HW0303

Yung-Jung Cheng

2025-03-08

Q17

The data file collegetown contains observations on 500 single-family houses sold in Baton Rouge, Louisiana, during 2009–2013. The data include sale price (in thousands of dollars), PRICE, and total interior area of the house in hundreds of square feet, SQFT.

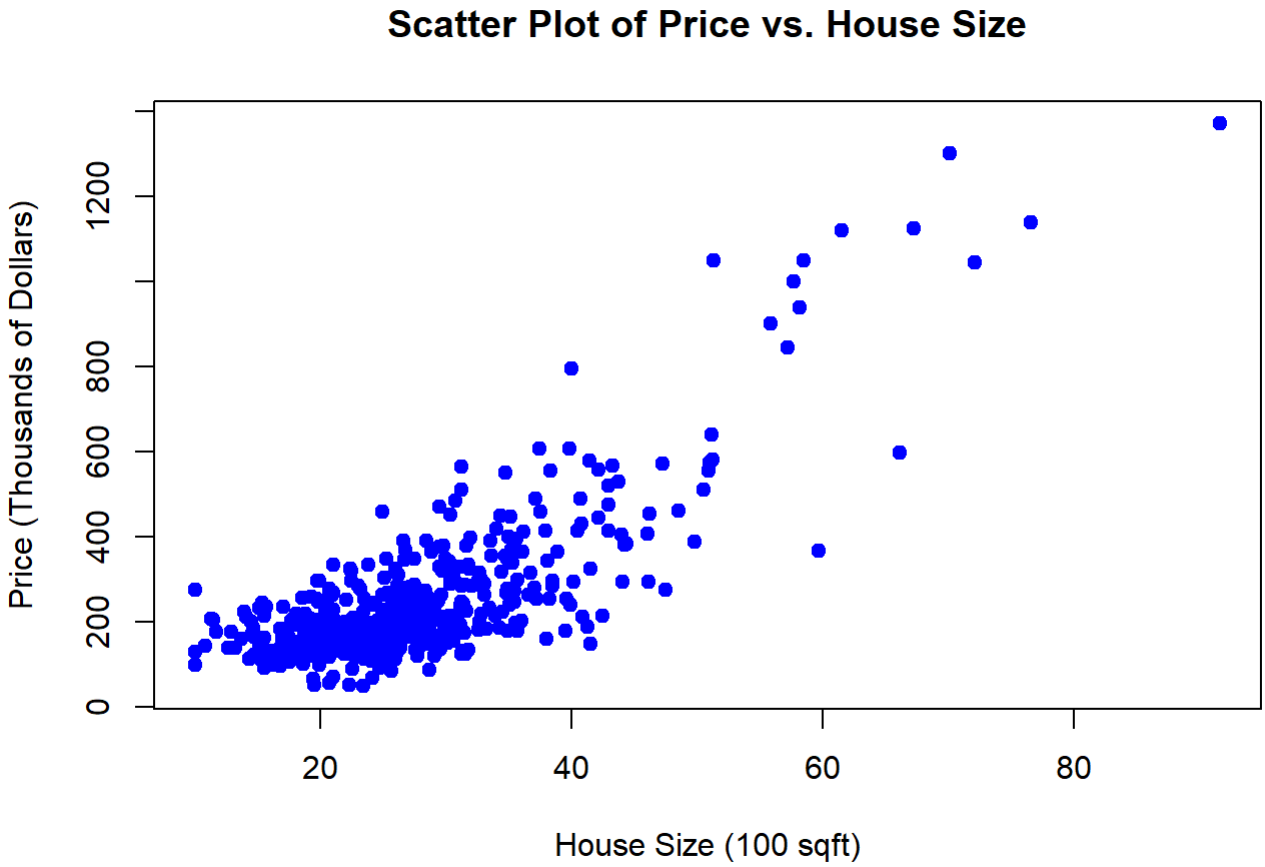
```
data("collegetown")
```

Q17 (a)

Plot house price against house size in a scatter diagram.

Ans

```
# 繪製房價對房屋面積的散點圖
plot(collegetown$sqft, collegetown$price,
     main="Scatter Plot of Price vs. House Size",
     xlab="House Size (100 sqft)",
     ylab="Price (Thousands of Dollars)",
     pch=19, col="blue")
```



Q17 (b)

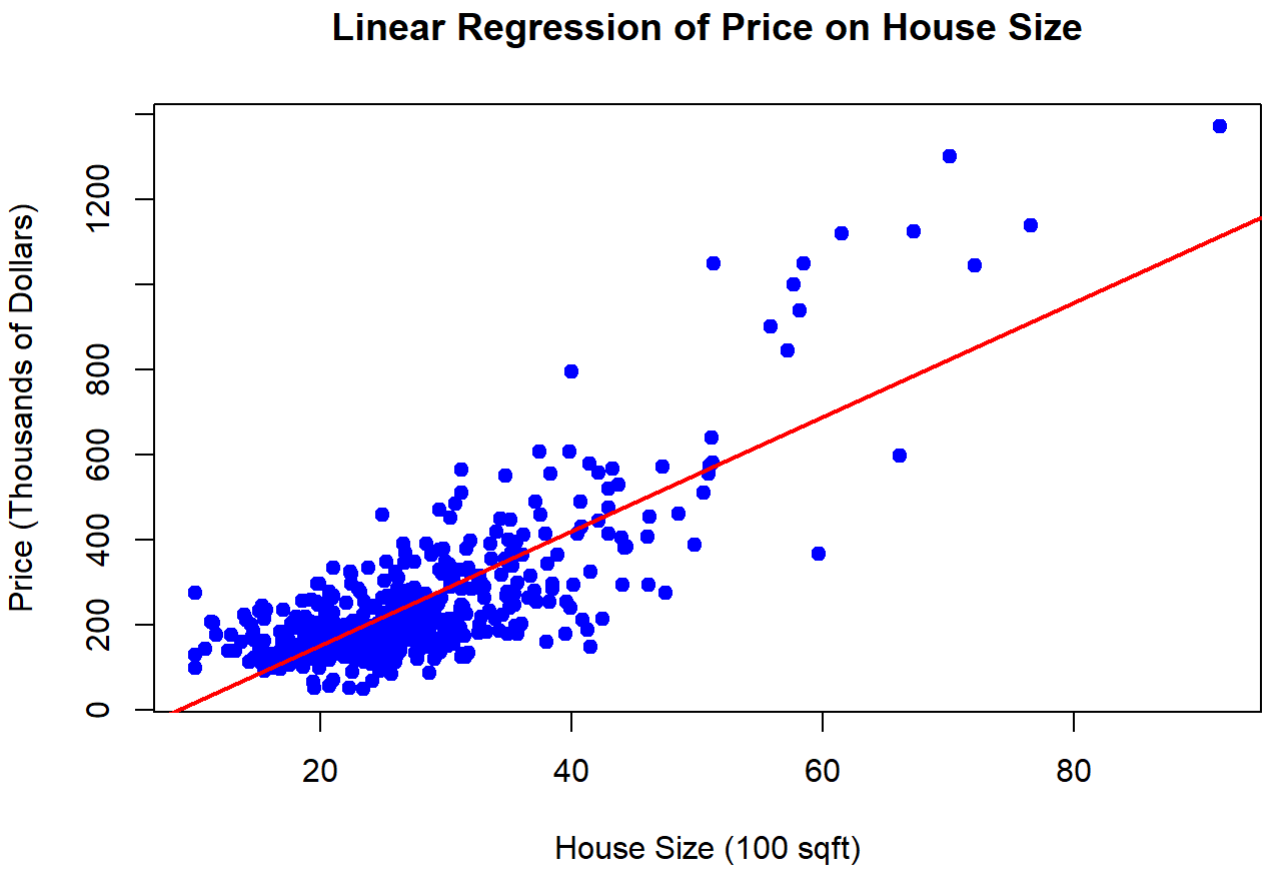
Estimate the linear regression model PRICE = β1 + β2SQFT + e. Interpret the estimates. Draw a sketch of the fitted line.

