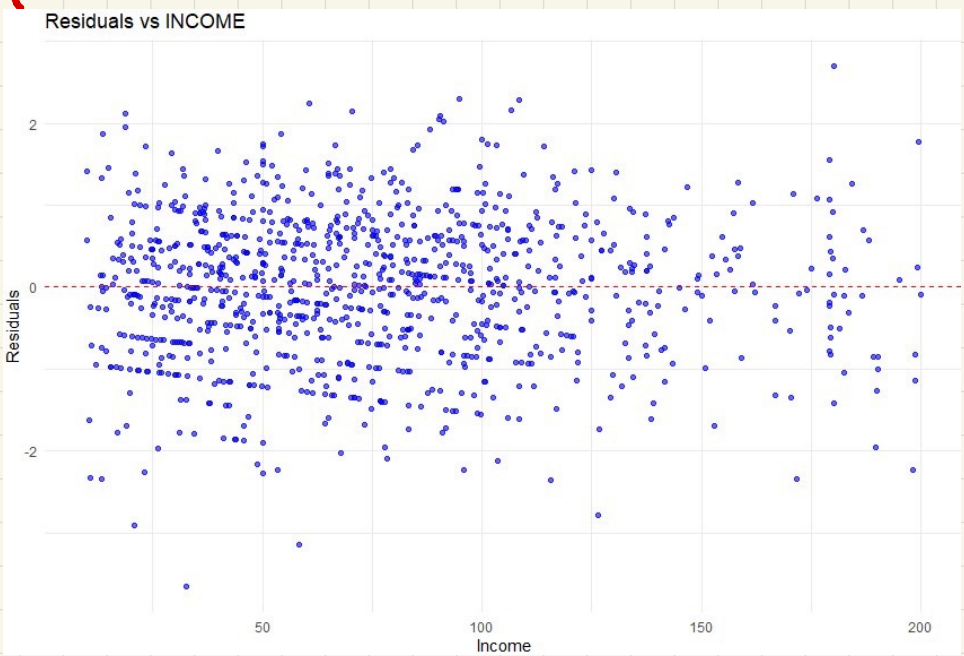
d.

```
> beta_2 <- coef(model)[2]
> cat("The slope is:", beta_2, "\n")
The slope is: 0.006901748
```

e.



f.



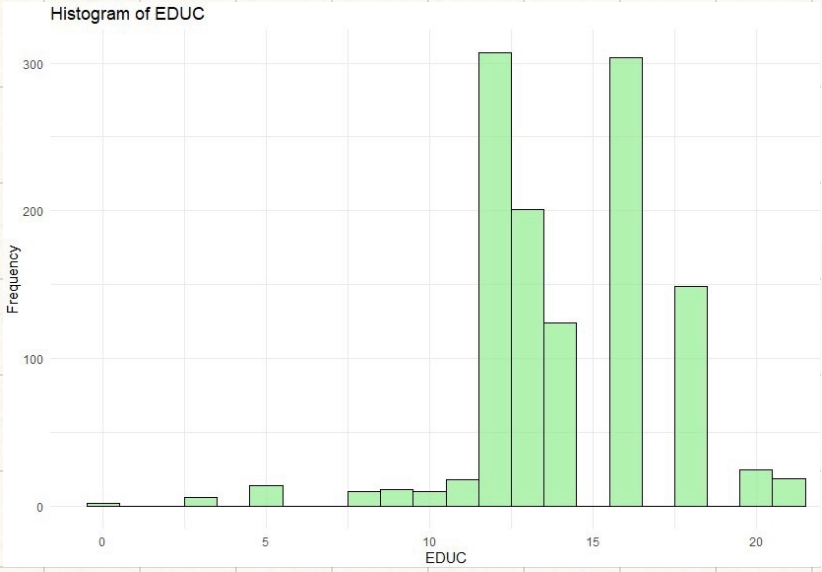
no unusual pattern, it seems random.

