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?

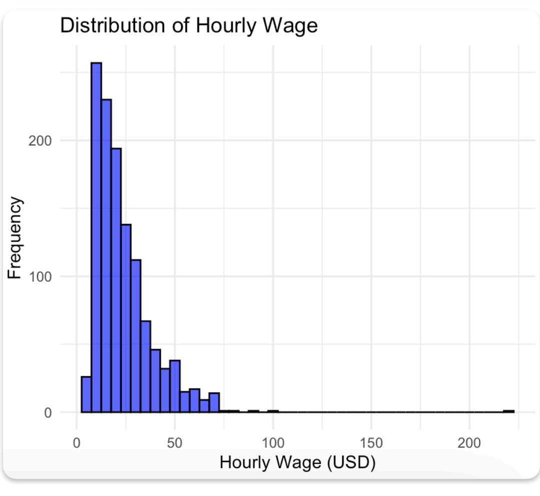
d. Estimate separate regressions for males, females, blacks, and whites. Compare the results.

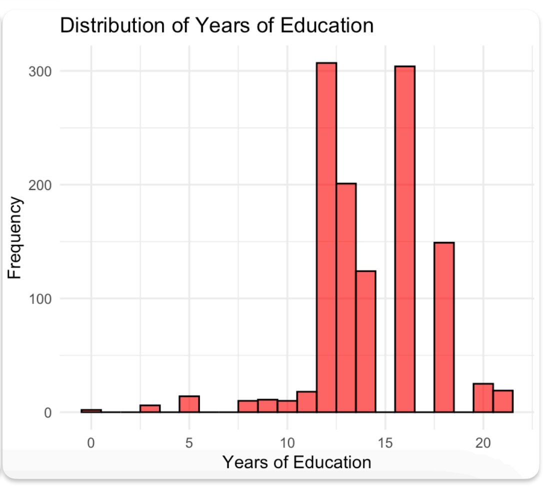
e. Estimate the quadratic regression $WAGE = \alpha_1 + \alpha_2 EDUC^2 + e$ and discuss the results. Estimate the marginal effect of another year of education on wage for a person with 12 years of education and for a person with 16 years of education. Compare these values to the estimated marginal effect 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.

```
# 摘要統計
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
```





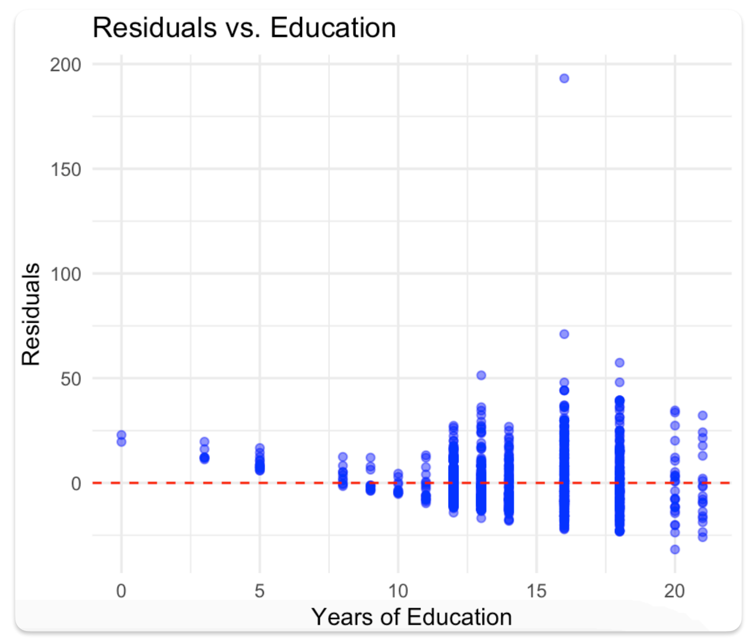
WAGE（工資）：呈右偏分佈（skewed right），表示少數人擁有極高的工資，而大部分人的工資相對較低。
EDUC（教育年數）：受教育年數多集中在 12-16 年，顯示多數人至少完成了高中或部分大學教育。
綜上，教育程度與工資之間可能存在正向關係。

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
教育年數的係數 β2 =2.3968，表示在其他條件不變下，教育年數每增加1年，預期每小時薪資增加 $1.2。
截距 -10.4，表示當教育年數為 0 年，預期每小時薪資為 $-10.4。（在此教育年數不可能為0，因薪資不可能為負數）

c.



