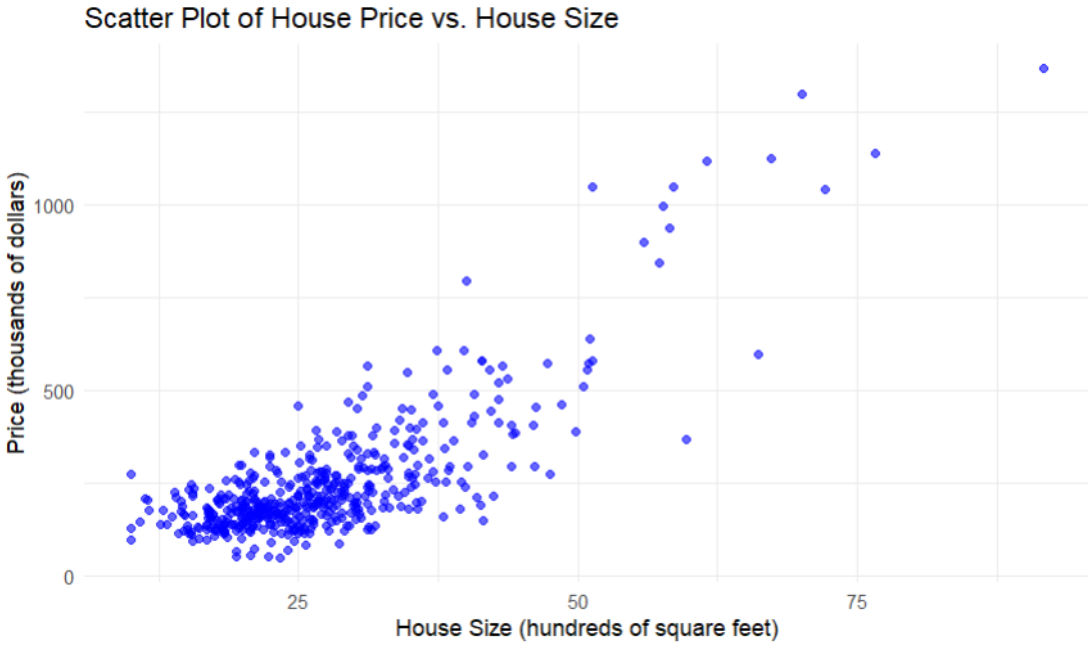
2.17 The data file *collegetown* contains observations on 500 single-family houses sold in Baton Rouge, Louisiana, during 2009–2013. The data include sale price (in thousands of dollars), *PRICE*, and total interior area of the house in hundreds of square feet, *SQFT*.

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

a.



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?

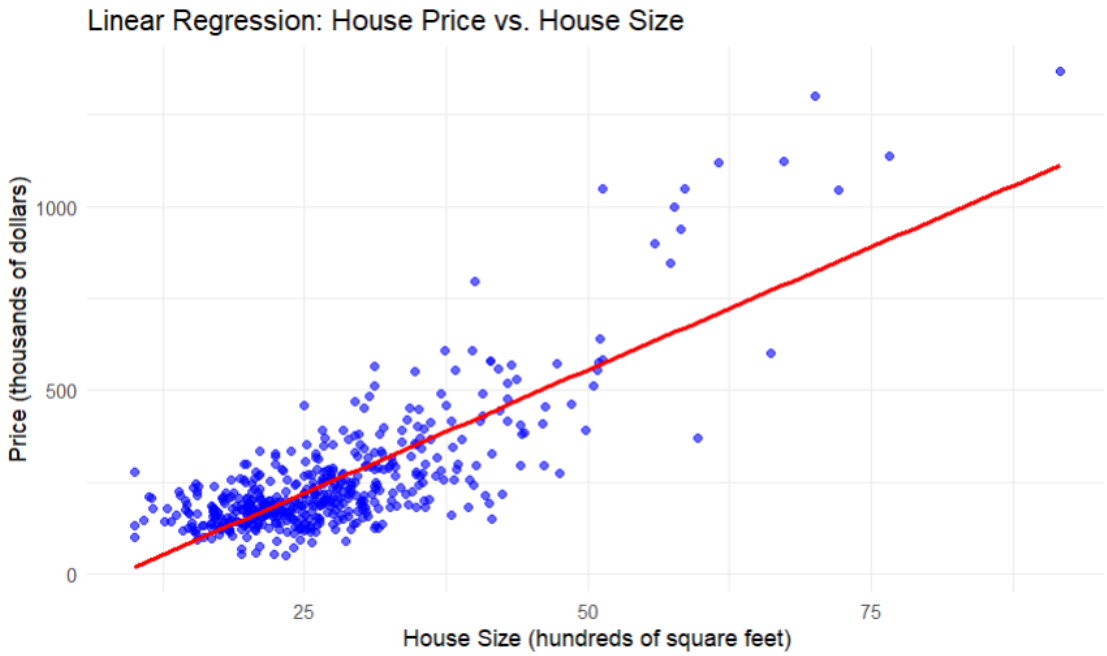
g. One basis for choosing between these two specifications is how well the data are fit by the model. Compare the sum of squared residuals (*SSE*) from the models in (b) and (c). Which model has a lower *SSE*? How does having a lower *SSE* indicate a "better-fitting" model?