Ans

```
# 建立線性迴歸模型
linear_model <- lm(price ~ sqft, data=collegetown)

# 查看模型摘要以解釋估計值
summary(linear_model)
```

```
##
## Call:
## lm(formula = price ~ sqft, data = collegetown)
##
## 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 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 102.8 on 498 degrees of freedom
## Multiple R-squared:  0.6413, Adjusted R-squared:  0.6406
## F-statistic: 890.4 on 1 and 498 DF,  p-value: < 2.2e-16
```

```
# 繪製數據和擬合線
plot(collegetown$sqft, collegetown$price,
     main="Linear Regression of Price on House Size",
     xlab="House Size (100 sqft)",
     ylab="Price (Thousands of Dollars)",
     pch=19, col="blue")
abline(linear_model, col="red", lwd=2)  # 添加擬合線
```



Q17 (c)

Estimate the quadratic regression model PRICE = α1 + α2SQFT 2 + e. Compute the marginal effect of an additional 100 square feet of living area in a home with 2000 square feet of living space.

Ans

```
# 建立二次迴歸模型
quadratic_model <- lm(price ~ sqft + I(sqft^2), data=collegetown)

# 查看模型摘要
summary(quadratic_model)
```

```
##
## Call:
## lm(formula = price ~ sqft + I(sqft^2), data = collegetown)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -386.71  -48.66   -8.78   37.64  472.80
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 124.68914   24.49255   5.091 5.07e-07 ***
## sqft         -1.87040    1.42603  -1.312     0.19
## I(sqft^2)     0.20796    0.01863  11.163  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 92.01 on 497 degrees of freedom
## Multiple R-squared:  0.7132, Adjusted R-squared:  0.7121
## F-statistic:   618 on 2 and 497 DF,  p-value: < 2.2e-16
```

```
# 計算在2000平方英尺時額外100平方英尺的邊際效果
# 邊際效果為導數 dPrice/dSQFT = α2 + 2*α3*SQFT 在 SQFT = 20（因為數據中的SQFT是以100為單位）
marginal_effect <- coef(quadratic_model)["sqft"] + 2 * coef(quadratic_model)["I(sqft^2)"] * 2
0
marginal_effect
```

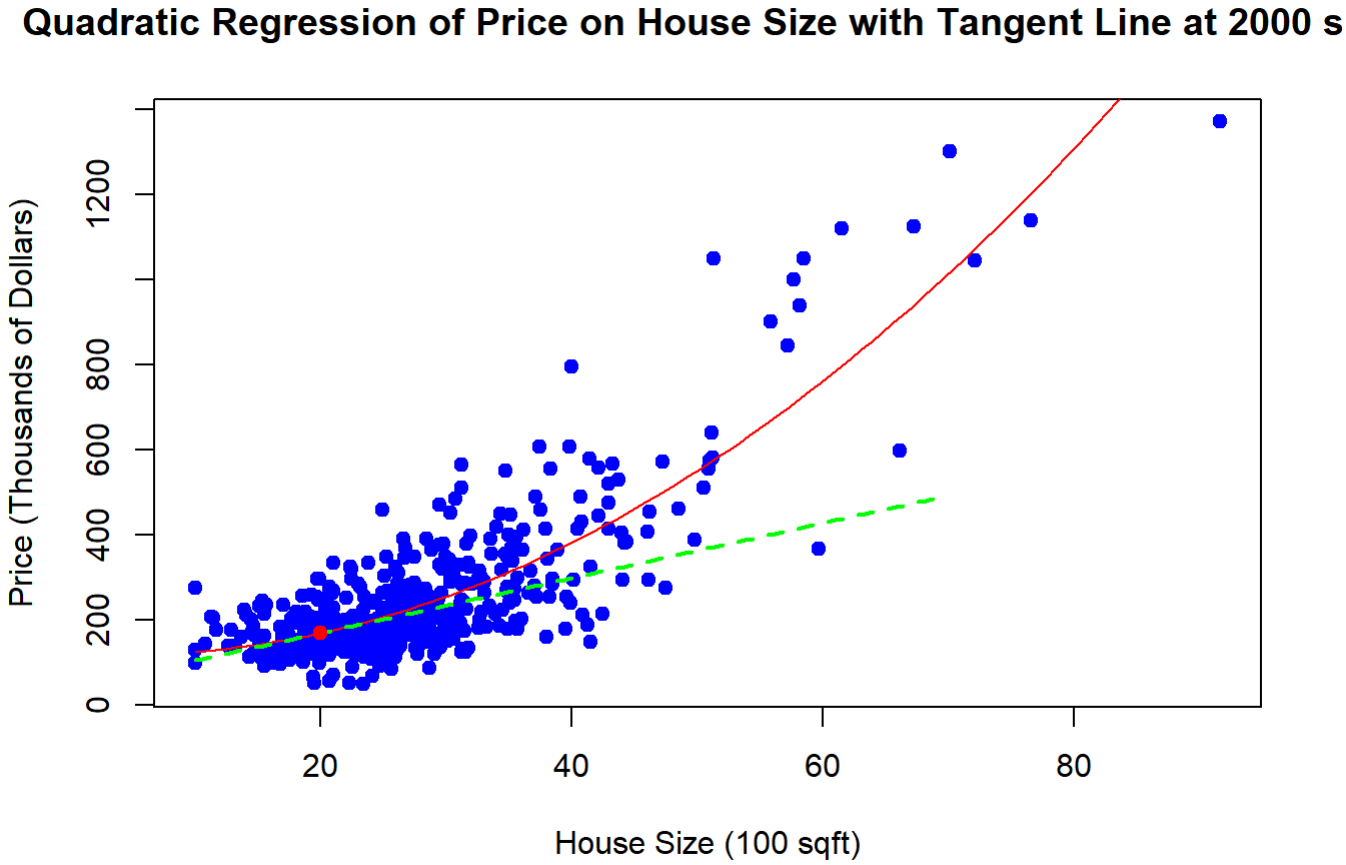
```
##     sqft
## 6.448092
```

Q17 (d)

Graph the fitted curve for the model in part (c). On the graph, sketch the line that is tangent to the curve for a 2000-square-foot house.

Ans

```
# 繪製數據點
plot(collegetown$sqft, collegetown$price,
     main="Quadratic Regression of Price on House Size with Tangent Line at 2000 sqft",
     xlab="House Size (100 sqft)",
     ylab="Price (Thousands of Dollars)",
     pch=19, col="blue")

# 繪製擬合曲線
curve(coef(quadratic_model)["(Intercept)"] +
      coef(quadratic_model)["sqft"] * x +
      coef(quadratic_model)["I(sqft^2)"] * x^2,
      from=min(collegetown$sqft), to=max(collegetown$sqft), add=TRUE, col="red")

# 計算切線斜率
slope <- coef(quadratic_model)["sqft"] + 2 * coef(quadratic_model)["I(sqft^2)"] * 20

# 繪製切線於2000平方英尺
tangent_line <- function(x) {
  intercept <- coef(quadratic_model)["(Intercept)"] +
               coef(quadratic_model)["sqft"] * 20 +
               coef(quadratic_model)["I(sqft^2)"] * 20^2 -
               slope * 20
  intercept + slope * x
}

# 添加切線到圖形
curve(tangent_line(x), from=10, to=70, add=TRUE, col="green", lwd=2, lty=2)
points(20, tangent_line(20), pch=19, col="red")  # 標記切點
```



Q17 (e)

For the model in part (c), compute the elasticity of PRICE with respect to SQFT for a home with 2000 square feet of living space.

