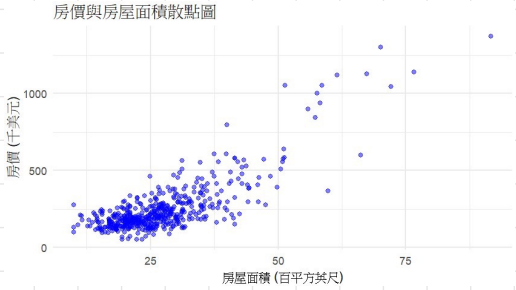
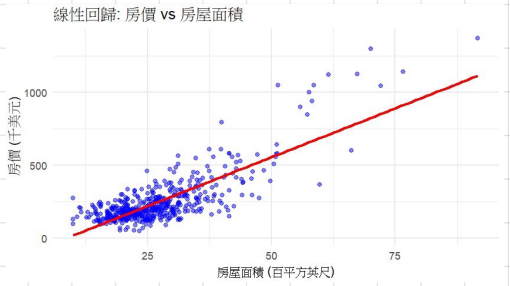
2.17

(a)



(b)



$\widehat{price} = -115.42 + 13.40\ SQFT + e$

說明 增加 100 square feet ⇒ price 上升 13.40 (千)

```
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 ***
```

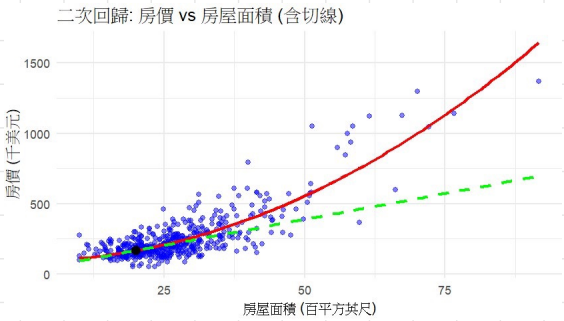
(c) (d)

```
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 ***
sqft2        0.184519   0.005256   35.11   <2e-16 ***

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

>
> sqft_2000 = 20
> mar_eff = coef(quad_model)["sqft2"] * 2 *sqft_2000
> print(paste("Marginal Effect at 2000 sqft:", round(mar_eff, 2),'
lars"))
[1] "Marginal Effect at 2000 sqft: 7.38 thousands of dollars"
```



$\hat{P} = 93.57 + 0.18 SQFT^2 + e$

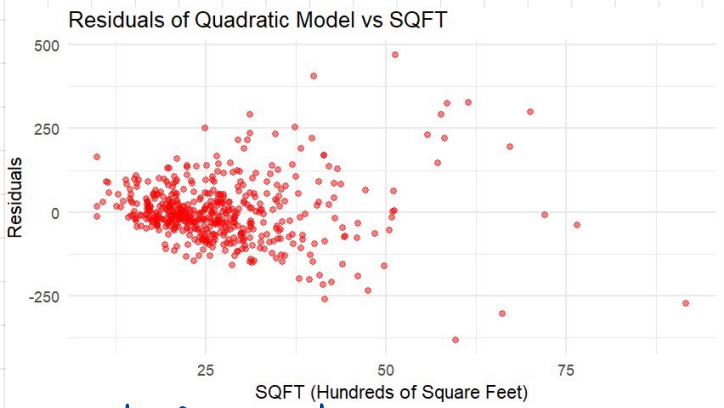
$\frac{\partial E(\hat{P}|SQFT)}{\partial SQFT}$

