8.6 Consider the wage equation

$$WAGE_i = \beta_1 + \beta_2 EDUC_i + \beta_3 EXPER_i + \beta_4 METRO_i + e_i \qquad \text{(XR8.6a)}$$

where wage is measured in dollars per hour, education and experience are in years, and $METRO = 1$ if the person lives in a metropolitan area. We have $N = 1000$ observations from 2013.

a. We are curious whether holding education, experience, and *METRO* constant, there is the same amount of random variation in wages for males and females. Suppose $\text{var}(e_i|\mathbf{x}_i, FEMALE = 0) = \sigma_M^2$ and $\text{var}(e_i|\mathbf{x}_i, FEMALE = 1) = \sigma_F^2$. We specifically wish to test the null hypothesis $\sigma_M^2 = \sigma_F^2$ against $\sigma_M^2 \neq \sigma_F^2$. Using 577 observations on males, we obtain the sum of squared OLS residuals, $SSE_M = 97161.9174$. The regression using data on females yields $\hat{\sigma}_F = 12.024$. Test the null hypothesis at the 5% level of significance. Clearly state the value of the test statistic and the rejection region, along with your conclusion.

(a)

$H_0: \sigma_M^2 = \sigma_F^2 \quad H_a: \sigma_M^2 \neq \sigma_F^2 \quad df_M = 577-4 = 573, \ df_F = 1000-577-4 = 419$

$\hat\sigma_M^2 = \dfrac{SSE_M}{df_M} = \dfrac{97161.9174}{573} = 169.57, \ F^* = \dfrac{\hat\sigma_M^2}{\hat\sigma_F^2} = \dfrac{169.57}{12.024^2} = 1.17$

$F(0.975, 573, 420) = 1.1966, \ F^* < F, \ F^* \in RR$, do not reject H_0

在95% 信心水準下, $\sigma_M^2 = \sigma_F^2$

b. We hypothesize that married individuals, relying on spousal support, can seek wider employment types and hence holding all else equal should have more variable wages. Suppose $\text{var}(e_i|\mathbf{x}_i, MARRIED = 0) = \sigma^2_{SINGLE}$ and $\text{var}(e_i|\mathbf{x}_i, MARRIED = 1) = \sigma^2_{MARRIED}$. Specify the null hypothesis $\sigma^2_{SINGLE} = \sigma^2_{MARRIED}$ versus the alternative hypothesis $\sigma^2_{MARRIED} > \sigma^2_{SINGLE}$. We add *FEMALE* to the wage equation as an explanatory variable, so that

$$WAGE_i = \beta_1 + \beta_2 EDUC_i + \beta_3 EXPER_i + \beta_4 METRO_i + \beta_5 FEMALE + e_i \qquad \text{(XR8.6b)}$$

Using $N = 400$ observations on single individuals, OLS estimation of (XR8.6b) yields a sum of squared residuals is 56231.0382. For the 600 married individuals, the sum of squared errors is 100,703.0471. Test the null hypothesis at the 5% level of significance. Clearly state the value of the test statistic and the rejection region, along with your conclusion.

$\hat\sigma^2_{single} = \dfrac{SSE_{single}}{400-5} = \dfrac{56231.0382}{400-5} \doteq 142.3571, \ \hat\sigma^2_{married} = \dfrac{100703.0471}{600-5} = 169.2488$

$F^* = \dfrac{\hat\sigma^2_{married}}{\hat\sigma^2_{single}} = \dfrac{169.2488}{142.3571} \doteq 1.1889, \ F(0.95, 595, 395) = 1.1647$

$F^* > F, \ F^* \in RR \Rightarrow$ Reject H_0, 在95% 信心水準下

c. Following the regression in part (b), we carry out the NR^2 test using the right-hand-side variables in (XR8.6b) as candidates related to the heteroskedasticity. The value of this statistic is 59.03. What do we conclude about heteroskedasticity, at the 5% level? Does this provide evidence about the issue discussed in part (b), whether the error variation is different for married and unmarried individuals? Explain.

$\chi^2(0.95,4)=9.4877$, $NR^2>\chi^2 \Rightarrow$ Reject H_0, 結婚和單身的樣本存在異質變異性

individuals? Explain.

d. Following the regression in part (b) we carry out the White test for heteroskedasticity. The value of the test statistic is 78.82. What are the degrees of freedom of the test statistic? What is the 5% critical value for the test? What do you conclude?