Ans

```
# 從二次迴歸模型中提取係數
alpha1 <- coef(quadratic_model)["(Intercept)"]
alpha2 <- coef(quadratic_model)["sqft"]
alpha3 <- coef(quadratic_model)["I(sqft^2)"]

# 計算在SQFT = 20 (即2000平方英尺)時的價格
price_at_2000_sqft <- alpha1 + alpha2 * 20 + alpha3 * 20^2

# 計算在SQFT = 20處的邊際效果(即導數)
marginal_effect_at_2000_sqft <- alpha2 + 2 * alpha3 * 20

# 計算彈性:(dPrice/Price) / (dSQFT/SQFT) = (邊際效果 * SQFT) / 價格
elasticity <- (marginal_effect_at_2000_sqft * 20) / price_at_2000_sqft

# 輸出彈性值
elasticity
```

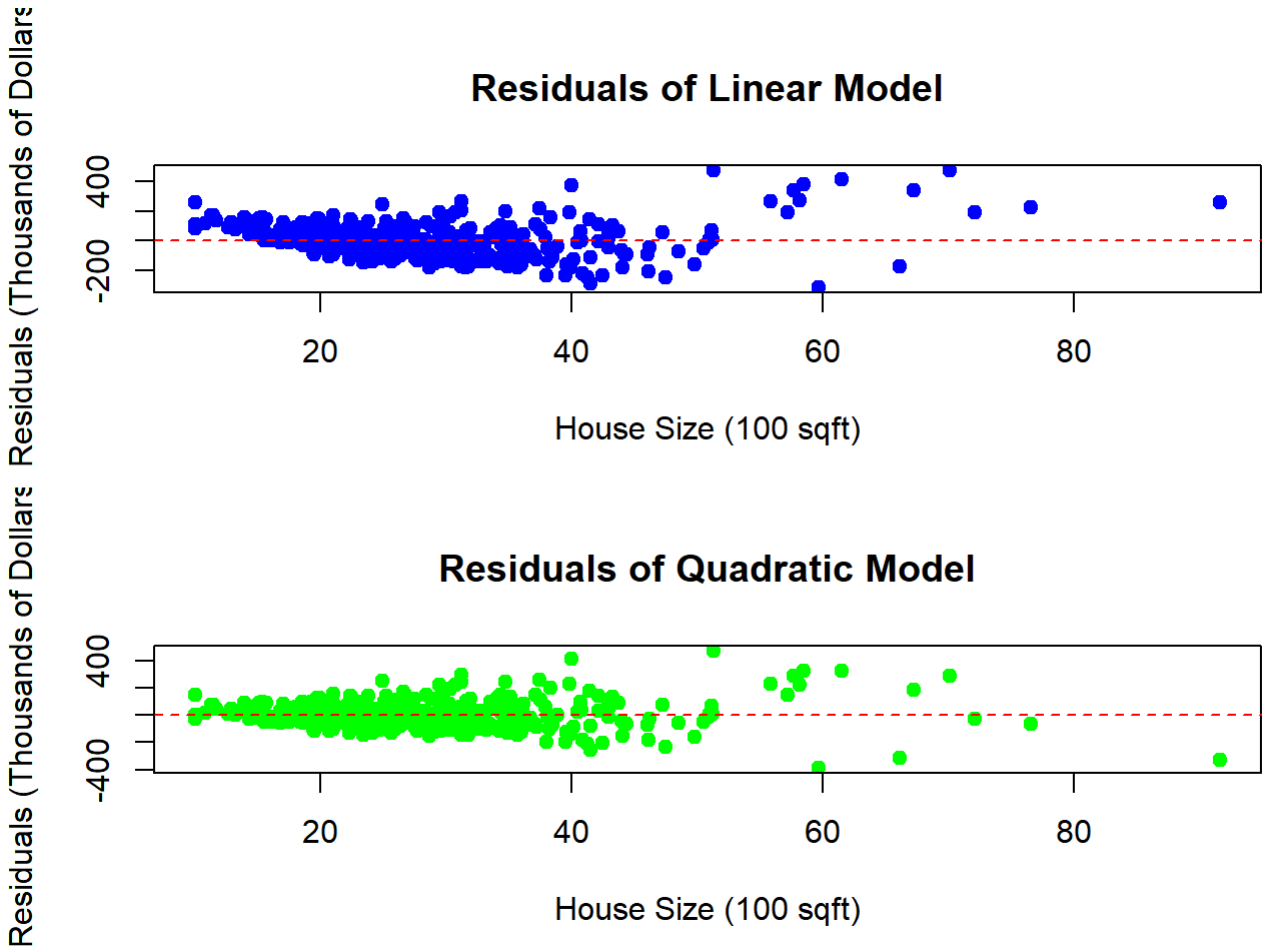
```
##      sqft
## 0.7565249
```

Q17 (f)

For the regressions in (b) and (c), compute the least squares residuals and plot them against SQFT. Do any of our assumptions appear violated?

Ans

```
# 線性模型的殘差
residuals_linear <- residuals(linear_model)

# 二次模型的殘差
residuals_quadratic <- residuals(quadratic_model)

# 繪製殘差圖
par(mfrow=c(2,1))  # 將圖形區域分為兩行一列

# 線性模型殘差
plot(collegetown$sqft, residuals_linear, main="Residuals of Linear Model",
     xlab="House Size (100 sqft)", ylab="Residuals (Thousands of Dollars)",
     pch=19, col="blue")
abline(h=0, col="red", lty=2)  # 添加水平線於殘差為0處

# 二次模型殘差
plot(collegetown$sqft, residuals_quadratic, main="Residuals of Quadratic Model",
     xlab="House Size (100 sqft)", ylab="Residuals (Thousands of Dollars)",
     pch=19, col="green")
abline(h=0, col="red", lty=2)  # 添加水平線於殘差為0處
```



Both appear to be acceptable.

##Q17 (g) One basis for choosing between these two specifications is how well the data are fit by the model. Compare the sum of squared residuals (SSE) from the models in (b) and (c). Which model has a lower SSE? How does having a lower SSE indicate a “better-fitting” model?

Ans

```
# 計算線性模型的 SSE
SSE_linear <- sum(residuals_linear^2)

# 計算二次模型的 SSE
SSE_quadratic <- sum(residuals_quadratic^2)

# 輸出 SSE 值
SSE_linear
```

```
## [1] 5262847
```

```
SSE_quadratic
```

```
## [1] 4207791
```

```
# 判斷哪個模型 SSE 較低
if (SSE_linear < SSE_quadratic) {
  print("Linear model has a better fit (lower SSE).")
} else {
  print("Quadratic model has a better fit (lower SSE).")
}
```

```
## [1] "Quadratic model has a better fit (lower SSE)."
```

A lower SSE means that the predicted values are, on average, closer to the actual values. And that indicate it is a “better-fitting” model

---

Q25

Consumer expenditure data from 2013 are contained in the file cex5_small. [Note: cex5 is a larger version with more observations and variables.] Data are on three-person households consisting of a husband and wife, plus one other member, with incomes between $1000 per month to $20,000 per month. FOODAWAY is past quarter’s food away from home expenditure per month per person, in dollars, and INCOME is household monthly income during past year, in $100 units.

