Chapter 8

6.

a.

Male : $n_M = 577$, $SSE_M = 97161.9174$

Female : $\hat{\sigma}_F = 12.024$, $\hat{\sigma}_F^2 = 144.5766$

$H_0 : \hat{\sigma}_M^2 = \hat{\sigma}_F^2$, $H_1 : \hat{\sigma}_M^2 \neq \hat{\sigma}_F^2$ $(\alpha = 5\%)$

$\hat{\sigma}_M^2 = \frac{SSE_M}{n_M} = \frac{97161.9714}{577} = 168.35$

By GQ test, $F = \frac{\hat{\sigma}_M^2}{\hat{\sigma}_F^2} = \frac{168.35}{144.5766} = 1.1729$

$\because F_{593,419,0.025} = 0.8377$ and $F_{593,419,0.975} = 1.1968$

$\therefore RR = \{ F < 0.8377$ and $1.1968 < F \}$

Since $F = 1.1729 \in RR$, we fail to reject H_0.

There's no sufficient evidence to show that $\hat{\sigma}_M^2$ is diff. from $\hat{\sigma}_F^2$.

b.

Single : $n_S = 400$, $SSE_S = 56231.0382$, $\hat{\sigma}_S^2 = \left(\frac{56231.0382}{400}\right)^2 = 140.58$

Married : $n_M = 600$, $SSE_M = 100703.0471$, $\hat{\sigma}_M^2 = \left(\frac{100703.0471}{600}\right)^2 = 167.84$

$H_0 : \hat{\sigma}_S^2 = \hat{\sigma}_M^2$, $H_1 : \hat{\sigma}_S^2 < \hat{\sigma}_M^2$ $(\alpha = 5\%)$

By GQ test, $F = \frac{\hat{\sigma}_M^2}{\hat{\sigma}_S^2} = \frac{167.84}{140.58} = 1.1943$

$\because F_{594,394,0.05} = 1.16$

$\therefore RR = \{ F > 1.16 \}$

Since $F = 1.1943 \in RR$, we reject H_0.

There's sufficient evidence to show that $\sigma_M^2 > \sigma_S^2$.

c.

$NR^2 = 59.03$ $(K = 5 , \alpha = 0.05)$

$\because \chi^2_{(5, 0.05)} = 11.07$

$\therefore NR^2 = 59.03 > \chi^2 = 11.07$

→ We reject H_0. There's sufficient evidence to show that heteroskedasticity exists in the model.

→ This is consistent with part (b).

The error variation is diff. for married and single individuals.

d.

The test statistic = 78.82

The degrees of freedom = $5 \times (5+1) \div 2 - 1 = 14$

The critical value : $\chi^2_{14, 0.05} = 23.68$

$\because 78.82 > \chi^2_{14, 0.05} = 23.68$

$\therefore$ We reject H_0. There's sufficient evidence to show that heteroskedasticity exists in the model.

→ This is consistent with part (b) and (c) .

The error variation is diff. for married and single individuals.

e.

Interval estimates is narrower: EXPER, METRO, FEMALE

Interval estimates is wider: intercept, EDUC

→ No contradiction, but it illustrates heteroskedasticity since it affects OLS standard errors unevenly.

f.

It is compatible with result (b).

This is because part (b) tests the heteroskedasticity but part (f) tests the influence of MARRIED to the model. There's no conflict between these two parts.

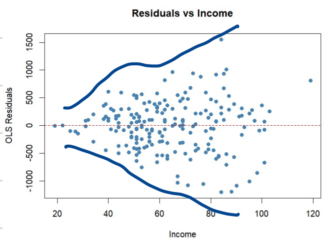
16.

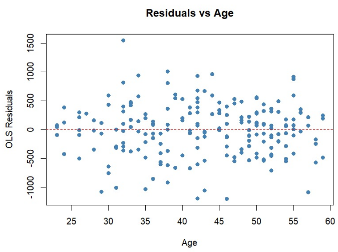
a.

```
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
```

→ 95% C.I. for KIDS = [-135.3298, -28.32302]

b.





Residuals v.s. Income — The spread of residuals gets larger as income increases. Therefore, the heteroskedasticity exists.

Residuals v.s. Age —— The spread of residuals is even with age. Therefore, no pattern exists.

c.

Let $\hat{\sigma}_2^2$ is for high-income and $\hat{\sigma}_1^2$ is for low-income.