b.

線性迴歸模型:PRICE= -115.4236+13.4029SQFT

β_1=控制其他變數不變下，以平均來說，當面積變動一單位，房價增加 13402.94 美元

α=面積為 0 的房子預期房價為-115423.6 美元

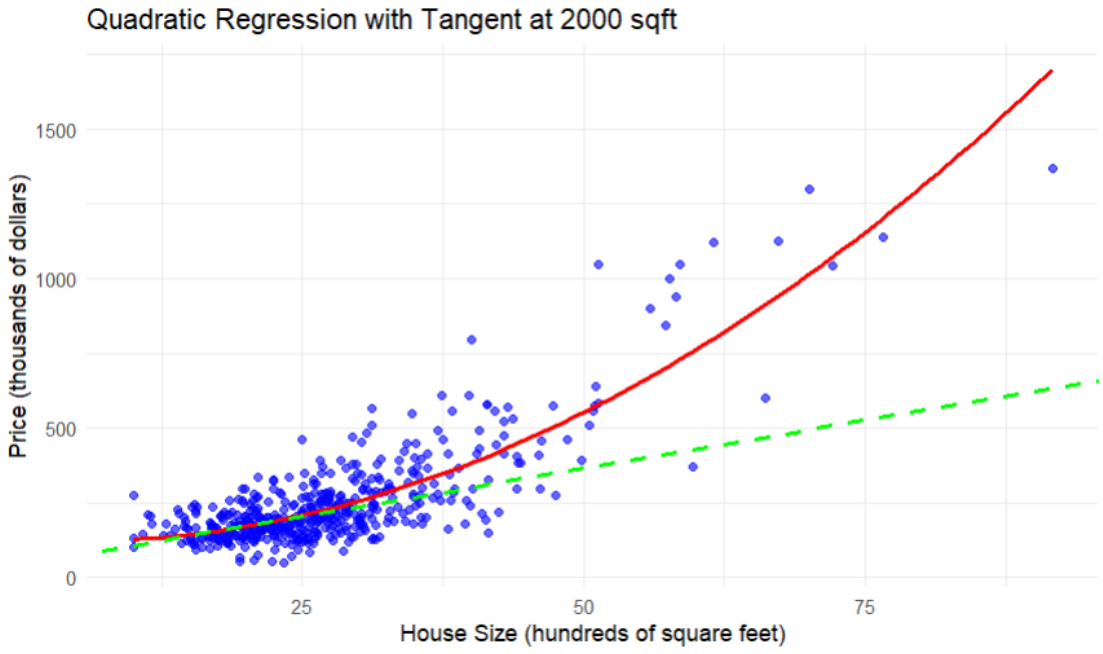


c.