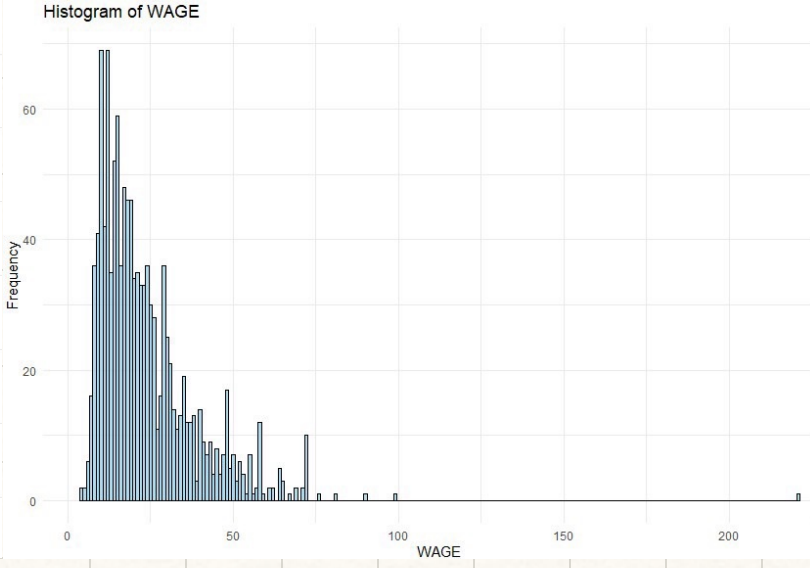
2.28 How much does education affect wage rates? The data file *cps5_small* contains 1200 observations on hourly wage rates, education, and other variables from the 2013 Current Population Survey (CPS). [Note: *cps5* is a larger version.]

- **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?

a.

```
警告： 正在使用 'POE5Rdata' 這個程式套件，因此不會被安裝
Summary Statistics for WAGE:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   3.94   13.00   19.30   23.64   29.80  221.10

Summary Statistics for EDUC:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0    12.0    14.0    14.2    16.0    21.0
```





多集中在 10~20 之間

大多數人工資較低，呈現右偏

b.

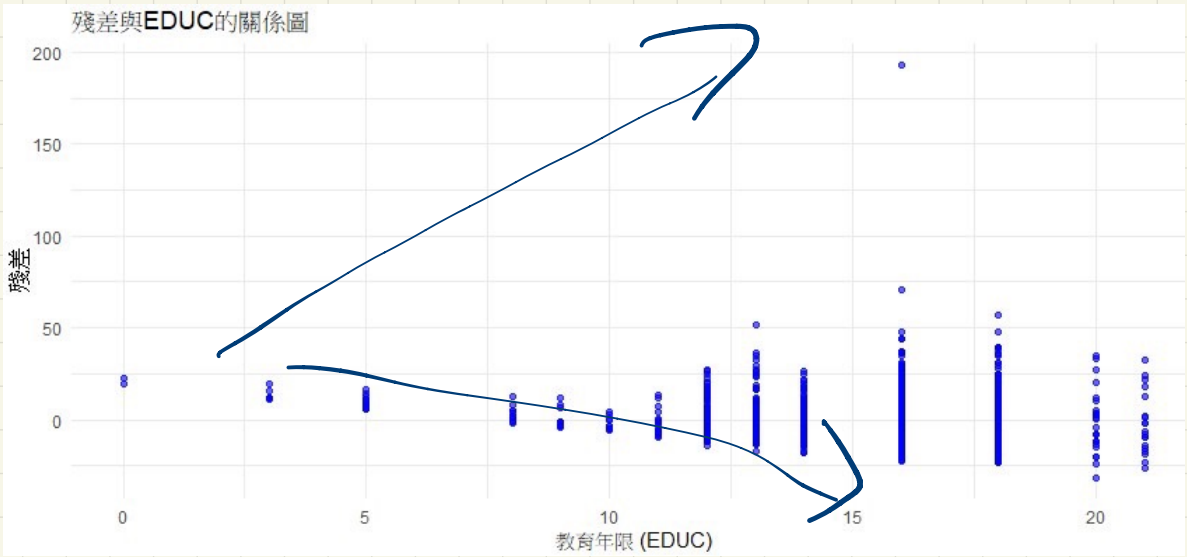
```
警告： 正在使用 'POE5Rdata' 這個程式套件，因此不會被安裝
截距 -10.39996
斜率 2.396761

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 ***
```

wage = -10.4 + 2.4 educ

when educ = 0 ⇒ wage = -10.4

educ ↑ 1 year ⇒ wage ↑ 2.4

c.



EDUC↑, residual↑, 不符合 Homoskedasticity (SR5)

d.