$H_0: \hat{\sigma}_1^2 = \hat{\sigma}_2^2$, $H_1: \hat{\sigma}_1^2 < \hat{\sigma}_2^2$

By GQ-test, $F = \dfrac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} = \dfrac{27160654/(90-4)}{8750040/(90-4)} = 3.1041$

$\because F_{86,\,0.05} = 1.4286 < F = 3.1041$

$\therefore$ We reject H_0. There's sufficient evidence to show that heteroskedasticity exists in the model.

d.

```
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
```

→ 95% C.I. for KIDS = [-138.969, -24.684]

→ Compared with part (a): [-135.3298, -28.32302]

the point estimate is same but the standard error gets larger since robust SE considers heteroskedasticity.

Therefore, the confidence interval becomes wider.

e.

```
Call:
lm(formula = miles ~ income + age + kids, data = vacation_gls,
    weights = 1/(income^2))

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
```

```
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
```

Model Version	Estimate	Std. Error	95% Confidence Interval
OLS (a)	-81.826	27.130	[-135.329, -28.323]
Robust OLS (d)	-81.826	29.154	[-138.969, -24.684]
GLS	-76.806	21.848	[-119.894, -33.718]
Robust GLS	-76.806	22.618	[-121.139, -32.474]

→ The point estimate of GLS and Robust GLS is little larger than OLS ones,

and the standard error of GLS ones are also smaller.

Therefore, Robust GLS, which offers narrower C.I., shows more accurate inference.

18.

a.

$H_0: \sigma_M^2 = \sigma_F^2$, $H_1: \sigma_M^2 \neq \sigma_F^2$

$$F = \frac{\sigma_M^2}{\sigma_F^2} = \frac{1208.99}{925.18} = 1.0538$$

$\because RR = \{F < 0.9453 \text{ and } 1.0581 < F\}$

$\therefore F = 1.0538 \notin RR$

→ We fail to reject H_0.

There's no sufficient evidence to show that σ_M^2 is diff. from σ_F^2.

b.

H_0 : The heteroskedasticity is related to these variables.

H_1 : The heteroskedasticity is not related to these variables.

$\because NR^2 = 23.5568 > \chi^2_{3,0.01} = 11.3449$

$\therefore$ We reject H_0. There's sufficient evidence to show that heteroskedasticity exists in the model.

c.

H_0 : The standard errors are constant.

H_1 : The standard errors are variable.

$\because$ the p-value of the test $< 2.2 \times 10^{-16} < 0.01$

$\therefore$ We reject H_0. There's sufficient evidence to show that standard errors are variable.

→ The model exists heteroskedasticity.

d.

變數	OLS.標準誤	Robust.標準誤	改變幅度....
(Intercept)	0.0321	0.0328	+2.12%
educ	0.0018	0.0019	+8.39%
exper	0.0013	0.0013	+1.12%
I(exper^2)	0.0000	0.0000	+4.71%
female	0.0095	0.0095	-0.43%
black	0.0169	0.0161	-5.01%
metro	0.0123	0.0116	-5.89%
south	0.0136	0.0139	+2.51%
midwest	0.0141	0.0137	-2.69%
west	0.0144	0.0146	+1.07%

→ No. Also, it increases the stability and accuracy of the model.

e.

變數	估計值	標準誤	X95..信賴區間
(Intercept)	1.1922	0.0316	[1.1303, 1.2541]
educ	0.1017	0.0018	[0.0982, 0.1051]
exper	0.0301	0.0013	[0.0275, 0.0326]
I(exper^2)	-0.0005	0.0000	[-5e-04, -4e-04]
female	-0.1662	0.0095	[-0.1848, -0.1476]
black	-0.1109	0.0170	[-0.1442, -0.0775]
metro	0.1178	0.0115	[0.0953, 0.1402]
south	-0.0448	0.0135	[-0.0713, -0.0183]
midwest	-0.0632	0.0140	[-0.0906, -0.0358]
west	-0.0055	0.0144	[-0.0337, 0.0227]

→ Many intervals don't contain 0.
Therefore, FGLS has effectively eliminate the error caused by heteroskedasticity.

f.

變數	FGLS.係數	FGLS.SE	Robust.SE	SE.變化....
(Intercept)	1.1922	0.0316	0.0324	+2.43%
educ	0.1017	0.0018	0.0019	+7.26%
exper	0.0301	0.0013	0.0013	+0.55%
I(exper^2)	-0.0005	0.0000	0.0000	+2.31%
female	-0.1662	0.0095	0.0094	-0.45%
black	-0.1109	0.0170	0.0159	-6.61%
metro	0.1178	0.0115	0.0116	+0.9%
south	-0.0448	0.0135	0.0138	+2.31%
midwest	-0.0632	0.0140	0.0137	-1.94%
west	-0.0055	0.0144	0.0145	+0.92%

→ The standard error of robust is very close to FGLS', meaning that heteroskedasticity is controlled.

9.

FGLS + Robust. These two can effectively correct the error brought with heteroskedasticity.