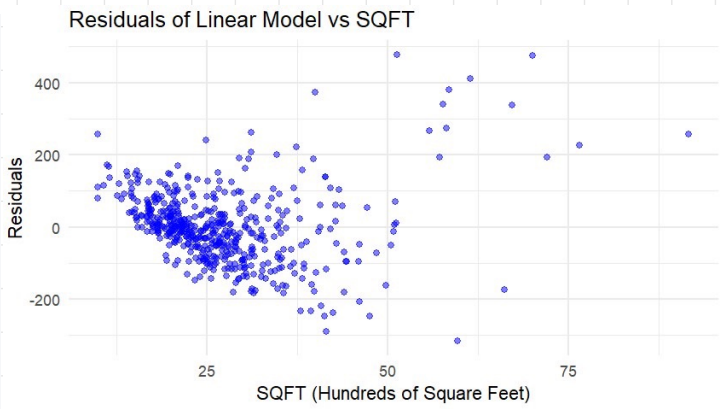
$= 2 \times 0.1845 \times 20 = 7.381$

(e)

```
[1] "Elasticity of PRICE with respect to SQFT at 2000 sqft: 0.882"
```

(f)





Both spread of residuals look like not random

⇒ the homoscedasticity assump is vidated

(g)

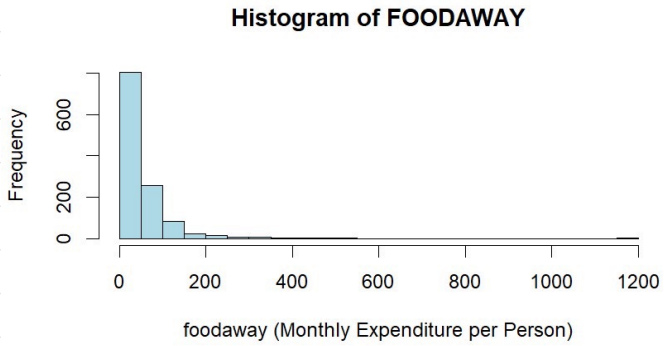
```
> print(paste("SSE for Linear Model:", round(linear_sse, 2)))
[1] "SSE for Linear Model: 5262846.95"
> print(paste("SSE for Quadratic Model:", round(quadratic_sse, 2)))
[1] "SSE for Quadratic Model: 4222356.35"
```

The model has lower SSE

⇒ fits the data more with less error value.

2.25

(a)

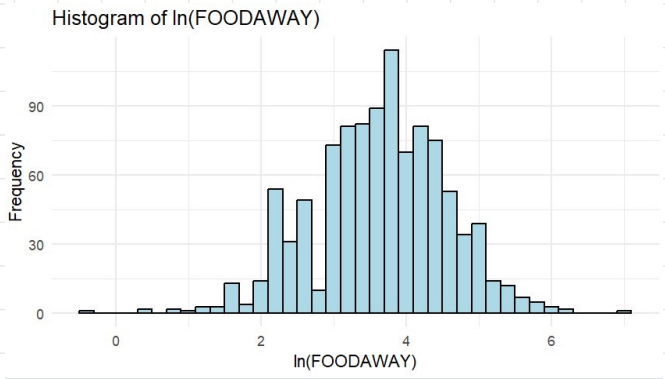



```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00   12.04   32.55   49.27   67.50 1179.00
> cat("Mean:", mean_foodaway, "\n")
Mean: 49.27085
> cat("Median:", median_foodaway, "\n")
Median: 32.555
> cat("25th percentile:", quantiles_foodaway[1], "\n")
25th percentile: 12.04
> cat("75th percentile:", quantiles_foodaway[2], "\n")
75th percentile: 67.5025
```

(b)

```
Mean and Median FOODAWAY:
Advanced Degree (advanced = 1) - Mean: 73.15494  Median: 48.15
College Degree but No Advanced (college = 1 & advanced = 0) - Mean: 48.59718  Medi
an: 36.11
No College or Advanced Degree (college = 0 & advanced = 0) - Mean: 39.01017  Media
n: 26.02
```

(c)




```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.3011  3.0759  3.6865  3.6508  4.2797  7.0724
> cat("Mean of ln(FOODAWAY):", mean_ln_foodaway, "\n")
Mean of ln(FOODAWAY): 3.650804
> cat("Median of ln(FOODAWAY):", median_ln_foodaway, "\n")
Median of ln(FOODAWAY): 3.686499
> cat("25th percentile of ln(FOODAWAY):", quantiles_ln_foodaway[1], "\n")
25th percentile of ln(FOODAWAY): 3.075929
> cat("75th percentile of ln(FOODAWAY):", quantiles_ln_foodaway[2], "\n")
75th percentile of ln(FOODAWAY): 4.279717
>
```

