8.16

A sample of 200 Chicago households was taken to investigate how far American households tend to travel when they take a vacation. Consider the model

$$MILES = \beta_1 + \beta_2 INCOME + \beta_3 AGE + \beta_4 KIDS + e$$

$MILES$ is miles driven per year, $INCOME$ is measured in $1000 units, AGE is the average age of the adult members of the household, and $KIDS$ is the number of children.

a. Use the data file $vacation$ to estimate the model by OLS. Construct a 95% interval estimate for the effect of one more child on miles traveled, holding the two other variables constant.

Ans.

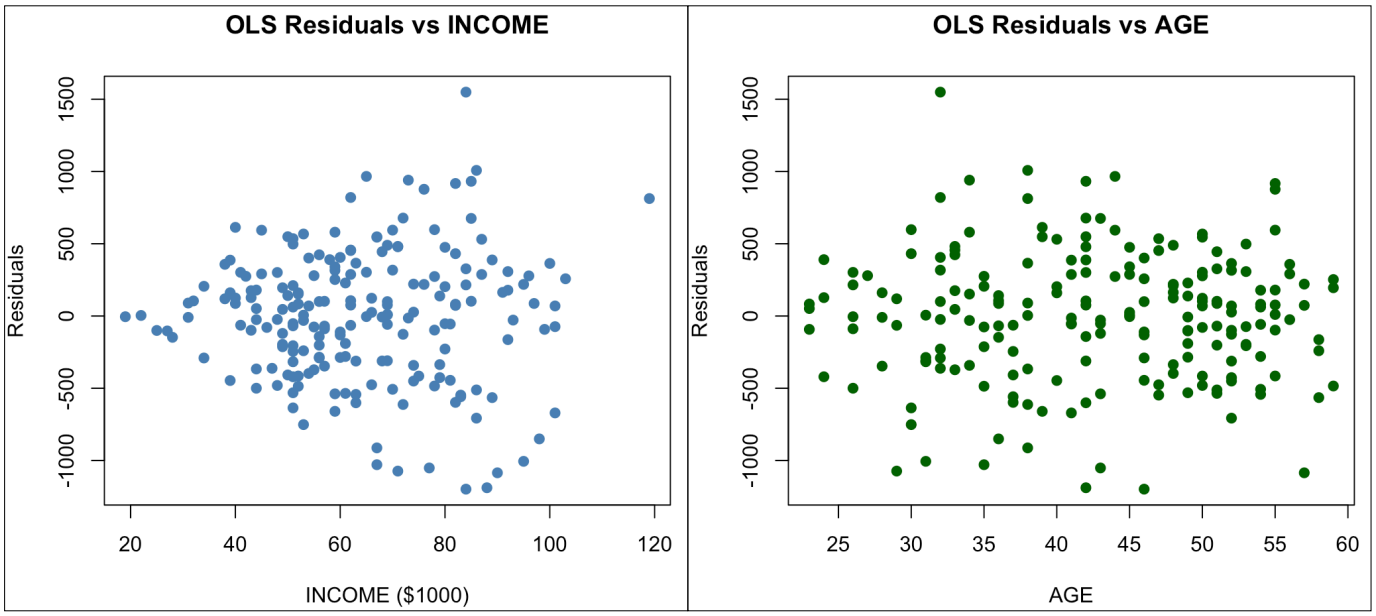
```
Call:
lm(formula = miles ~ income + age + kids, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-1198.14  -295.31    17.98   287.54  1549.41

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -391.548    169.775  -2.306   0.0221 *
income        14.201      1.800   7.889 2.10e-13 ***
age           15.741      3.757   4.189 4.23e-05 ***
kids         -81.826     27.130  -3.016   0.0029 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 452.3 on 196 degrees of freedom
Multiple R-squared:  0.3406,    Adjusted R-squared:  0.3305
F-statistic: 33.75 on 3 and 196 DF,  p-value: < 2.2e-16
```

```
> confint(model_ols, level = 0.95)
                 2.5 %    97.5 %
(Intercept) -726.36871 -56.72731
income        10.65097  17.75169
age            8.33086  23.15099
kids        -135.32981 -28.32302
```

kids 的係數為 −81.826，表示在控制收入和年齡不變的情況下，每多一個孩子，家庭平均每年旅行里程會減少約 81.826英里。kids 的 95% 信賴區間是 [−135.32981, −28.32302]。

b. Plot the OLS residuals versus $INCOME$ and AGE. Do you observe any patterns suggesting that heteroskedasticity is present?

Ans.



觀察「OLS 殘差 vs INCOME」散點圖，隨著 INCOME 增加，殘差範圍變大，有異質變異數的跡象。

觀察「OLS 殘差 vs AGE」散點圖，殘差沒有明顯的擴散或收斂趨勢。

c. **Sort the data according to increasing magnitude of income. Estimate the model using the first 90 observations and again using the last 90 observations. Carry out the Goldfeld–Quandt test for heteroskedastic errors at the 5% level. State the null and alternative hypotheses.**

Ans.