由上圖可知：

- 殘差的變異數隨 EDUC 增加而變大。
- 存在極端值（Outliers）。
- 殘差未完全隨機分佈，在特定教育年數（如 12-16 年）可能較集中。

模型可能違反 OLS 假設（SR1-SR5），特別是：

- SR2：E(e | X) = 0。
  由上圖可知，殘差在某些教育年數範圍內明顯偏離 0。
- SR3：Homoskedasticity，Var(e | X) = σ^2。
  由上圖可知，殘差的變異數隨 EDUC 增加，表示存在 Heteroskedasticity。

d. 男性：wage = -8.2849 + 2.3785*educ

```
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
```

女性：wage = -16.6028 + 2.6595*educ

```
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
```

黑人：wage = -6.2541 + 1.9233*educ

```
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
```

白人：wage = -10.475 + 2.418*educ

```
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
```

回歸結果摘要：

族群	截距 (β_1)	教育係數 (β_2)	R^2	p 值
男性	-8.2849	**2.3785**	0.1927	< 2.2e-16
女性	-16.6028	**2.6595**	0.2764	< 2.2e-16
黑人	-6.2541	**1.9233**	0.1846	4.79e-06
白人	-10.475	**2.418**	0.2072	< 2.2e-16

不同性別比較：女性的教育回報率較高（斜率較高），但當 educ = 0 時，女性的起薪較低（截距較低），可能存在性別薪資差距。

不同種族比較：黑人的教育回報率較低（斜率較低），當 educ = 0 時，白人的起薪較低（截距較低），但白人工資提升的邊際效益較高（斜率較高）。

e.

```
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 ***
---
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
```

$\widehat{wage} = 4.9165 + 0.0891\, educ^2$

Marginal Effect (ME) : $\dfrac{d(wage)}{d(educ)} = 2 \times 0.0891\, educ$

when $educ = 12$, $ME = 2 \times 0.0891 \times 12 = 2.1384$

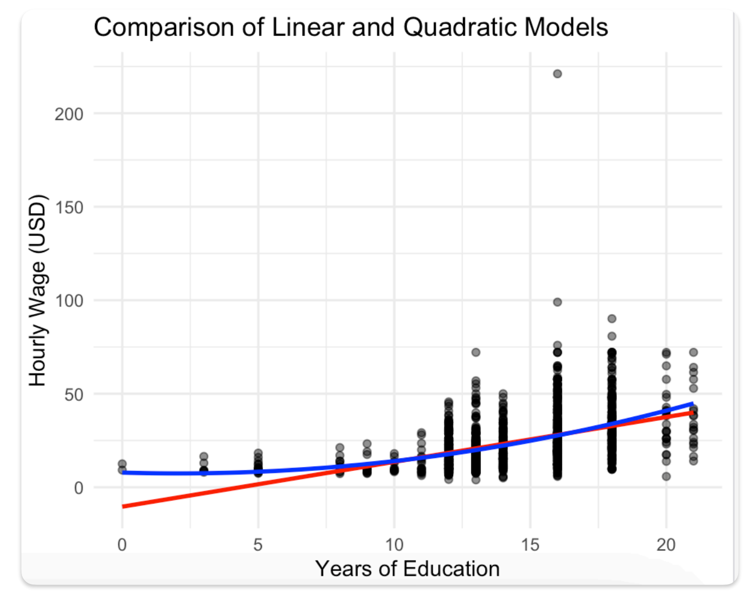
when $educ = 16$, $ME = 2 \times 0.0891 \times 16 = 2.8512$

與 (b) 相比，在 線性回歸 中，Marginal Effect 是常數，ME = 2.3968
但在 二次回歸 中，Marginal Effect 會變動：

- 當 EDUC = 12 時，ME = 2.1384（比線性回歸小）。
- 當 EDUC = 16 時，ME = 2.8512（比線性回歸大）。

線性回歸假設固定回報，但二次回歸顯示教育對工資的影響可能是遞增的。

f.



二次回歸（藍色）比線性回歸（紅色）更能捕捉數據的變化趨勢：
在低教育水準（EDUC < 10），線性回歸（紅色）對低教育年數的擬合偏低，特別是在 EDUC < 5 時，預測的工資接近 0 或負數，不合理。二次回歸（藍色）曲線更貼合數據點。