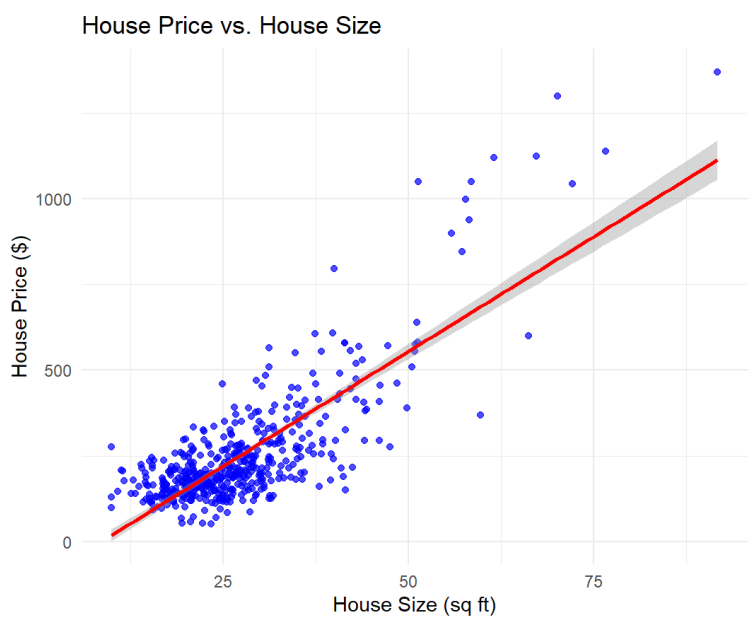


2.17(a)、(b)



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 ***

$$\widehat{PRICE} = -115.4236 + 13.40294SQFT$$

因此可以知道線性回歸為： (se) (13.0882) (0.4492)

2.17(c)

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 ***

$$\widehat{PRICE} = 93.5659 + 0.1845SQFT^2$$

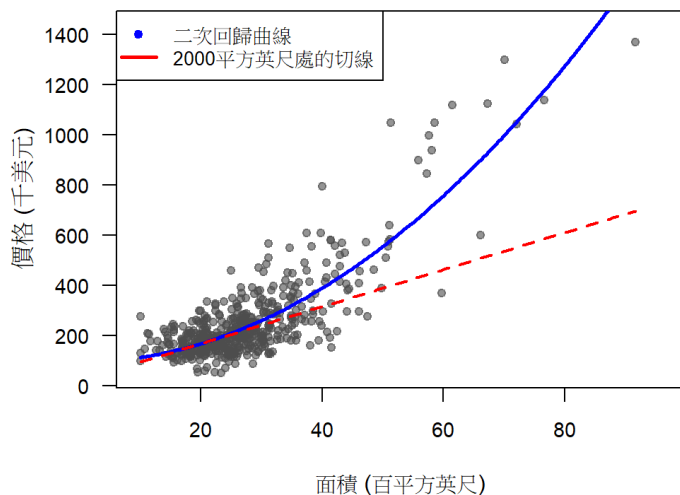
而二次回歸為： (se) (6.0722) (0.00525)

邊際效應計算結果：

```
> cat("在2000平方英尺的房屋中，增加100平方英尺的邊際效應 =",
+     round(marginal_effect, 4), "千美元 = $",
+     format(round(marginal_effect * 1000, 1), nsmall = 1), "\n")
在2000平方英尺的房屋中，增加100平方英尺的邊際效應 = 7.3808 千美元 = $ 7380.8
```

2.17(d)

房屋價格與面積的二次回歸模型



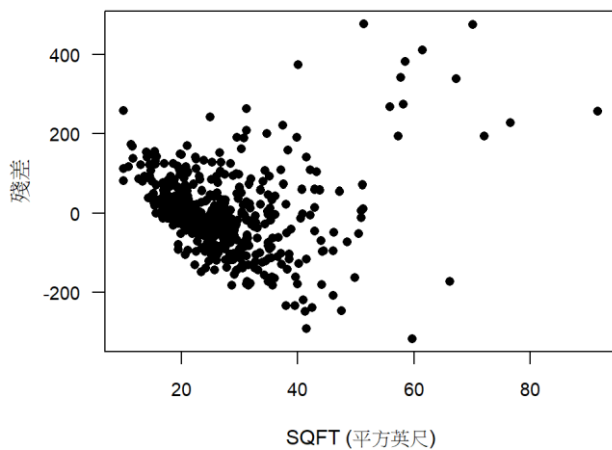
2.18(e)

彈性計算結果：

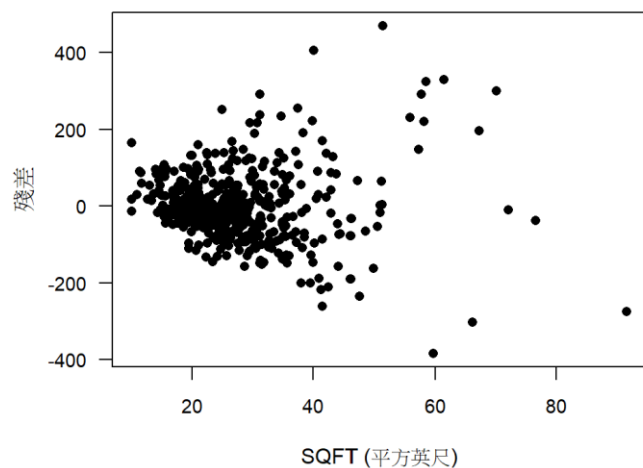
```
> cat("在2000平方英尺的房屋中，房價對面積的彈性 =", round(elasticity, 4), "\n")  
在2000平方英尺的房屋中，房價對面積的彈性 = 0.882
```