線性迴歸模型:PRICE= 93.5659+0.1845SQFT^2

每增加 100 平方英尺的面積，以 2000 平方公尺的面積來說，在其他條件不變的情況下，預期房價的邊際效果將增加 $7,380.80。

d.



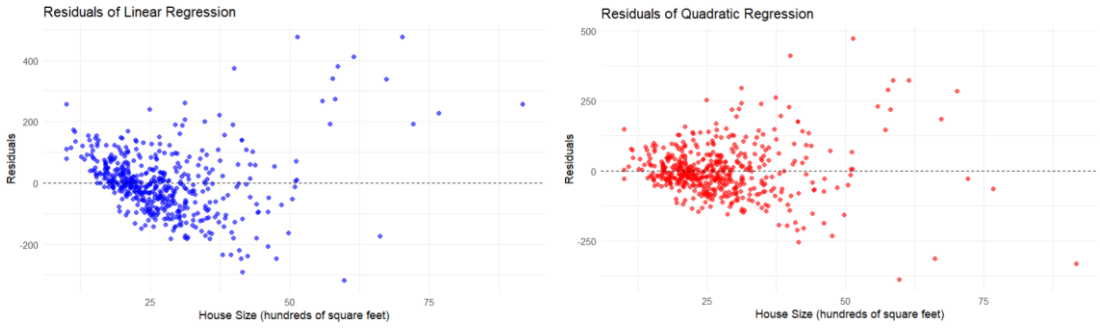
e.

$$\varepsilon = \frac{\partial p}{\partial s} \times \frac{s}{p}$$

$$= 2 \times 0.184519 \times 20 \times \frac{20}{p}$$

$$= 0.881951$$

f.



g.

The SSE linear model is 5,262,846.9. The SSE for the quadratic model is 4,222,356.3. In this case the quadratic model has the lower SSE. The lower SSE means that the data values are closer to the fitted line for the quadratic model than for the linear model.

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.

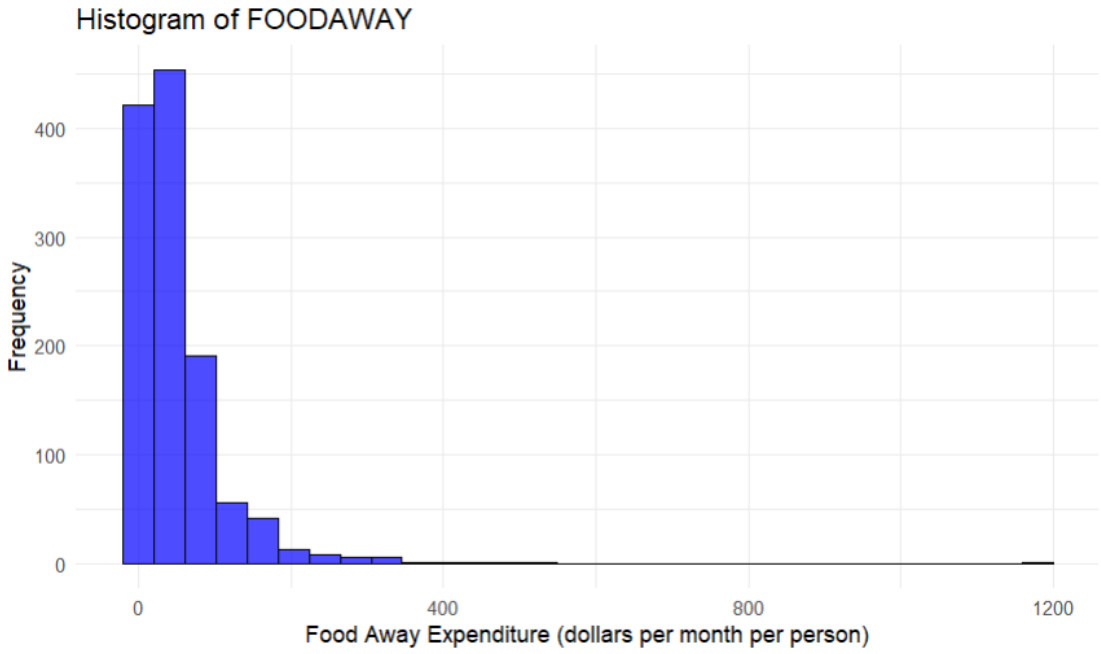
d. Estimate the linear regression $\ln(FOODAWAY) = \beta_1 + \beta_2 INCOME + e$. Interpret the estimated slope.