H_0 : 殘差變異數相等，即 $\sigma^2_{high} = \sigma^2_{low}$

H_1 : 高收入組的殘差變異數大於低收入組，即 $\sigma^2_{high} > \sigma^2_{low}$

$\alpha = 0.05 \qquad df_{high} = n_{high} - k = 90 - 4 = 86 \qquad df_{low} = n_{low} - k = 90 - 4 = 86$

Goldfeld–Quandt F statistic: 3.1041

F critical value: 1.4286

由於 F 統計量 (3.1041) > 臨界值 (1.4286)，落在拒絕域，我們拒絕虛無假設 $\sigma^2_{high} = \sigma^2_{low}$。

結論：在 5% 顯著水平下，有足夠的證據表明高收入組的殘差變異數大於低收入組。

```
        Goldfeld-Quandt test

data:  miles ~ income + age + kids
GQ = 3.1041, df1 = 86, df2 = 86, p-value = 1.64e-07
alternative hypothesis: variance increases from segment 1 to 2
```

d. **Estimate the model by OLS using heteroskedasticity robust standard errors. Construct a 95% interval estimate for the effect of one more child on miles traveled, holding the two other variables constant. How does this interval estimate compare to the one in (a)?**

Ans.

```
t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept) -391.5480   142.6548 -2.7447 0.0066190 **
income        14.2013     1.9389  7.3246 6.083e-12 ***
age           15.7409     3.9657  3.9692 0.0001011 ***
kids         -81.8264    29.1544 -2.8067 0.0055112 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> coefci(model_ols, vcov. = robust_se, level = 0.95)
                   2.5 %     97.5 %
(Intercept) -672.883378 -110.21263
income        10.377633   18.02503
age            7.919934   23.56191
kids        -139.322973  -24.32986
```

模型版本	kids 係數	kids 標準誤	kids 95% 信賴區間
OLS (a)	−81.826	27.130	[−135.32981, −28.32302]
Robust OLS (d)	−81.826	29.154	[−139.32297, −24.32986]

kids 係數：Robust SE 不改變 OLS 係數，兩者一致，皆為 −81.826。

kids 標準誤：Robust SE 較大（29.1544 vs 27.130），Robust SE 考慮了 heteroskedasticity 影響。

kids 的 95% 信賴區間：Robust SE [−139.3230, −24.32986] 比 OLS [−135.32981, −28.32302] 略寬。

反映出模型存在異質變異時，傳統 OLS 可能低估標準誤。Robust SE 提供更保守且穩健的推論結果。

e. **Obtain GLS estimates assuming $\sigma_i^2 = \sigma^2 INCOME_i^2$. Using both conventional GLS and robust GLS standard errors, construct a 95% interval estimate for the effect of one more child on miles traveled, holding the two other variables constant. How do these interval estimates compare to the ones in (a) and (d)?**

Ans.

```
Call:
lm(formula = miles ~ income + age + kids, data = data, weights = weights)

Weighted Residuals:
     Min       1Q   Median       3Q      Max
-15.1907  -4.9555   0.2488   4.3832  18.5462

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -424.996    121.444  -3.500 0.000577 ***
income        13.947      1.481   9.420  < 2e-16 ***
age           16.717      3.025   5.527 1.03e-07 ***
kids         -76.806     21.848  -3.515 0.000545 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.765 on 196 degrees of freedom
Multiple R-squared:  0.4573,    Adjusted R-squared:  0.449
F-statistic: 55.06 on 3 and 196 DF,  p-value: < 2.2e-16
```

```
> confint(model_gls, level = 0.95)
                  2.5 %     97.5 %
(Intercept) -664.50116 -185.49119
income        11.02744   16.86718
age           10.75260   22.68240
kids        -119.89450  -33.71808
```

```
t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept) -424.9962    95.8035 -4.4361 1.526e-05 ***
income        13.9473     1.3470 10.3545 < 2.2e-16 ***
age           16.7175     2.7974  5.9761 1.061e-08 ***
kids         -76.8063    22.6186 -3.3957 0.0008286 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> coefci(model_gls, vcov. = gls_robust_se, level = 0.95)
                  2.5 %     97.5 %
(Intercept) -613.93428 -236.05807
income        11.29086   16.60376
age           11.20062   22.23438
kids        -121.41339  -32.19919
```

模型版本	kids 係數	kids 標準誤	kids 95% 信賴區間
OLS (a)	−81.826	27.130	[−135.32981, −28.32302]
Robust OLS (d)	−81.826	29.154	[−139.32297, −24.32986]
GLS	−76.806	21.848	[−119.89450, −33.71808]
Robust GLS	−76.806	22.619	[−121.41339, −32.19919]

kids 係數：GLS 係數比 OLS 略小，因為 GLS 使用權重 $\frac{1}{INCOME_i^2}$ ，給予低 INCOME 觀測值更大權重，改變了估計。Robust GLS 係數與 GLS 相同，因為 Robust SE 不影響點估計。

kids 標準誤：GLS 模型因考慮 heteroskedasticity 結構，標準誤相對較小，推論更有效率。

kids 的 95% 信賴區間：

GLS 信賴區間 [−119.89450, −33.71808] 比 OLS [−135.322981, −28.323002] 窄，估計更精確

Robust GLS 信賴區間 [−121.41339, −32.19919] 比 GLS 略寬，但仍優於 OLS 與 Robust OLS。

GLS方法（無論是否使用Robust SE）都提供了更窄的信賴區間，表明這些估計更有效率。