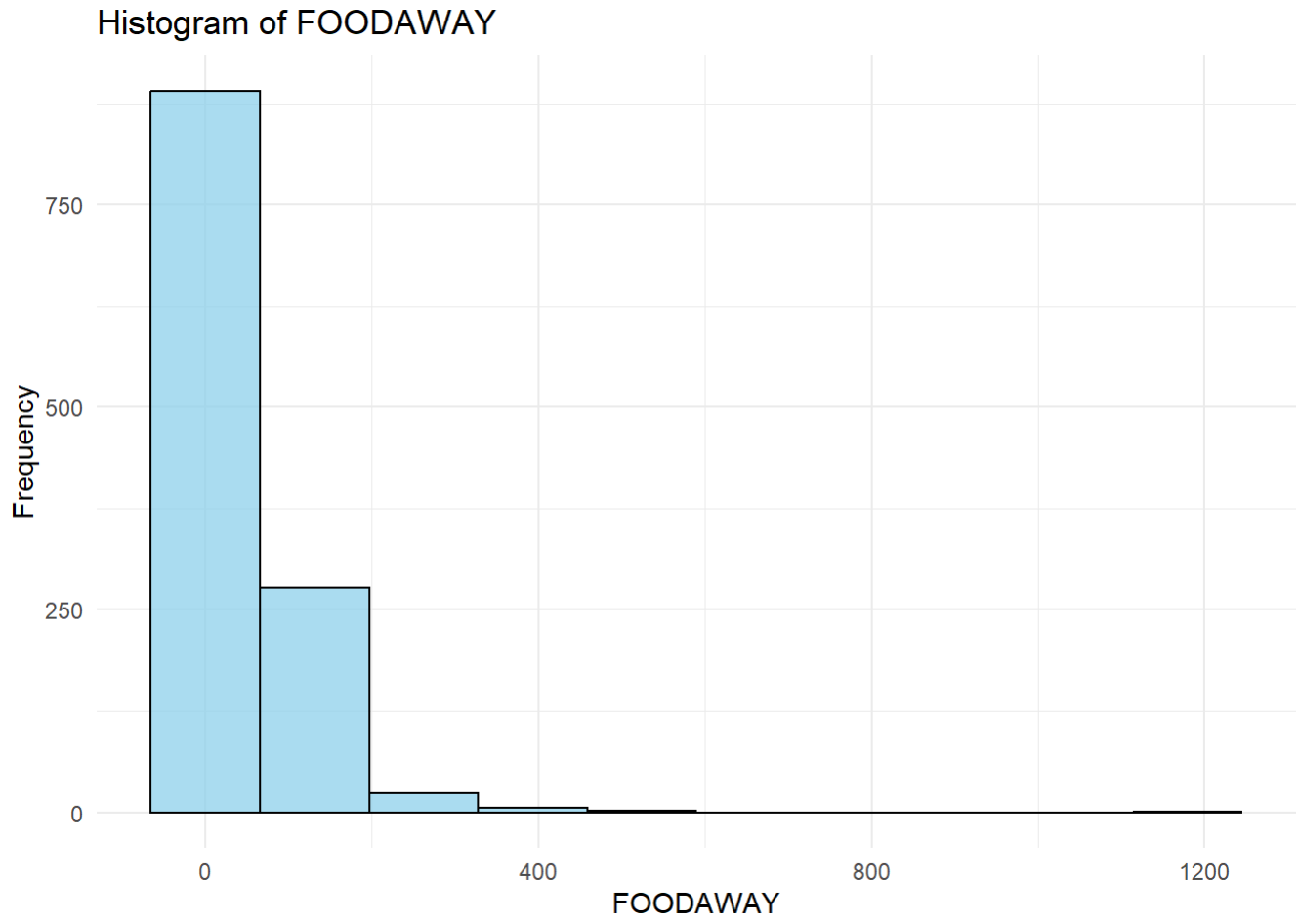
```
data("cex5_small")
```

Q25(a)

Construct a histogram of FOODAWAY and its summary statistics. What are the mean and median values? What are the 25th and 75th percentiles?

Ans

```
# 繪製 FOODAWAY 直方圖
ggplot(cex5_small, aes(x = foodaway)) +
  geom_histogram(bins = 10, fill = "skyblue", color = "black", alpha = 0.7) +
  labs(title = "Histogram of FOODAWAY", x = "FOODAWAY", y = "Frequency") +
  theme_minimal()
```




```
# 計算統計摘要
summary(cex5_small$foodaway)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   12.04   32.55   49.27   67.50 1179.00
```

```
# 計算 25th 和 75th 百分位數
quantile(cex5_small$foodaway, probs = c(0.25, 0.75))
```

```
##     25%     75%
## 12.0400 67.5025
```

Q25(b)

What are the mean and median values of FOODAWAY for households including a member with an advanced degree? With a college degree member? With no advanced or college degree member?

Ans

```
# 計算不同教育程度的 FOODAWAY 平均數與中位數，並移除分組
foodaway_stats <- cex5_small %>%
  group_by(advanced, college) %>%
  summarise(
    mean_foodaway = mean(foodaway, na.rm = TRUE),
    median_foodaway = median(foodaway, na.rm = TRUE),
    .groups = "drop"  # 這行讓輸出變成一般的 dataframe，而非分組資料
  )

# 顯示結果
print(foodaway_stats)
```

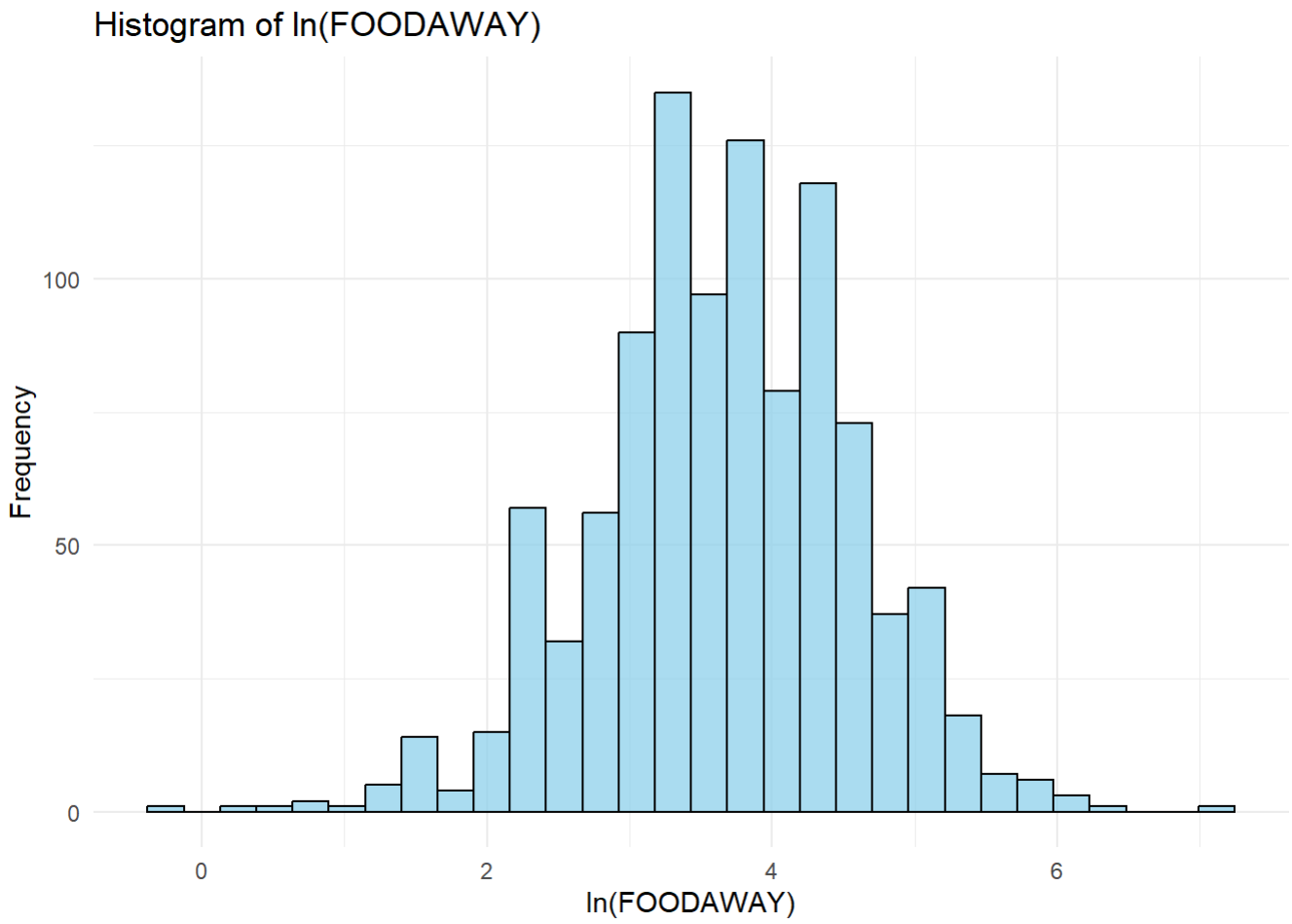
```
## # A tibble: 3 × 4
##   advanced college mean_foodaway median_foodaway
##      <int>   <int>         <dbl>           <dbl>
## 1        0       0          39.0            26.0
## 2        0       1          48.6            36.1
## 3        1       0          73.2            48.2
```

Q25(c)

Construct a histogram of ln(FOODAWAY) and its summary statistics. Explain why FOODAWAY and ln(FOODAWAY) have different numbers of observations.