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.

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.00   12.04   32.55   49.27   67.50 1179.00
```

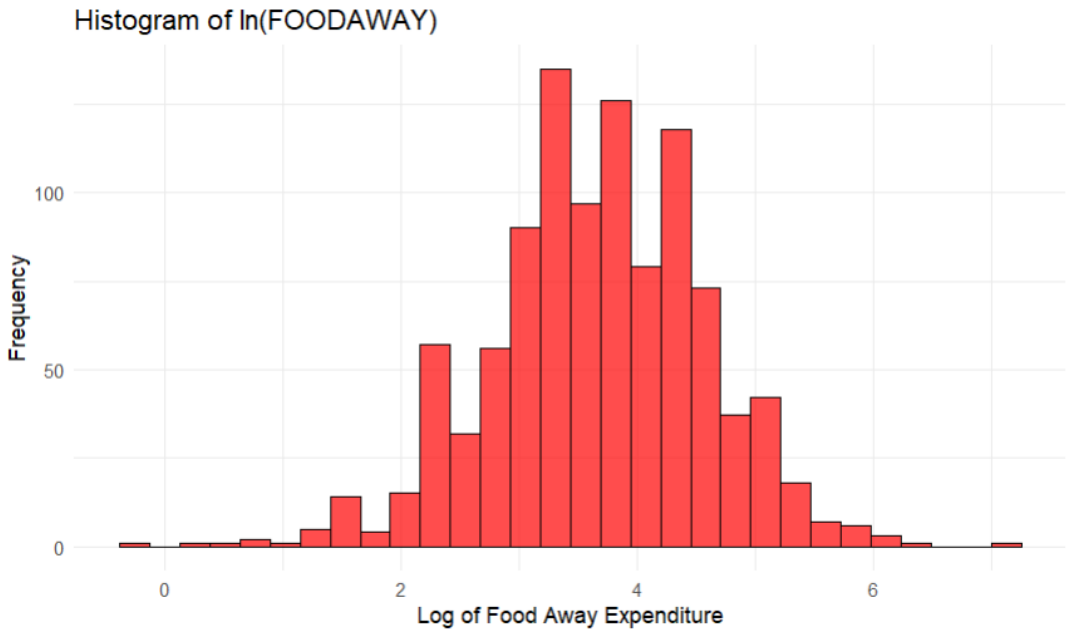


b.

education_group	count	mean_foodaway	median_foodaway
<chr>	<int>	<dbl>	<dbl>
Advanced Degree	257	73.2	48.2
College Degree	369	48.6	36.1
No College Degree	574	39.0	26.0

c.因有 178 個家庭在外飲食的支出為 0，導致有 178 個缺失值。

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.3011  3.0759  3.6865  3.6508  4.2797  7.0724
```

