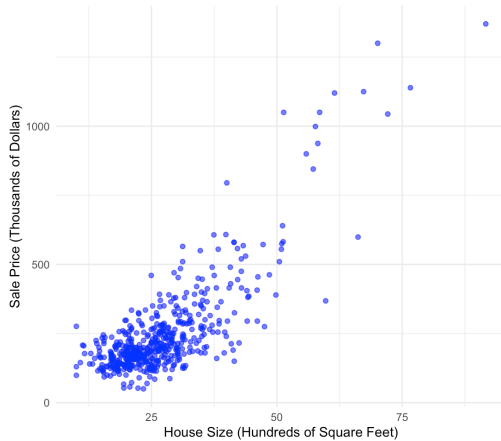
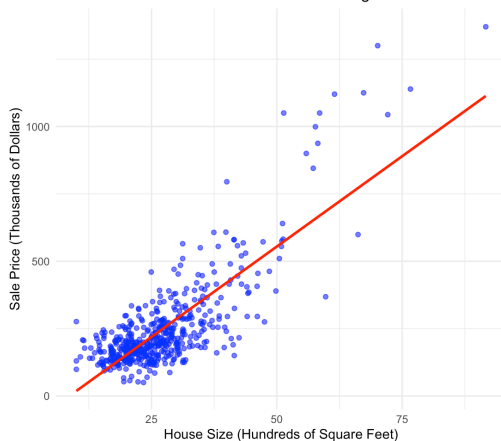


Q 2.17**a.**

Scatter Plot of House Price vs. House Size

**b.** 回歸方程式： $PRICE = -115.4236 + 13.4029 * SQFT$

House Price vs. House Size with Fitted Regression Line



當其他條件不變時，每增加 100 平方英尺的居住面積，預期房價將上升 13,402.94 美元。
估計的截距為 -115,423.60，代表一棟面積為零的房屋之預期價格將為 -115,423.60 美元。

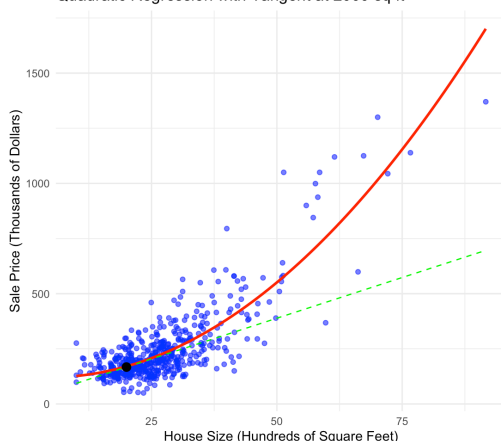
c. $PRICE = 93.5659 + 0.1845 * SQFT^2$

$$\text{marginal effect} = 2 \times 0.1845 \times 20 = 7.38 \text{ (千元)}$$

在房屋總面積為2000平方英尺的情況，額外增加100平方英尺的居住空間，預期房價上升 \$7,380.80

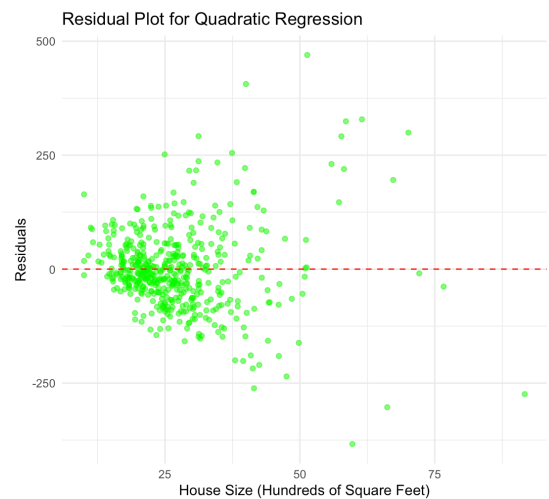
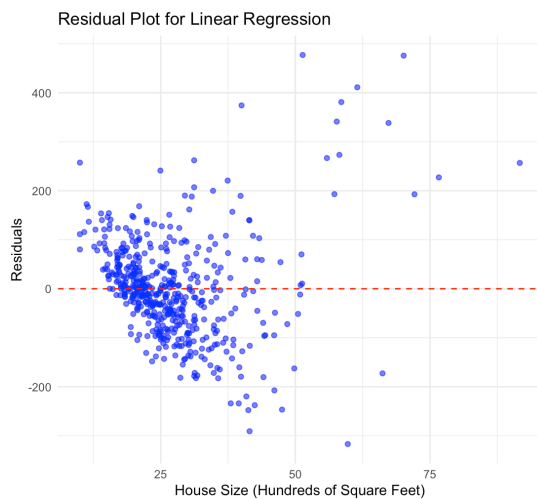
d.