Ans

```
# 先過濾掉 FOODAWAY 為 0 的觀測值，因為 ln(0) 不存在
cex5_small_filtered <- cex5_small %>%
  filter(foodaway > 0) %>%
  mutate(ln_foodaway = log(foodaway))

# 繪製 ln(FOODAWAY) 直方圖
ggplot(cex5_small_filtered, aes(x = ln_foodaway)) +
  geom_histogram(bins = 30, fill = "skyblue", color = "black", alpha = 0.7) +
  labs(title = "Histogram of ln(FOODAWAY)", x = "ln(FOODAWAY)", y = "Frequency") +
  theme_minimal()
```




```
# 計算 ln(FOODAWAY) 的統計摘要
summary(cex5_small_filtered$ln_foodaway)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.3011  3.0759  3.6865  3.6508  4.2797  7.0724
```

Since $ln(0)$ does not exist, we need to remove these observations. As a result, $FOODAWAY$ and $ln(FOODAWAY)$ have different numbers of observations.

Q25(d)

Estimate the linear regression ln(FOODAWAY)=β1 + β2INCOME + e. Interpret the estimated slope.

Ans

```
# 確保只使用 foodaway > 0 的數據，因為 ln(0) 無效
cex5_small_filtered <- cex5_small %>%
  filter(foodaway > 0) %>%
  mutate(ln_foodaway = log(foodaway))

# 進行線性回歸
lm_model <- lm(ln_foodaway ~ income, data = cex5_small_filtered)

# 顯示回歸結果
summary(lm_model)
```

```
##
## Call:
## lm(formula = ln_foodaway ~ income, data = cex5_small_filtered)
##
## 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
## Multiple R-squared:  0.09826,    Adjusted R-squared:  0.09738
## F-statistic: 111.1 on 1 and 1020 DF,  p-value: < 2.2e-16
```

For every 1-unit increase in INCOME (which is 100 USD), the predicted value of $ln(FOODAWAY)$ increases by 0.0069. The percentage increase in FOODAWAY can be calculated as: $e^{0.0069} - 1 = 0.69$

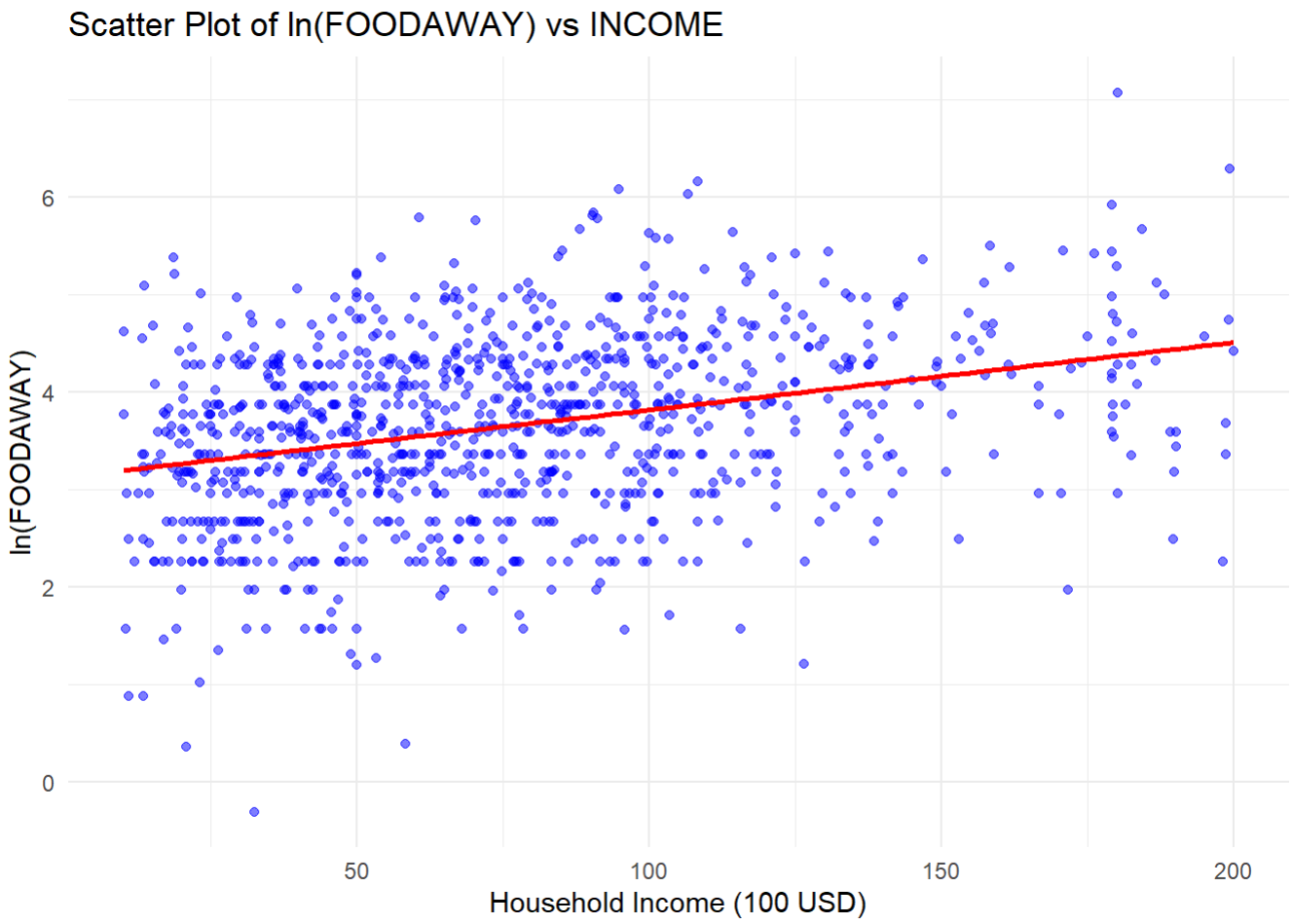
Q25(e)

Plot $ln(FOODAWAY)$ against $INCOME$, and include the fitted line from part (d).