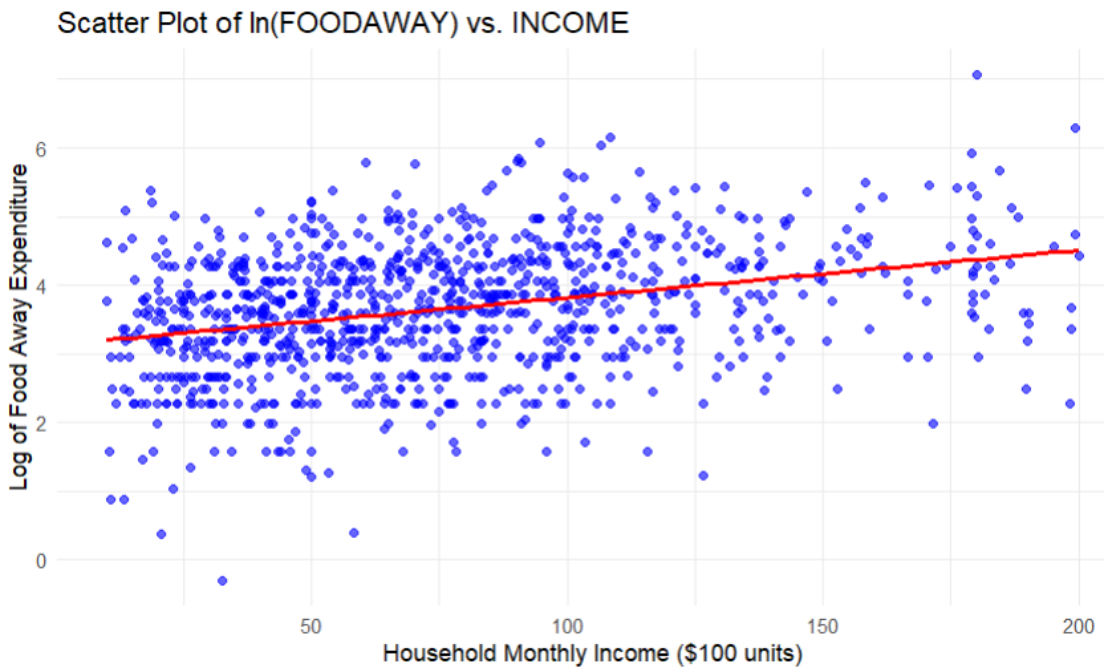


d.

ln(FOODAWAY)=3.1293+0.0069×INCOME

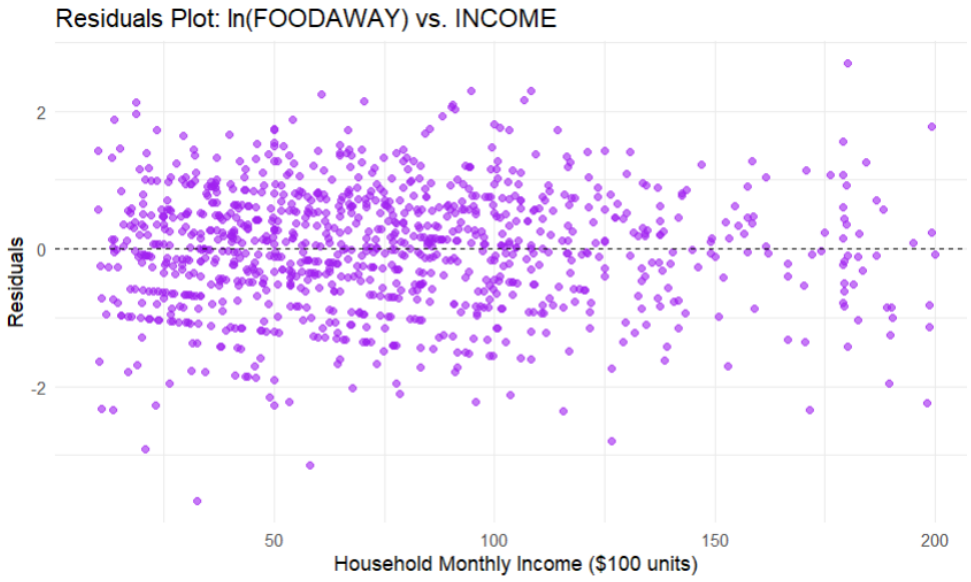
每增加$100 的收入，預計外食支出會增加 0.69%，假設其他變數不變下。

e.



f.