Quadratic Regression with Tangent at 2000 sq ft



e. elasticity=0.8819511

f.

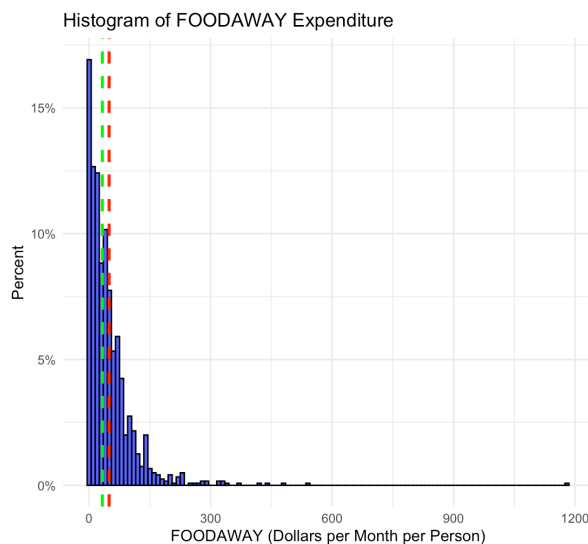


殘差模式並不呈現隨機分佈。隨著 SQFT（平方英尺） 增加，殘差的變異性也隨之增加，表示可能不滿足homoskedasticity的假設。

g. 線性回歸的殘差平方和為 5,262,846.9，而二次回歸的殘差平方和為 4,222,356.3。二次模型的 SSE 較低，表示數據點與二次模型的擬合線更接近，和線性模型的擬合線相比，誤差較小，代表二次模型能更好地解釋數據的變異性。

Q 2.25

a.



平均值 (Mean): 49.27085

25th 百分位數 (Q1): 12.04

中位數 (Median): 32.555

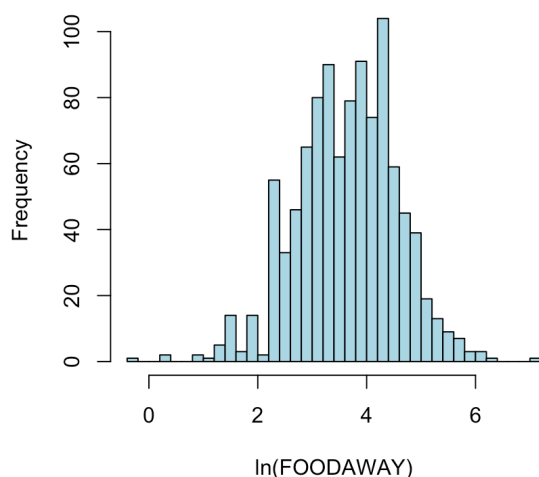
75th 百分位數 (Q3): 67.5025

b.

	N	Mean	Median
Advanced=1	257	73.2	48.2
College=1	369	48.6	36.1
None	574	39.01	26.0

c.

Histogram of ln(FOODAWAY)



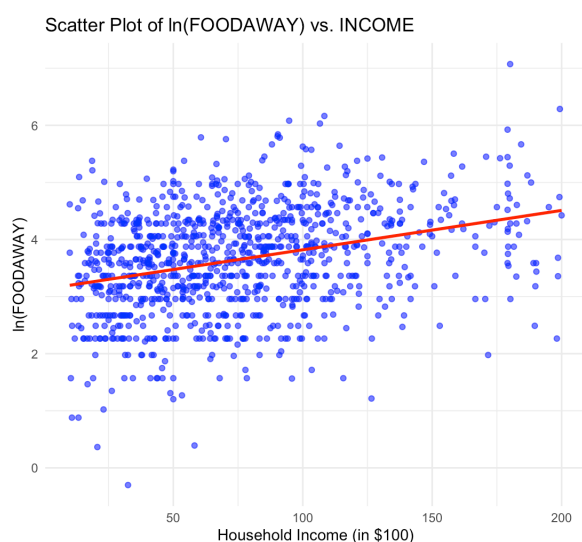
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
-0.3011	3.0759	3.6865	3.6508	4.2797	7.0724	178