Ans

```
# 繪製散佈圖與回歸線
ggplot(cex5_small_filtered, aes(x = income, y = ln_foodaway)) +
  geom_point(alpha = 0.5, color = "blue") +  # 繪製資料點
  geom_smooth(method = "lm", color = "red", se = FALSE) +  # 加上回歸線
  labs(title = "Scatter Plot of ln(FOODAWAY) vs INCOME",
       x = "Household Income (100 USD)",
       y = "ln(FOODAWAY)") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

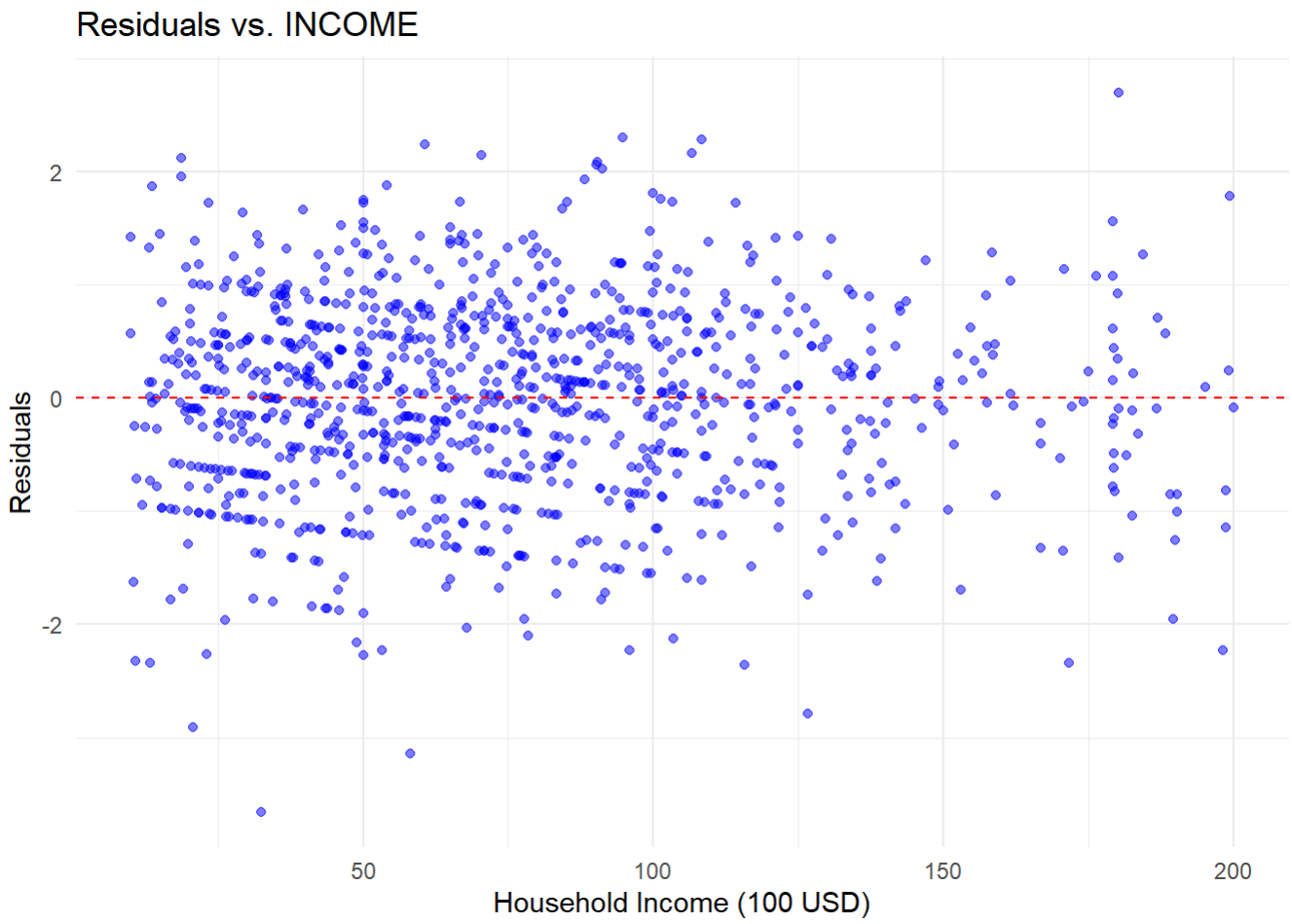


Q25(f)

Calculate the least squares residuals from the estimation in part (d). Plot them vs. INCOME. Do you find any unusual patterns, or do they seem completely random?

Ans

```
# 計算殘差
cex5_small_filtered$residuals <- lm_model$residuals

# 繪製殘差圖
ggplot(cex5_small_filtered, aes(x = income, y = residuals)) +
  geom_point(alpha = 0.5, color = "blue") +  # 繪製資料點
  geom_hline(yintercept = 0, color = "red", linetype = "dashed") +  # 添加 y=0 的參考線
  labs(title = "Residuals vs. INCOME",
       x = "Household Income (100 USD)",
       y = "Residuals") +
  theme_minimal()
```



it seem completely random!

Q28

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.]

```
data("cps5_small")
```

Q28(a)

Obtain the summary statistics and histograms for the variables WAGE and EDUC. Discuss the data characteristics.

Ans

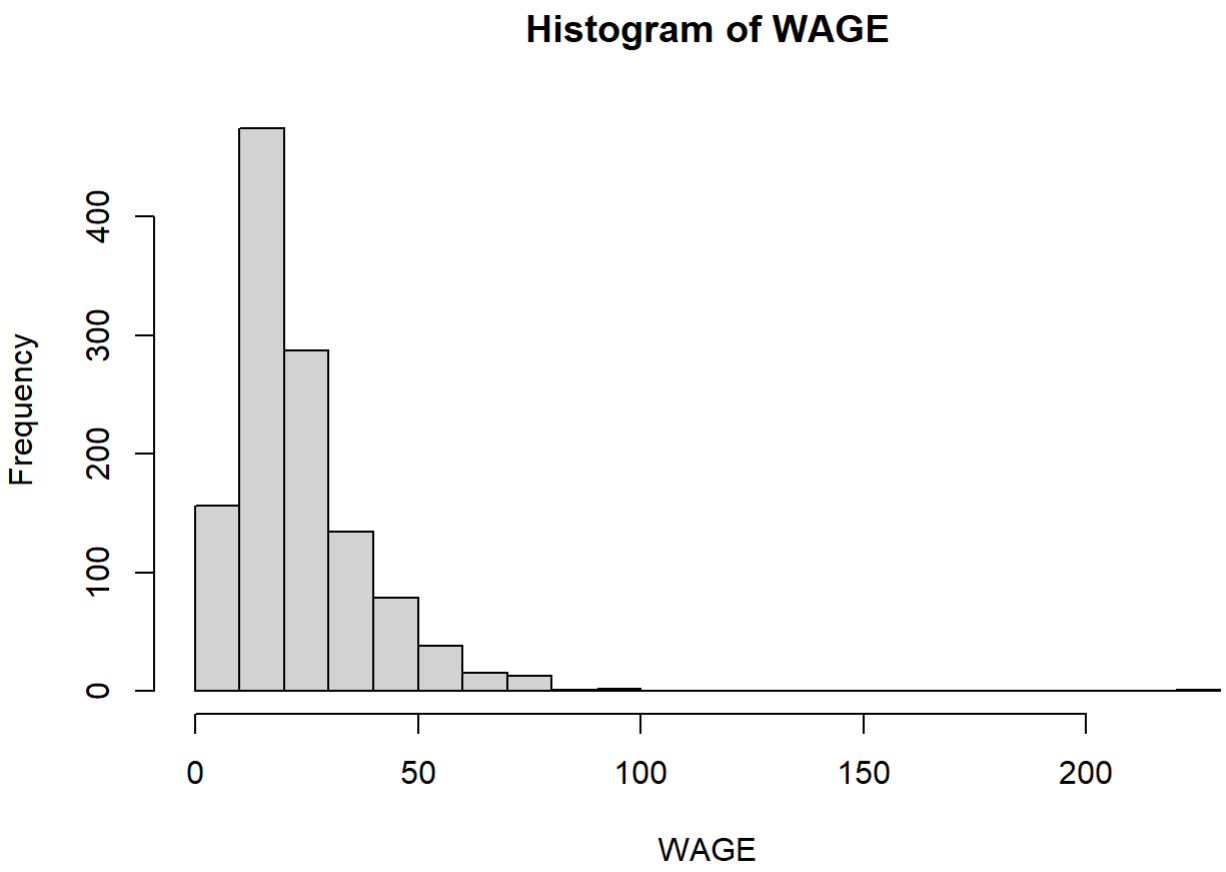
```
# 敘述統計
summary(cps5_small$wage)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.94   13.00   19.30   23.64   29.80  221.10
```

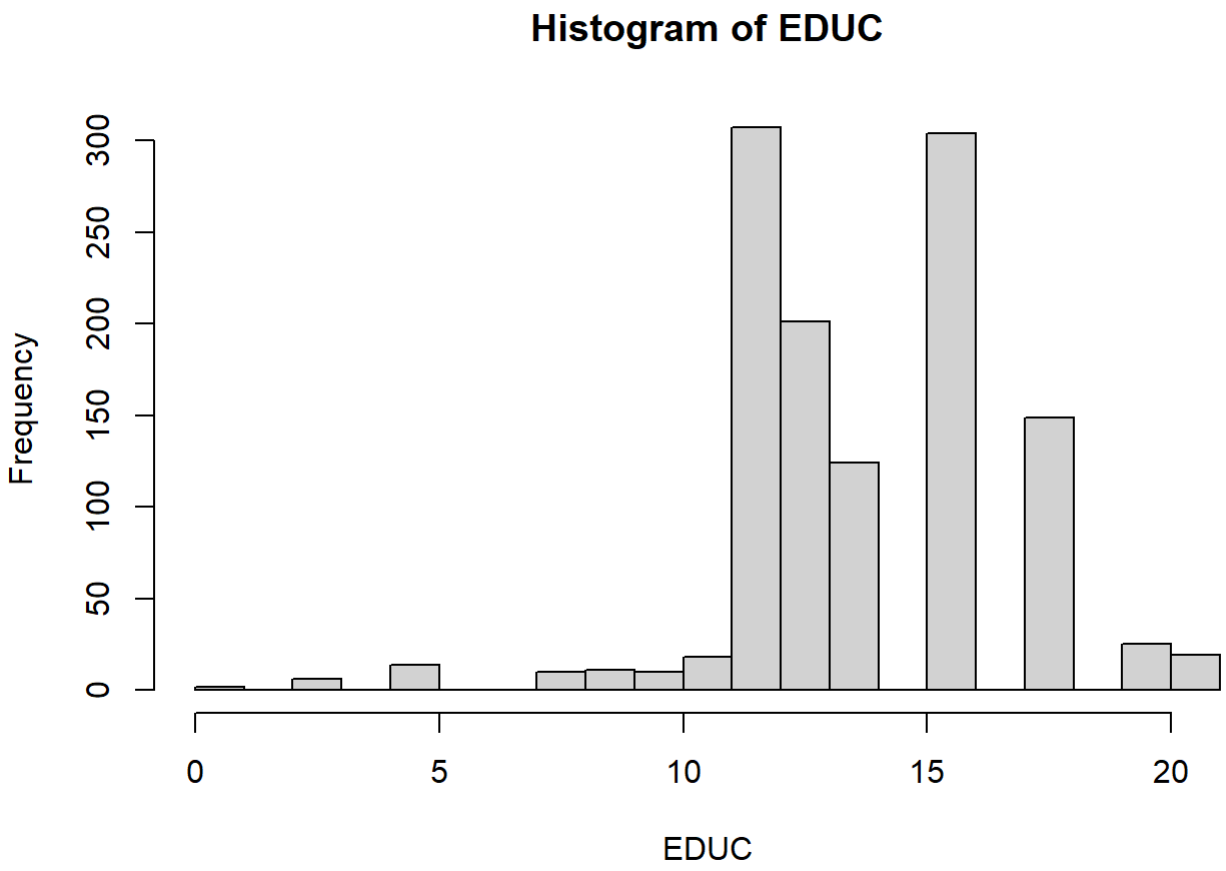
```
summary(cps5_small$educ)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0    12.0    14.0    14.2    16.0    21.0
```