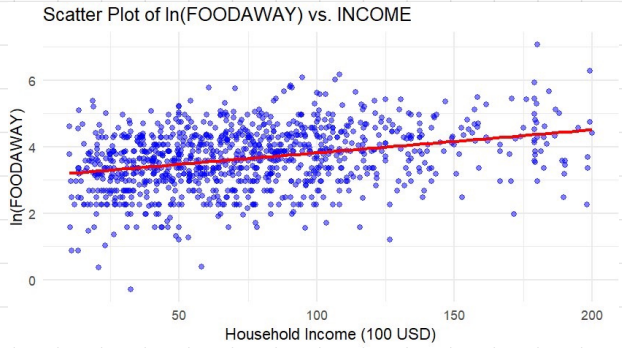
Number of observations of ln(FOODWAY) < 1.98.
⇒ the 198 datas ≦ 0

(d)

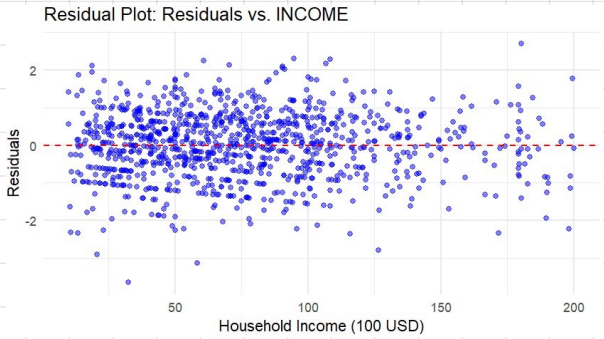
```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.1293004  0.0565503   55.34   <2e-16 ***
income      0.0069017  0.0006546   10.54   <2e-16 ***
```

$\ln(FOODWAY) = 3.1293 + 0.0069\,INCOME + e$

(e)



(f)

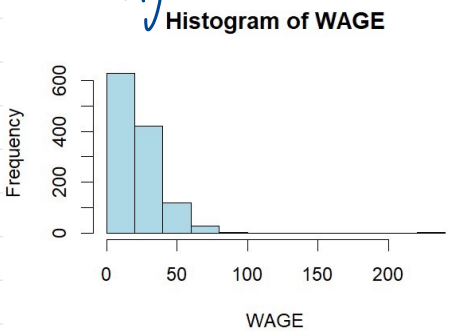


Residuals 沒有明顯的 pattern.
⇒ 有可能是 random distribution.

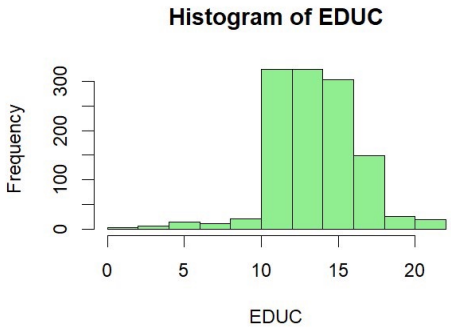
2.28

(a)

right skewness



left skewness or clustering around EDUC levels.




```
> 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
```

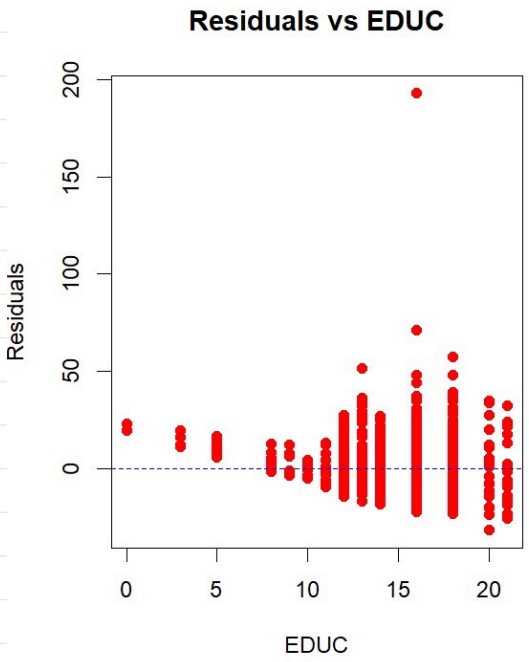
(b)

```
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 ***
```