2.18(f)

線性模型殘差 vs SQFT



二次模型殘差 vs SQFT



在這兩個模型中，殘差模式都不是隨機出現的。殘差的變化隨著 SQFT 的增加而增加，表明可能違反同方差假設。

2.18(g)

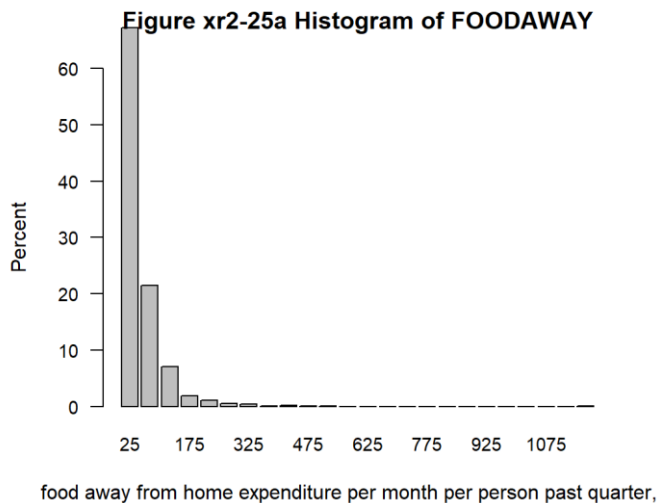
線性模型 (mod1) 的 SSE: 5262847

```
> cat("二次模型 (mod_quad) 的 SSE:", sse_mod_quad, "\n")
```

二次模型 (mod_quad) 的 SSE: 4222356

SSE 較低意味著二次模型的資料值比線性模型的資料值更接近擬合線。

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)

education_category	N	MEAN	MEDIAN
<chr>	<int>	<dbl>	<dbl>
None	574	39.0	26.0
advanced = 1	257	73.2	48.2
college = 1	369	48.6	36.1

2.25(c)

```
foodaway 觀察值數量: 1200
> cat("ln_foodaway 觀察值數量:", n_ln_foodaway, "\n")
ln_foodaway 觀察值數量: 1022
> cat("foodaway 為零的觀察值數量:", n_zeros, "\n")
foodaway 為零的觀察值數量: 178
> cat("因取對數而移除的觀察值數量:", n_foodaway - n_ln_foodaway, "\n\n")
因取對數而移除的觀察值數量: 178
```

$\ln(0)$ 未定義。它創建了軟體無法在迴歸中使用的「缺失值」

2.25(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 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8761 on 1020 degrees of freedom

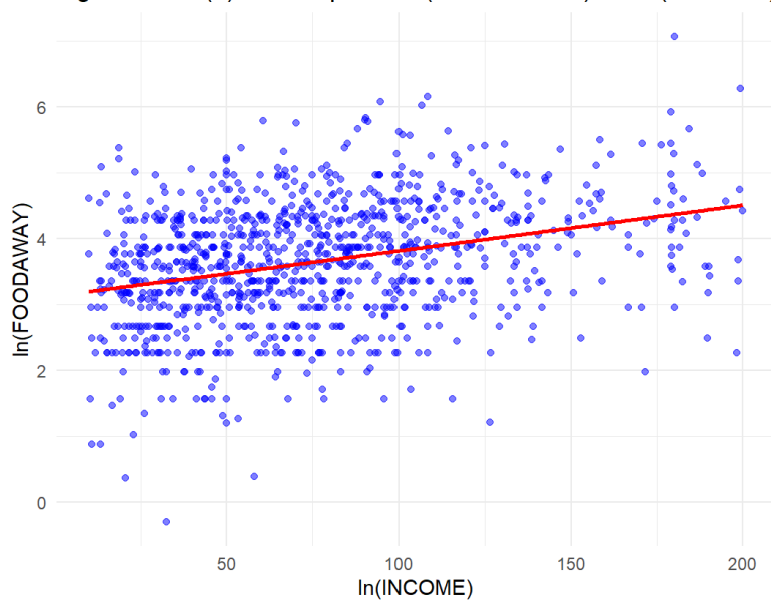
(因為不存在 178 個觀察量被刪除了)