SSE=5262846.9471，沒特別的分布

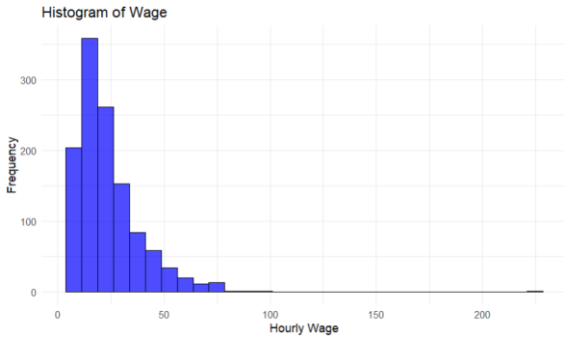


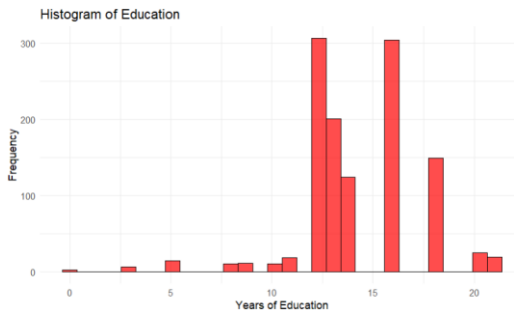
2.28 How much does education affect wage rates? The data file *cps5_small* contains 1200 observations on hourly wage rates, education, and other variables from the 2013 Current Population Survey (CPS). [Note: *cps5* is a larger version.]

- **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.

```
summary(cps5_small$wage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   3.94   13.00   19.30   23.64   29.80  221.10
summary(cps5_small$educ)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0    12.0    14.0    14.2    16.0    21.0
```





wage 呈現右偏分佈，大部分的時薪集中在 $0 - $50 區間內，極端高薪的人數極少，可能代表少數高收入者。educ 並不是均勻分佈，而是集中在特定年數。

b.

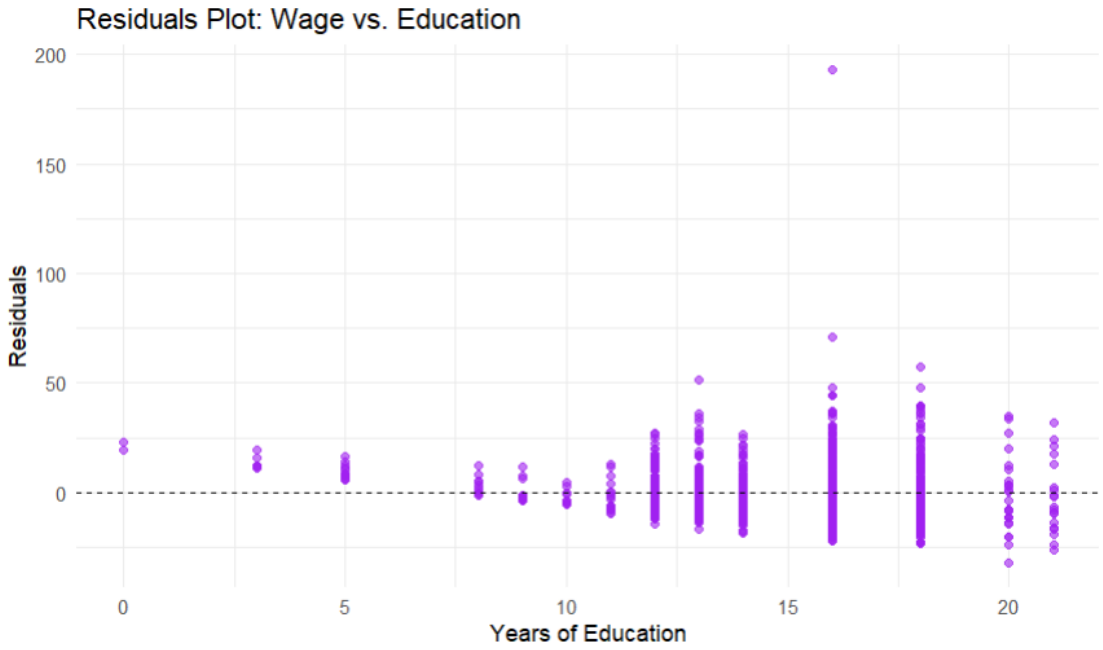
wage=−10.4+2.3968×educ

β_1 ：每增加 1 年教育，預計時薪增加約 2.40 美元，其他條件不變的情況下。

α ：當 educ = 0 (即沒有受過教育時)，預測工資為 -10.4，但這樣的討論無實際意義，因為教育年數不可能為 0。

c.