$\widehat{WAGE} = -10.4 + 2.3968 EDUC + e$

(c)



It shows significant heterogeneous variability and the possibility of nonlinear relationship.

⇒ violating to SR1, SR5

If it doesn't violate, it should be random distribution.

(d)

Males

```
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
```

Females

```
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
```

Blacks

```
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
```

Not Blacks

```
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
```

教育回報最高：Females（但起薪低）

教育回報最低：Blacks（R^2也最低）

⇒學歷不是Blacks影響Wage的主要原因

(e)

```
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

>
> educ12 = 12
> educ16 = 16
> me_educ12 = 2*coef(quad_modle)[2]*educ12
> me_educ16 = 2*coef(quad_modle)[2]*educ16
> me_educ12
   educ2
2.139216
> me_educ16
   educ2
2.852288
```

EDUC increase ⇒ ME increase

quadratic regression is more emphasize the effect on wage from the increasing in education.

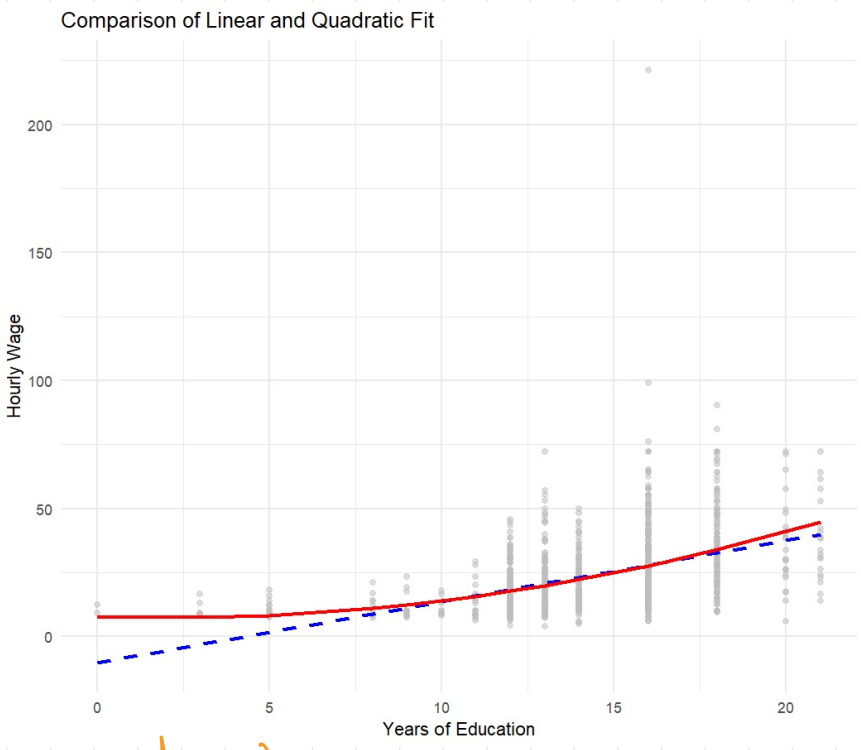
$\widehat{Wage} = 4.916477 + 0.089134\,EDUC^2 + e$

The ME is better when EDUC = 16

$2 \times 0.089134 \times 12 = 2.139216$

$2 \times 0.089134 \times 16 = 2.852288$

(f)



more fit with wage for high-education level (EDUC > 15)

(red one)

The quadratic line is more fit with data.

⇒ closer to the real value when EDUC < 10