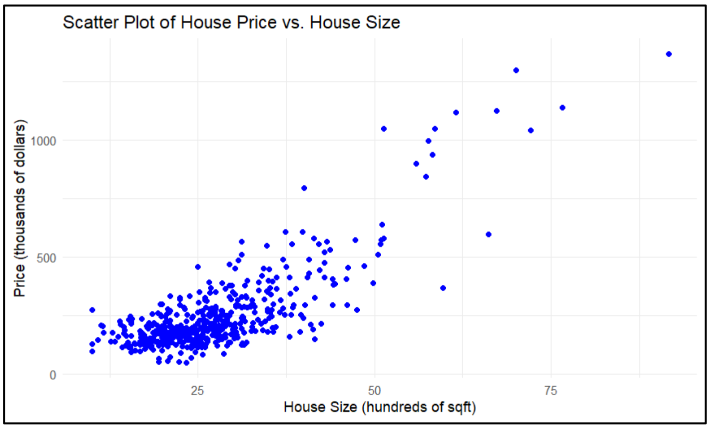
HW0303

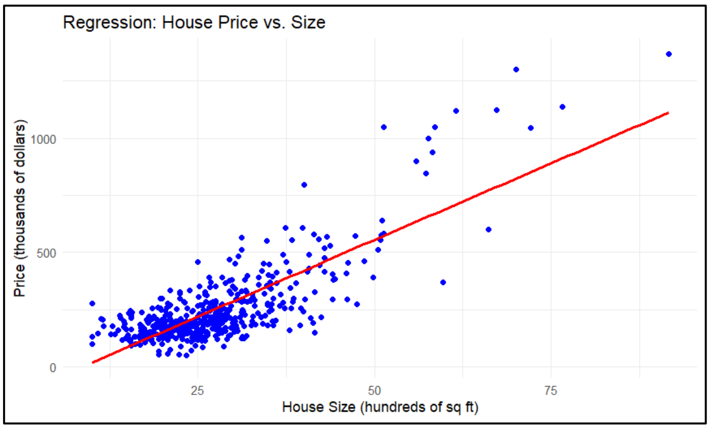
2.17 (a)



2.17 (b)

PRICE = -115.4236 + 13.4029 * SQFT

With other conditions unchanged, when the house size increases by 100 square feet, the house price will increase by 13.4029 thousand. When the house size equals zero, the house price will become -115.4236 thousand.



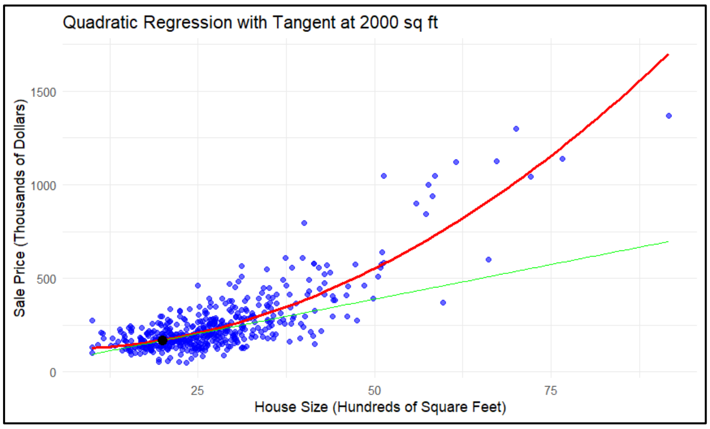
2.17 (c)

PRICE = 93.5659 + 0.1845 * SQFT^2

Margin effect = 2*0.1845*20 = 7.3808

When an additional 100 square feet of living area is added to a home with 2000 square feet of living space, the house price will increase by 7.3808 thousand.

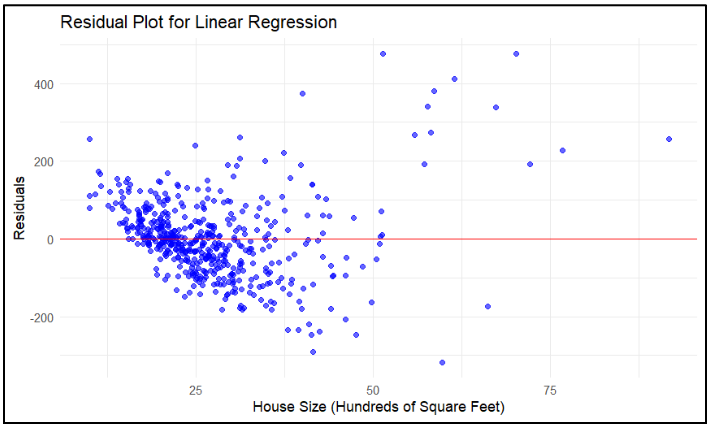
2.17 (d)