$df=5-1=4$, $\chi^2(0.95,4)=9.4877$, $78.82>9.4877 \Rightarrow$ Reject H_0,

至少一個(或不只一個)誤差項不為零

e. The OLS fitted model from part (b), with usual and robust standard errors, is

$$\begin{array}{lllllll}
\widehat{WAGE} = & -17.77 & +\ 2.50EDUC & +\ 0.23EXPER & +\ 3.23METRO & -\ 4.20FEMALE \\
\text{(se)} & (2.36) & (0.14) & (0.031) & (1.05) & (0.81) \\
\text{(robse)} & (2.50) & (0.16) & (0.029) & (0.84) & (0.80)
\end{array}$$

For which coefficients have interval estimates gotten narrower? For which coefficients have interval estimates gotten wider? Is there an inconsistency in the results?

窄: EXPER, METRO, FEMALE $\Rightarrow$ 無一致性

寬: EDUC, 截距

f. If we add *MARRIED* to the model in part (b), we find that its t-value using a White heteroskedasticity robust standard error is about 1.0. Does this conflict with, or is it compatible with, the result in (b) concerning heteroskedasticity? Explain.

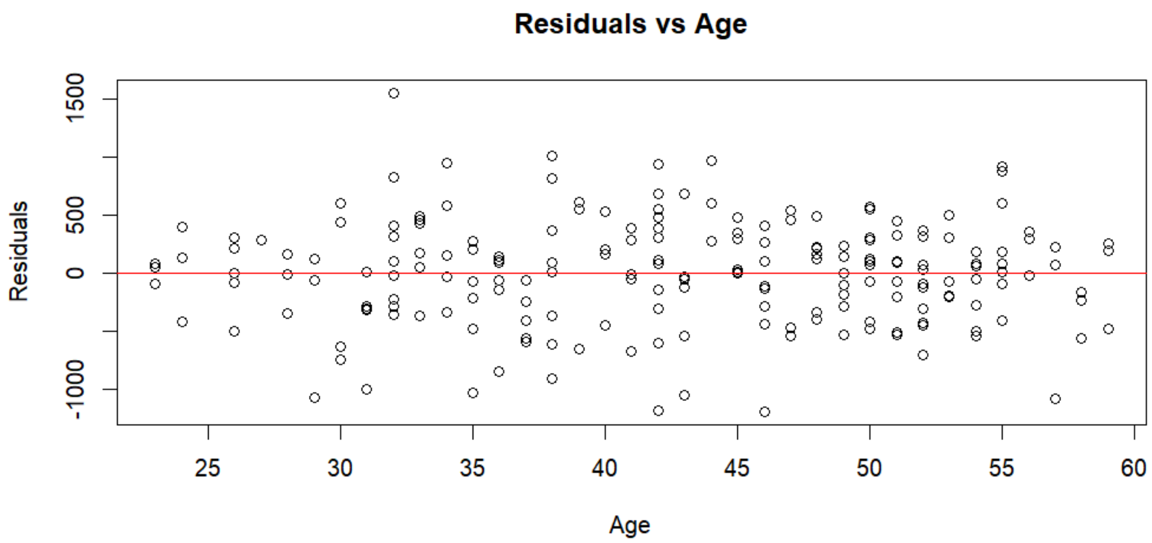
與(b)之結論不衝突 $\Rightarrow$ white estimator 可用於估計同/異質變異性