ln(FOODAWAY) 值減少了 178 個，因為有 178 個家庭報告每人在外出食物上的支出為 0 美元，並且 $\ln(0)$ 未定義，所以產生了無法執行的「缺失值」。

d. $\ln(\text{FOODAWAY}) = 3.1293 + 0.0069\text{INCOME}$

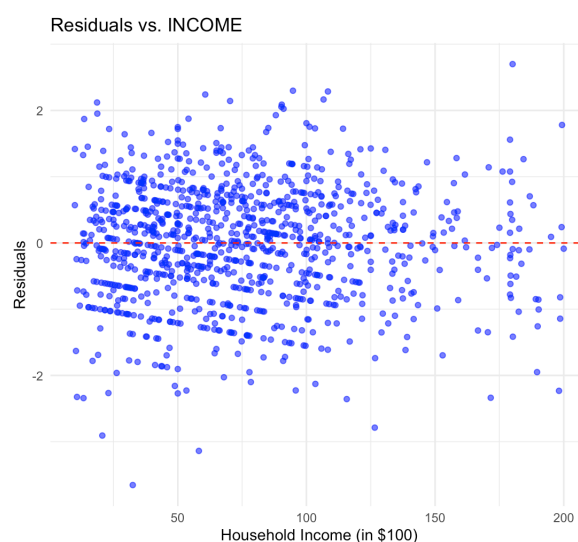
每增加 \$100 的家庭收入，每人的外出用餐支出將增加約 0.69%

e.



ln(FODAWAY)和Income呈現正相關

f.

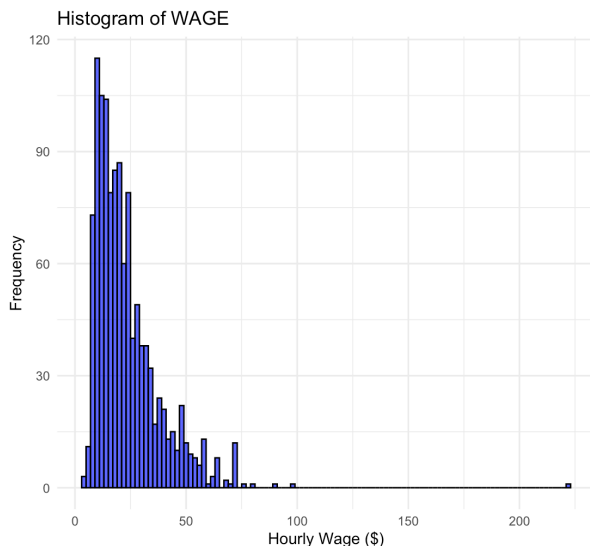


OLS 殘差看起來是隨機分佈

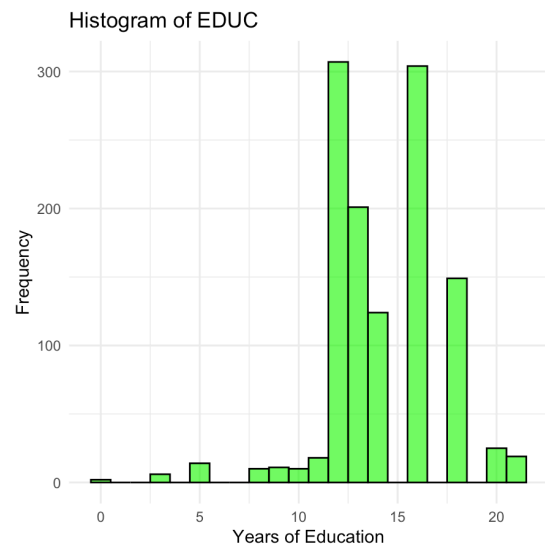
Q 2.28

a.