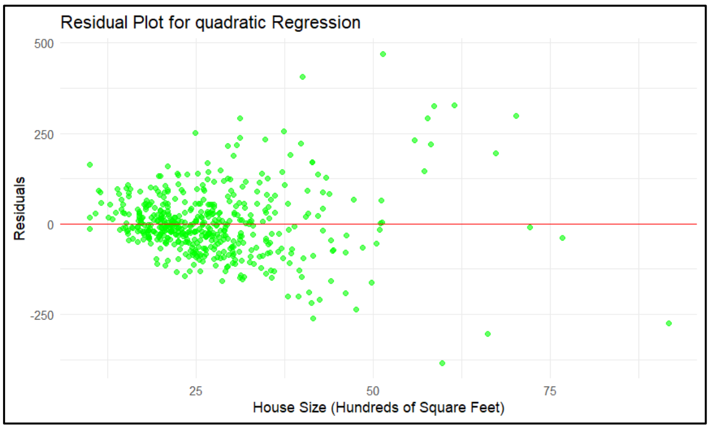


2.17 (e)

Elasticity = 0.8819511

2.17 (f)





The errors increase as SQFT increases, which violates the homoscedasticity assumption.

2.17 (g)

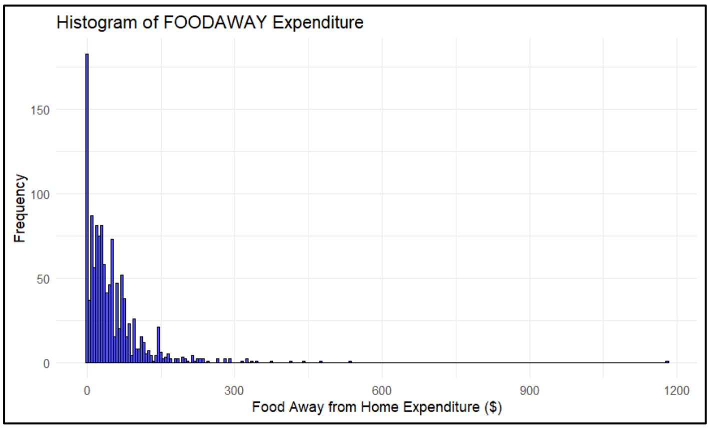
SSE of linear model：5262847

SSE of quadratic model：4222356

平方項的回歸模型 SSE 較小，表示平方項的回歸模型比線性模型擬合數據的效果更好，可能原因是它考慮了線性模型無法捕捉到的非線性關係。

2.25 (a)

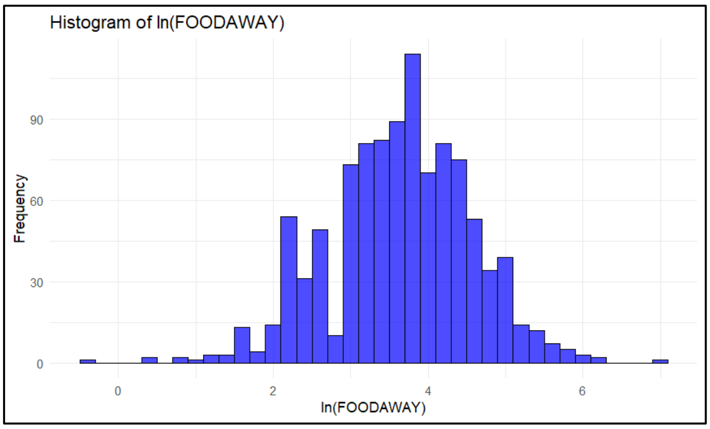
```
> summary(cex5_small$foodaway)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.00   12.04   32.55   49.27   67.50 1179.00 
```



2.25 (b)

```
Table: Summary of FOODAWAY by Education Level

|Group           | Numbers| Mean_Foodaway| Median_Foodaway|
|:---------------|-------:|-------------:|---------------:|
|Advanced Degree |     257|      73.15494|           48.15|
|College Degree  |     369|      48.59718|           36.11|
|No Degree       |     574|      39.01017|           26.02|
```

2.25 (c)




```
> print(summary_stats)
   mean_ln median_ln    q25_ln    q75_ln Numbers_ln Numbers_original
1 3.650804  3.686499 3.075929 4.279717       1022             1200
```

2.25 (d)