隨著教育年數增加，殘差越大，代表模型具有異質性，違反誤差項需為同質性的假設。



d.

male：wage^=−8.2849+2.3785×educ

female：wage^=−16.6028+2.6595×educ

black：wage^=−6.2541+1.9233×educ

white：wage^=−10.475+2.4180×educ

1.教育對女性的薪資影響最大

教育每增加 1 年，女性的時薪增加 $2.66，是所有群體中最高的。

2.黑人群體的教育回報最低

每多 1 年教育，黑人時薪僅增加 $1.92，遠低於其他群體。

e.

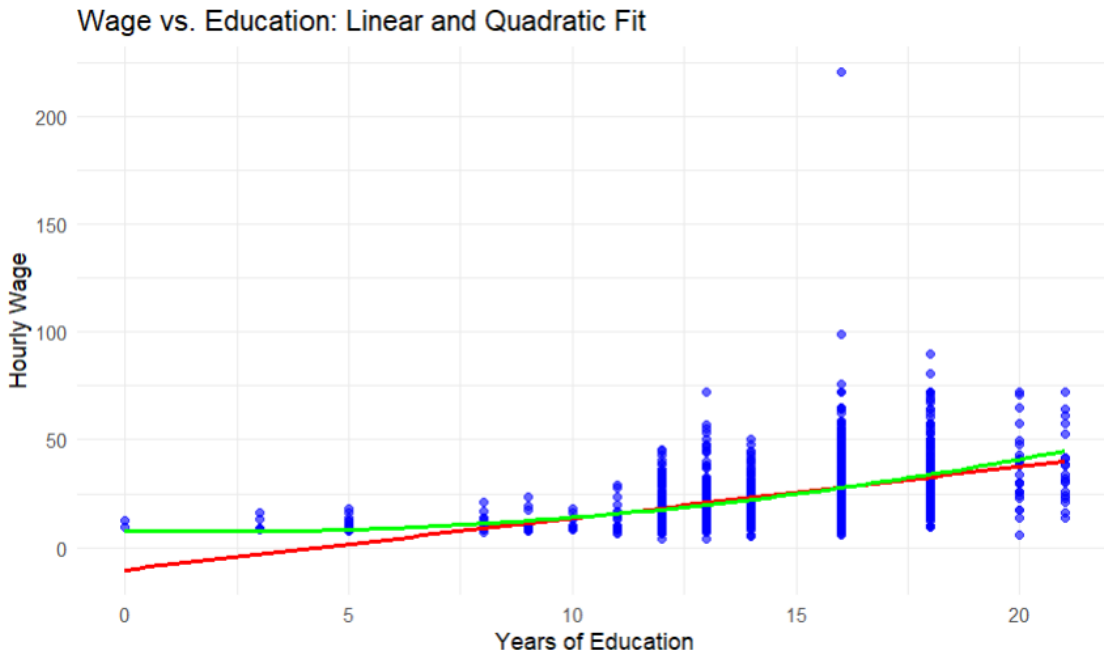
wage^=4.9165+0.0891×educ^2

> marginal_effect_12 <- 2 * 0.0891 * edu_12= 2.1392162355261

> marginal_effect_16 <- 2 * 0.0891 * edu_16= 2.8522883140348

教育對薪資的影響是遞增的，隨著受教育年數增加，薪資的增長幅度加大，而非如 b 小題所預估的線性增加。

f.



線性模型低估了高學歷者的薪資增長，而二次模型能更好地捕捉這種邊際效應。