男性回歸結果:

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

女性回歸結果:

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

黑人回歸結果:

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

白人回歸結果:

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

女性教育回報最高(2.65)

黑人教育回報最低(1.92)

低教育水平下黑人薪資高:截距為-6.25

女性薪資低:截距為-16.6

e、

```
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
```

```
> cat("12年教育的邊際效應:", marginal_effect_12, "\n")
12年教育的邊際效應: 2.139216
> cat("16年教育的邊際效應:", marginal_effect_16, "\n")
16年教育的邊際效應: 2.852288
```

```
> cat("線性回歸的邊際效應:", beta_linear, "\n")
線性回歸的邊際效應: 2.396761
```

二次回歸隨EDUC↑, marginal effect↑

marginal effect = $2 \times \beta_2 \times EDUC$

↓

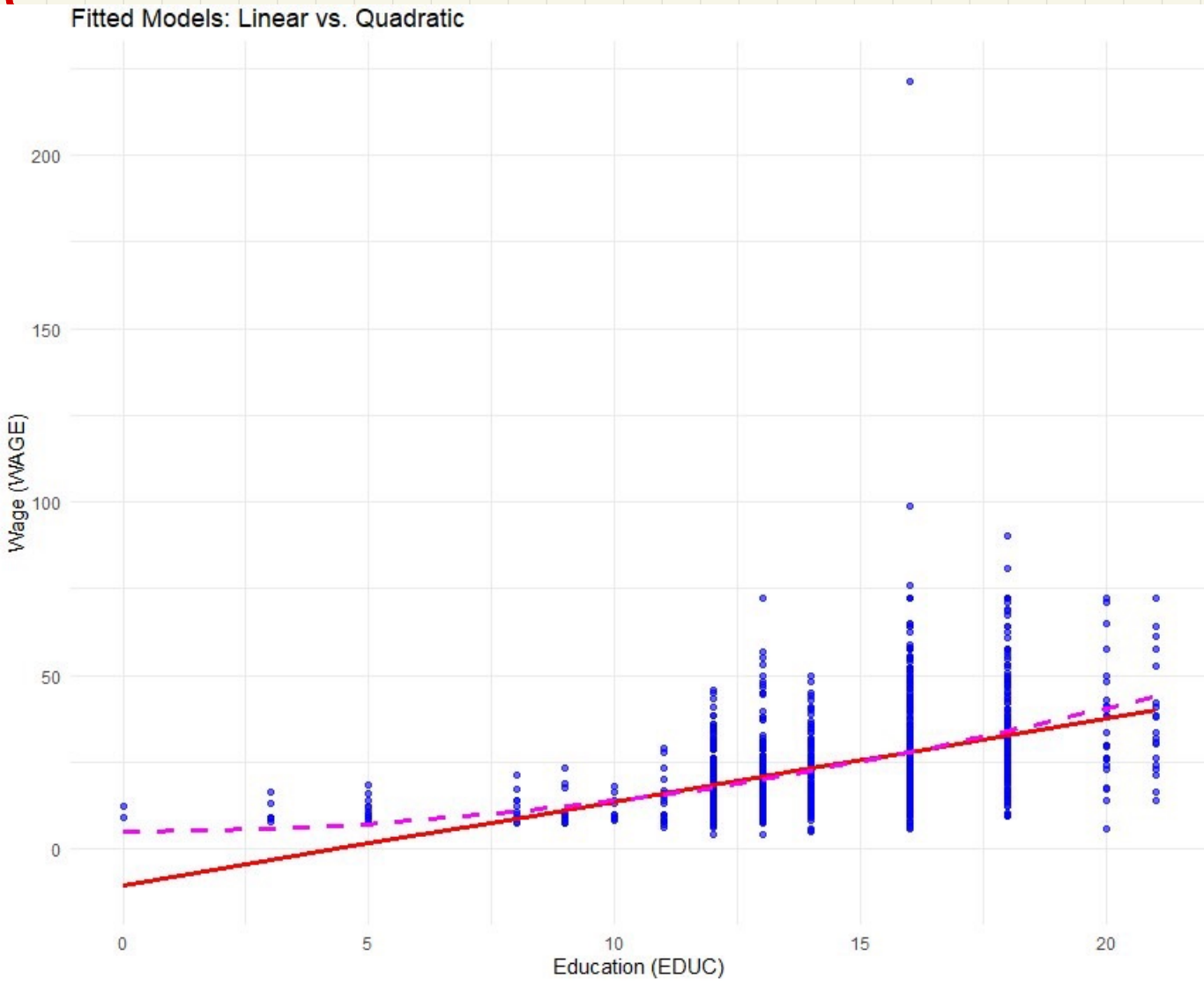
0.089134

線性回歸的marginal effect固定為β_2

↓

2.3967

f1



由图片左方可看出二次回歸擬合度較高

右方幾乎無差異

故選擇二次回歸較佳