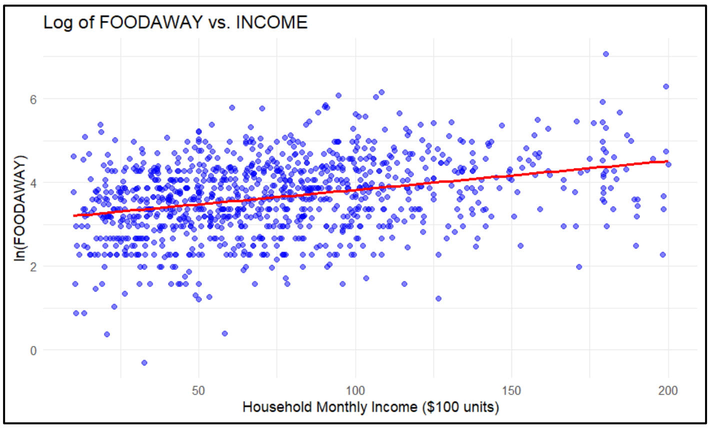
```
Call:
lm(formula = ln_foodaway ~ income, data = cex5_small)

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
  (因為不存在，178 個觀察量被刪除了)
Multiple R-squared:  0.09826,	Adjusted R-squared:  0.09738 
F-statistic: 111.1 on 1 and 1020 DF,  p-value: < 2.2e-16
```

Regression equation: ln(FOODAWAY) = 3.1293 + 0.0069 * INCOME

When household income is zero, the predicted value of ln(FOODAWAY) is 3.1293.

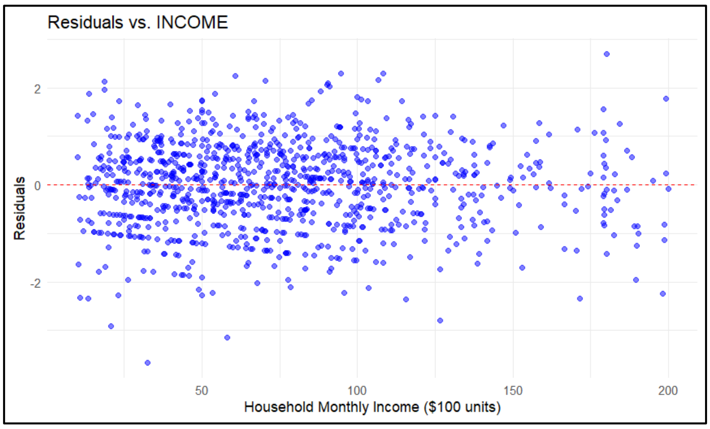
For every $100 increase in household income, FOODAWAY increases by 0.69%.

2.25 (e)



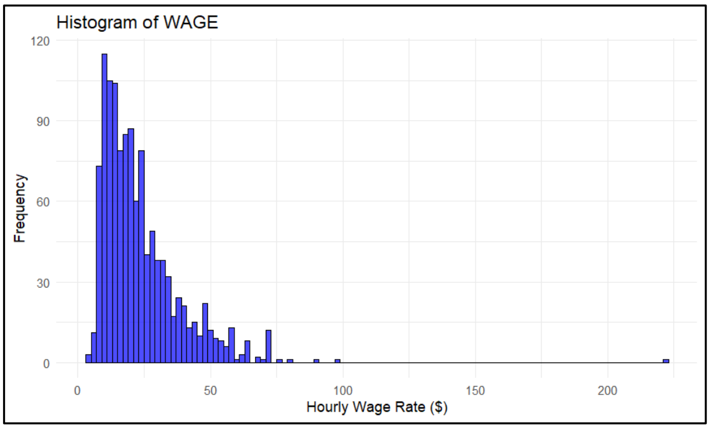
Income and ln(FOODAWAY) are positively related.

2.25 (f)



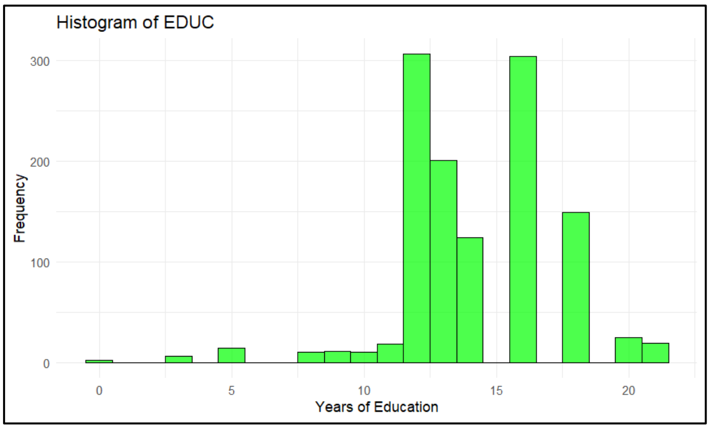
There is no obvious pattern or systematic trend in the plot of residuals versus income. So, the residuals appear to be randomly distributed.