Multiple R-squared: 0.09826, Adjusted R-squared: 0.09738

F-statistic: 111.1 on 1 and 1020 DF, p-value: < 2.2e-16

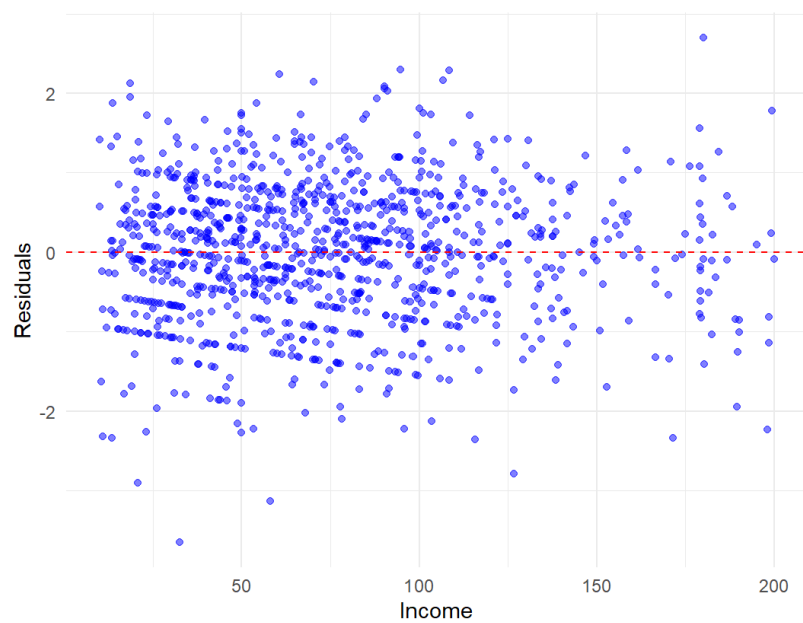
2.25(e)

Figure xr2.25(d) Scatter plot of $\ln(\text{FOODAWAY})$ vs. $\ln(\text{INCOME})$



2.25(f)

Residuals vs. Income



OLS 殘差確實隨機分佈，沒有明顯的模式。收入較高的觀察較少，因此有較多的「空白」。

2.28(a)

```
> print(summary_wage)
      N      Mean  Median      SD  Min  Max
1 1200 23.64004   19.3 15.21655  3.94 221.1
> print(summary_educ)
      N      Mean  Median      SD  Min  Max
1 1200 14.2025    14  2.890811    0   21
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
```

2.28(b)

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

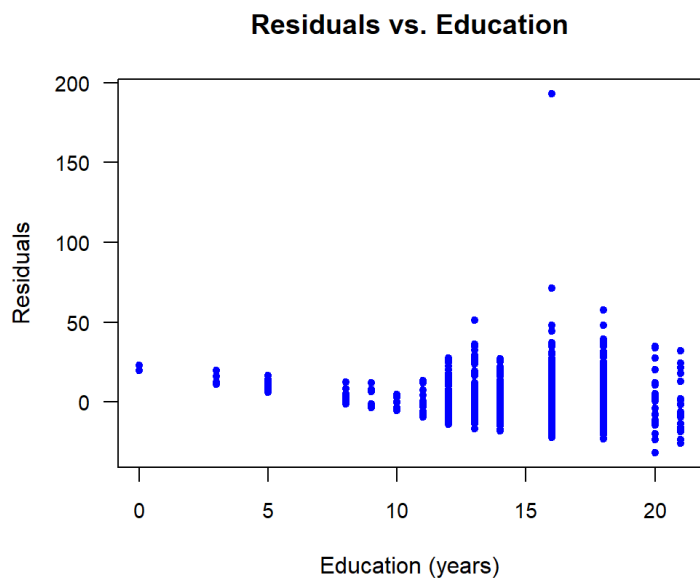
Residuals:

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 ***

2.28(c)



殘差在教育年數較高時（大約 15-20 年）出現較大的變異，有一個異常點在教育 15 年時殘差約 200，殘差的分佈範圍隨教育年數增加而變大，可能代表異質性

這可能違反 殘差變異數恆定性，因為殘差在較高的教育年數範圍內較為分散。

表示回歸模型沒有完全捕捉到教育與應變數之間的關係

2.28(d)

```
> summary(mod1_black)

Call:
lm(formula = wage ~ educ, data = cps5_black)

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

Coefficients:
(Intercept) -6.2541  1.9233
educ         1.9233  0.3983  4.829  4.79e-06 ***

> summary(mod1_female)

Call:
lm(formula = wage ~ educ, data = cps5_female)

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

Coefficients:
(Intercept) -16.6028  2.6595
educ         2.6595  0.1876 14.174 < 2e-16 ***

> summary(mod1_male)

Call:
lm(formula = wage ~ educ, data = cps5_male)

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

Coefficients:
(Intercept) -8.2849  2.3785
educ         2.3785  0.1881 12.648 < 2e-16 ***
```

2.28(e)

```
> cat("只有平方項的模型中，12年教育的邊際效應:", round(me_12, 4), "\n")
只有平方項的模型中，12年教育的邊際效應: 2.1392
> cat("只有平方項的模型中，16年教育的邊際效應:", round(me_16, 4), "\n")
只有平方項的模型中，16年教育的邊際效應: 2.8523
> cat("線性模型中的邊際效應:", round(me_linear, 4), "\n")
線性模型中的邊際效應: 2.3968
```

2.28(f)

二次模型可以更好的擬和我們的誤差項

工資與教育關係：線性模型 vs. 二次模型