	Min.	1st Qu.	Median	Mean	3rd Qu	Max.
WAGE	3.94	13.00	19.30	23.64	29.80	221.10
EDUC	0.0	12.0	14.0	14.2	16.0	21.0



工資呈現右偏分佈，表示大多數人工資相對低



教育方面大多集中在12~16年，表示大部分的人有高中、大學學歷

b. $WAGE = -10.4 + 2.3968EDUC$

截距 (Intercept) 為 -10.4000，當教育年數 (educ) 為 0 時，預測薪資為 -10.4（但教育年數不可能為 0）

教育 (educ) 每增加一年，薪資 (wage) 平均增加 2.3968 單位（應為美元），這個係數在 1% 顯著水準下顯著

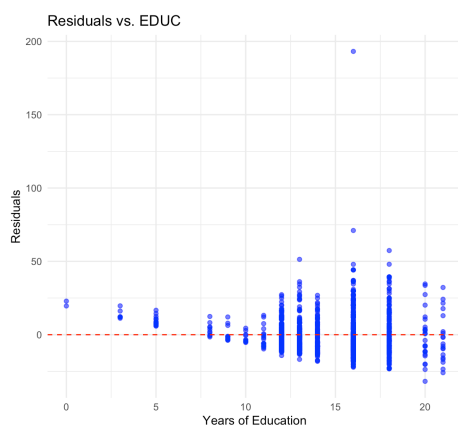
```
Call:
lm(formula = wage ~ educ, data = cps5_small)

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   -5.3138 1.38e-07 ***
educ         2.3968     0.1354   17.7121 < 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
```

c.



隨著 EDUCATION 變大，殘差也變大
⇒ 不滿足 Homoskedasticity (SR5)

如果 OLS 假設 (SR1-SR5) 成立
殘差應該在各個 EDUC 水平下均勻分佈，變異性不應該隨著 EDUC 增加而改變

d.

男性 回歸方程： $wage = -8.2849 + 2.3785 \times educ$

女性 回歸方程： $wage = -16.6028 + 2.6595 \times educ$

黑人 回歸方程： $wage = -6.2541 + 1.9233 \times educ$

白人 回歸方程： $wage = -10.4747 + 2.4178 \times educ$

解釋：

教育對薪資的影響在女性群體中最大(2.6595)，表示每多一年教育，女性薪資平均增加2.66;在黑人群體的教育回報率最低(1.9233)，表示教育每增加一年，薪資的增幅相對較小。女性的截距最低 (-16.6028)，表示即使在教育年數為 0 的情況下，預測薪資也比男性、黑人和白人都低。

性別 (男性 vs. 女性)

女性的教育回報 (2.6595) 高於男性 (2.3785)，表示女性在更高教育水平的職業中有較高的薪資成長機會。但女性的截距 (-16.6028) 明顯低於男性 (-8.2849)，表示在低教育水平下，女性的薪資較低。

種族 (黑人 vs. 白人)

白人教育回報 (2.4178) 高於黑人 (1.9233)，顯示黑人族群的教育邊際收益較低。

黑人截距 (-6.2541) 高於白人 (-10.4747)，但因為黑人教育回報較低，在較高教育水準時，薪資增幅趨於落後。

e. $wage = 4.916477 + 0.089134 \times educ^2$

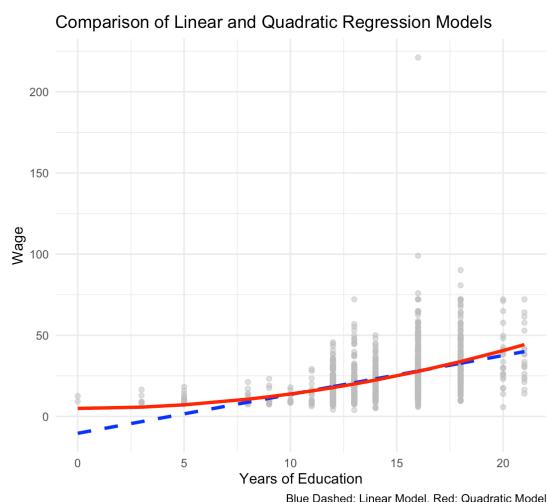
$$marginal\ effect = 2 \times 0.089134 \times educ$$

二次回歸的話，邊際影響裡含有education這個變數會隨著education的值有不同的邊際效果，隨著EDUCATION變大，邊際影響變大。

邊際影響 (EDUC = 12) = $2 \times 0.089134 \times 12 = 2.139$

邊際影響 (EDUC = 16) = $2 \times 0.089134 \times 16 = 2.852$

f.



二次回歸直線 (紅色實線) 比線性回歸直線更能更準確的預測變化，在 (EDUC<10) 紅色線也更貼近實際值。