2.28 (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 
```

從直方圖可以看到薪水呈現右尾分布，多數人的薪水相對低。




```
> summary(cps5_small$educ)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    0.0    12.0    14.0    14.2    16.0    21.0
```

從圖中可以觀察到多數人的教育年數落在 12 年到 16 年之間，表示多數人皆有高中及大學學歷。

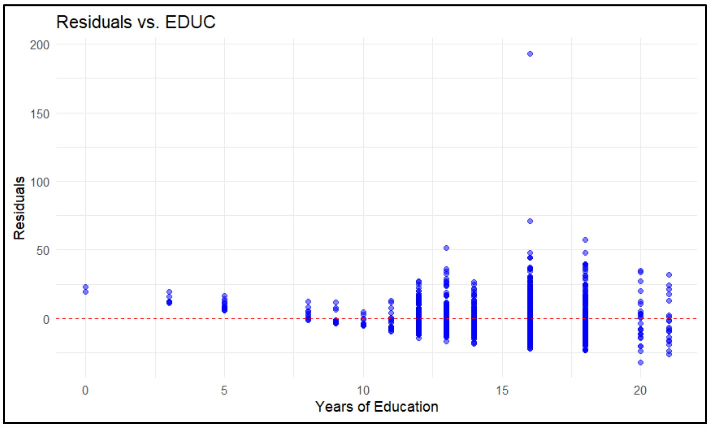
2.28 (b)

```
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
```

WAGE = -10.4 + 2.3968 * EDUC

When an individual has zero years of education, his predicted hourly wage is -10.4. And each additional year of education increases the hourly wage by $1.20 on average.

2.28 (c)



The residuals increase as years of education increase, indicating that the homos cedasticity assumption of the model may be violated.

2.28 (d)

```
> summary(model_male)

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
```

```
> summary(model_female)

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
```

```
> summary(model_black)

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
```

```
> summary(model_white)

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
```

Male Model: WAGE = -8.2849 + 2.3785 * EDUC

Female Model: WAGE = -16.6028 + 2.6595 * EDUC

Black Model: WAGE = -6.2541 + 1.9233 * EDUC

White Model: WAGE = -10.4747 + 2.4178 * EDUC

Male Model v.s Female Model

- 女性模型的截距項比男性模型低，代表在低教育水準下，女性的起薪低於男性。
- 女性模型的斜率大於男性模型的斜率，表示教育對女性的薪資影響程度較大，因此女性可以透過高等教育而獲得更好的工作機會減少性別帶來的薪資差距。

Black Model v.s White Model

- 黑人模型的截距項較高，與事實不符，回歸模型的結果可能受到其他變數影響。
- 白人模型的斜率高於黑人模型的斜率，表示教育對白人的薪資影響更大，而黑人的薪資成長速度較慢，這反映了種族差異導致的薪資歧視。

2.28 (e)

```
> summary(model_quadratic)

Call:
lm(formula = wage ~ EDUC2, 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 ***
EDUC2       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
```

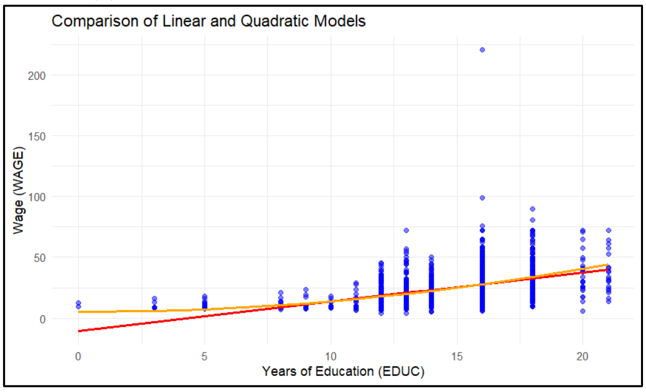
WAGE = 4.9165 + 0.0891 * EDUC^2

Marginal Effect （EDUC = 12）：2.1392

Marginal Effect（EDUC = 16）：2.8523

在（b）小題中的線性模型的邊際效果是固定的，而這裡的邊際效果是會隨教育年數的不同而改變，當教育年數越高時邊際效果會越高。

2.28 (f)



橘色線比紅色線更貼近數據的分布，表示二次模型更能擬合數據，因為平方項能捕捉到非線性效果。