```
# 繪製直方圖
hist(cps5_small$wage, main="Histogram of WAGE", xlab="WAGE", breaks=20)
```




```
hist(cps5_small$educ, main="Histogram of EDUC", xlab="EDUC", breaks=15)
```



- WAGE:
  - The distribution of wages is right-skewed, with most values clustered at the lower end and a few exceptionally high wages.
  - The median wage is significantly lower than the mean, indicating the presence of outliers or extreme high values.
  - Wages range from a minimum of approximately $3.94 to a maximum of $221.10.
- EDUC:
  - Education years show a concentrated distribution mostly between 12 to 16 years.
  - The median and mean are close, around 14 years, indicating a symmetric distribution around the middle school to college education level.
  - The data spans from no formal education (0 years) to advanced degrees (21 years).

Q28(b)

Estimate the linear regression WAGE = β1 + β2EDUC + e and discuss the results.

Ans

```
linear_model <- lm(wage ~ educ, data=cps5_small)
summary(linear_model)
```

```
##
## 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
```

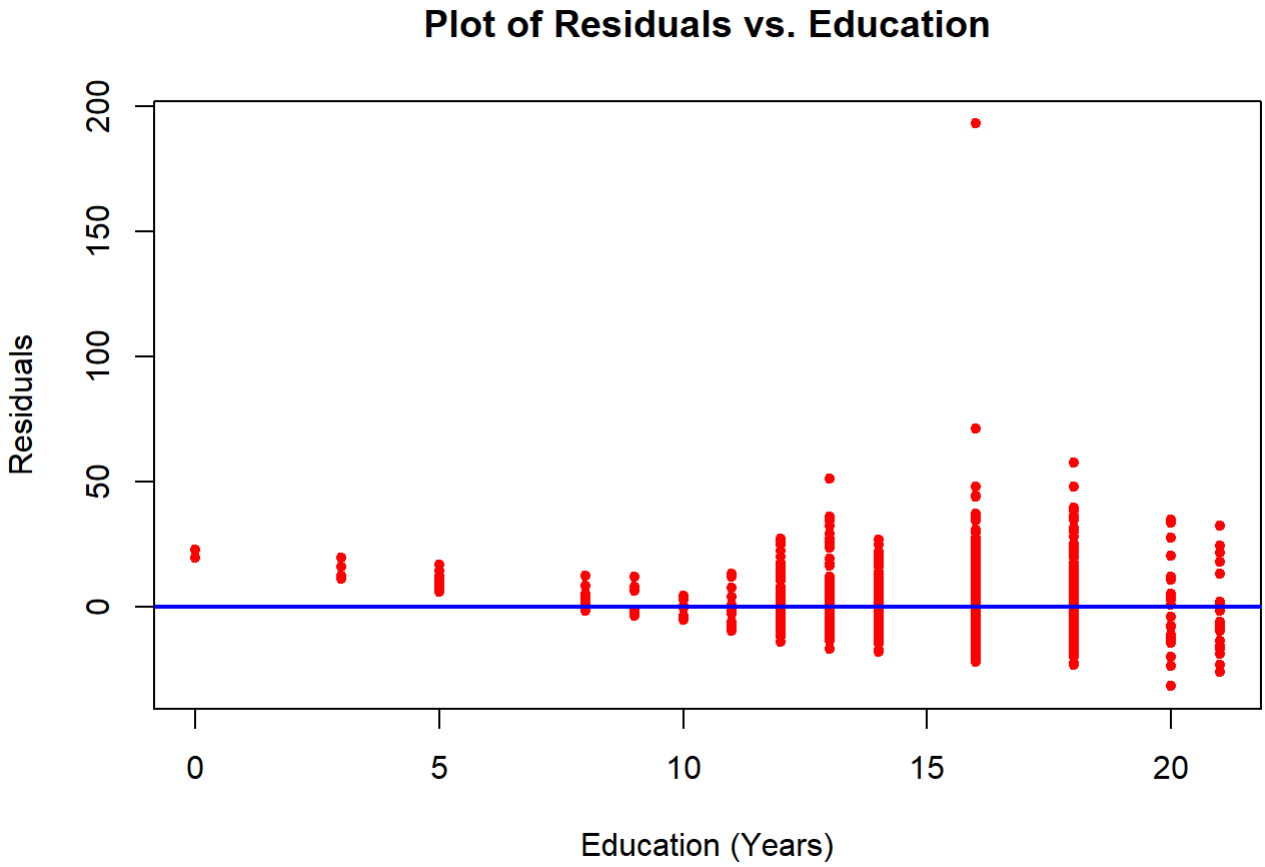
- **Equation**: $\mathrm{WAGE} = -10.4 + 2.3968 \times \mathrm{EDUC}.$
- **Slope**: Each additional year of education increases wages by approximately $2.3968, significant (p < 2e-16).
- **R-squared**: 0.2073, indicating the model explains about 21% of wage variance.
- **Residuals**: Range widely, suggesting outliers or extreme values. The model demonstrates a significant positive relationship between education and wages.

Q28(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?

Ans

```
# 計算殘差
residuals <- residuals(linear_model)

# 繪製殘差與教育年數的散點圖
plot(cps5_small$educ, residuals,
     xlab = "Education (Years)",
     ylab = "Residuals",
     main = "Plot of Residuals vs. Education",
     pch = 20, col = "red")

# 添加水平線於殘差=0的位置
abline(h = 0, col = "blue", lwd = 2)
```



Yes, it seem to be close to 0 with some outliers. If SR1 to SR5 hold, the residuals should be randomly distributed around zero with no systematic patterns or changes in variance across the range of education.

Q28(d)

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

Ans

```
# 針對男性和女性分別進行迴歸分析
male_model <- lm(wage ~ educ, data = cps5_small, subset = (female == 0))
female_model <- lm(wage ~ educ, data = cps5_small, subset = (female == 1))

# 針對黑人和白人分別進行迴歸分析
black_model <- lm(wage ~ educ, data = cps5_small, subset = (black == 1))
white_model <- lm(wage ~ educ, data = cps5_small, subset = (black == 0))

# 查看每個模型的摘要結果
summary(male_model)
```

```
##
## 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(female_model)
```

```
##
## 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(black_model)
```

```
##
## 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(white_model)
```

```
##
## 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
```

- Coefficients:
  - **Males**: For each additional year of education, wages increase by approximately $2.38.
  - **Females**: Each additional year increases wages by about $2.66.
  - **Blacks**: Each additional year increases wages by about $1.92.
  - **Whites**: Each additional year increases wages by about $2.42.
- Model Fit:
  - **Females** show the highest model explanatory power with an R-squared of 0.2764, indicating that education explains about 27.64% of wage variability.
  - **Males**, **Blacks**, and **Whites** have R-squared values of 0.1927, 0.1846, and 0.2072 respectively.
- Summary:

- Education impacts wages most significantly for females and least for blacks. The models vary in effectiveness, with females showing the highest explanatory power.

Q28(e)

Estimate the quadratic regression WAGE = α1 + α2EDUC2 + 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).

Ans

```
# 建立包含教育年數平方的二次迴歸模型
quadratic_model <- lm(wage ~ educ + I(educ^2), data = cps5_small)

# 查看模型摘要
summary(quadratic_model)
```

```
##
## Call:
## lm(formula = wage ~ educ + I(educ^2), data = cps5_small)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.219  -8.047  -2.708   5.307 193.439
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.82200    4.62512   1.691   0.0911 .
## educ        -0.42951    0.66438  -0.646   0.5181
## I(educ^2)    0.10434    0.02402   4.344 1.52e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.45 on 1197 degrees of freedom
## Multiple R-squared:  0.2196, Adjusted R-squared:  0.2183
## F-statistic: 168.4 on 2 and 1197 DF,  p-value: < 2.2e-16
```

```
# 計算在12年和16年教育時的邊際效應
educ_values <- c(12, 16)
marginal_effects <- coef(quadratic_model)[2] + 2 * coef(quadratic_model)[3] * educ_values

# 顯示邊際效應
names(marginal_effects) <- paste("Marginal effect at EDUC =", educ_values)
marginal_effects
```

```
## Marginal effect at EDUC = 12 Marginal effect at EDUC = 16
##                     2.074698                     2.909434
```

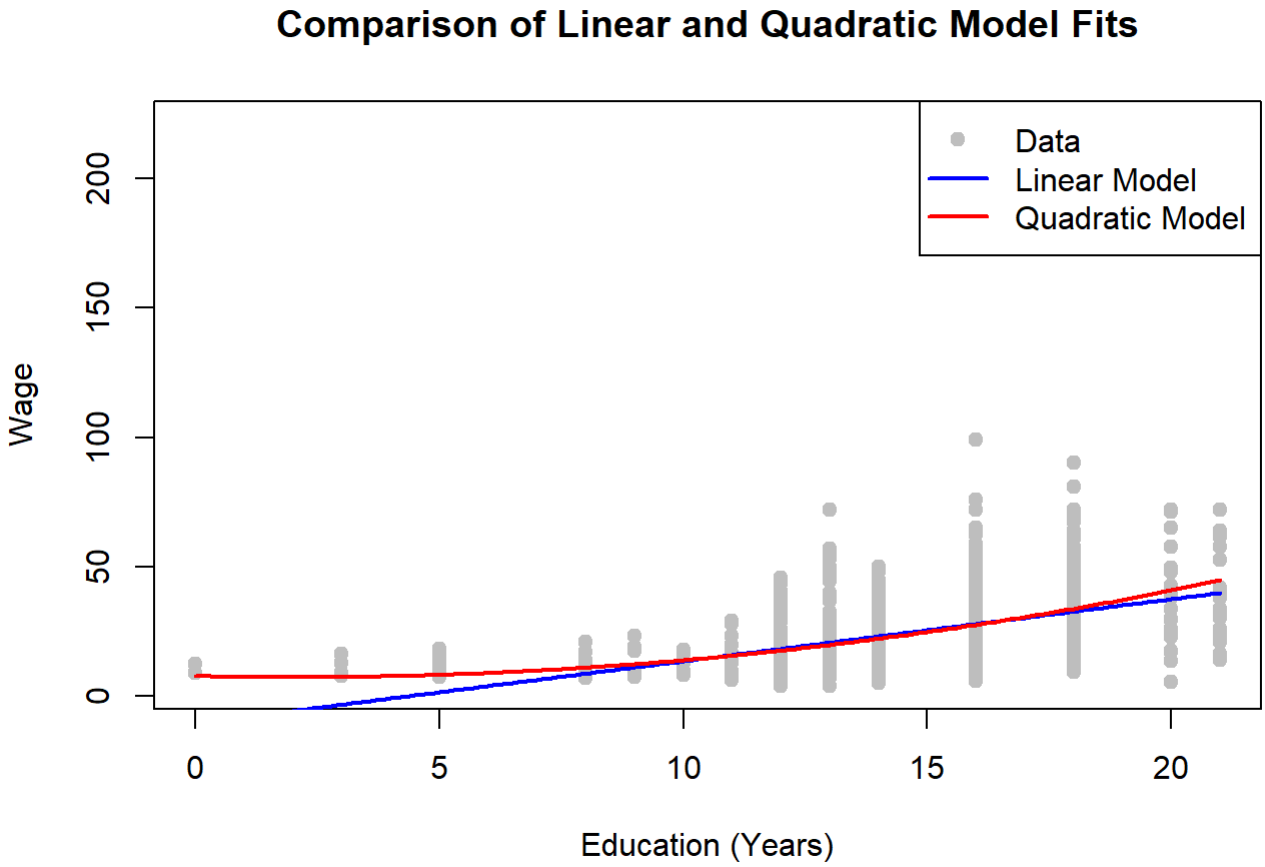
The quadratic regression shows increasing marginal effects of education on wages: 2.0747 for 12 years and 2.9094 for 16 years. Compared to the linear model's consistent increase of about $2.3968 per year, the quadratic model suggests that the wage benefit of education grows with higher levels of education.

Q28(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?

Ans

```
# 線性模型
linear_model <- lm(wage ~ educ, data = cps5_small)

# 二次模型
quadratic_model <- lm(wage ~ educ + I(educ^2), data = cps5_small)

# 生成教育年數的預測範圍
educ_range <- seq(from = min(cps5_small$educ), to = max(cps5_small$educ), by = 0.1)

# 計算線性模型的預測值
linear_predictions <- predict(linear_model, newdata = list(educ = educ_range))

# 計算二次模型的預測值
quadratic_predictions <- predict(quadratic_model, newdata = list(educ = educ_range))

# 繪製數據點
plot(cps5_small$educ, cps5_small$wage, xlab = "Education (Years)", ylab = "Wage",
     main = "Comparison of Linear and Quadratic Model Fits",
     pch = 19, col = "gray")

# 繪製線性模型的擬合線
lines(educ_range, linear_predictions, col = "blue", lwd = 2)

# 繪製二次模型的擬合線
lines(educ_range, quadratic_predictions, col = "red", lwd = 2)

# 添加圖例
legend("topright", legend = c("Data", "Linear Model", "Quadratic Model"),
       col = c("gray", "blue", "red"), pch = c(19, NA, NA), lty = c(NA, 1, 1), lwd = c(NA, 2,
2))
```



the quadratic model from part (e) appears to fit the data better.