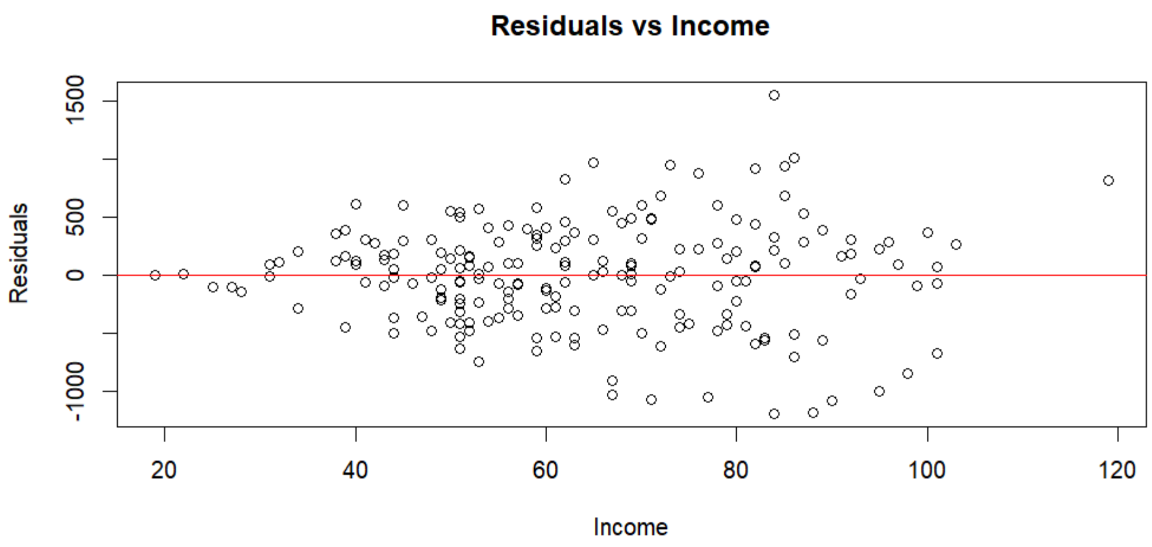
Q16

(a) kids係數的95%信賴區間

```
> confint(ols_model, "kids", level = 0.95)
          2.5 %    97.5 %
kids -135.3298 -28.32302
```

(b)殘差與Income、Age關係圖





(c)p-value<0.05，拒絕虛無假設:不存在異質變異性。即可能存在異質變異性(Heterskedastic)

```
	Goldfeld-Quandt test

data:  miles ~ income + age + kids
GQ = 3.4142, df1 = 76, df2 = 76, p-value = 1.106e-07
alternative hypothesis: variance increases from segment 1 to 2
```

(d)使用Robust SE重新估計Kids的係數95%信賴區間，此區間範圍較a小題廣一點

```
> # kids 的信賴區間（手動計算）
> beta_kids <- coef(ols_model)["kids"]
> se_kids <- sqrt(vcovHC(ols_model, type = "HC1")["kids", "kids"])
> ci_robust <- beta_kids + c(-1, 1) * 1.96 * se_kids
> ci_robust
[1] -138.96900  -24.68383
```

(e)使用GLS估計，範圍被縮小，可能表示估計更精準。

```
> gls_ci$coef["kids", ]
     lower        est.      upper
-119.89450  -76.80629  -33.71808
```

Q18

(a)統計量不落在拒絕域，無充足證具拒絕虛無假說，即無法證明男女變異數不同。

```
> cat("F統計量值：", F_stat,"\n")
F統計量值： 1.05076
> cat("5%顯著水準的臨界值：", F_critical_lower,"和", F_critical_upper, "\n")
5%顯著水準的臨界值： 0.9452566 和 1.058097
```

(b)紙放入三個變數跟所有變數的結果皆在1%信心水準拒絕虛無假設，代表薪水的變異數的確因膚色、性別、居住地有所差別。且存在異質變異性。

```
使用 METRO、FEMALE、BLACK:
> cat("NR^2:", NR2, "\n")
NR^2: 23.55681
> cat("1%顯著水準的χ²臨界值 =", (
1%顯著水準的χ²臨界值 = 11.34487
> cat("使用所有解釋變數: \n")
使用所有解釋變數:
> cat("NR^2:", NR2_all, "\n")
NR^2: 109.4243
> cat("1%顯著水準的χ²臨界值 =", (
1%顯著水準的χ²臨界值 = 21.66599
```

(c)R語言中沒有white test故使用擴增型BP-Test來模擬White Test，結論與(b)相同，統計量落在拒絕域，即數據存在在5%顯著水準下存在異質變異性。

```
> cat("White 檢定 NR^2:", white_test$statistic, "\n")
White 檢定 NR^2: 194.4447
> cat("自由度:", white_test$parameter, "\n")
自由度: 44
> cat("5% 顯著水準的臨界值:", critical_value_white, "\n")
5% 顯著水準的臨界值: 60.48089
```

(d)female、black、metro以及midwest之Robust_SE變小，代表CI範圍也變窄(可更精準預測)其餘變數則標準誤變大、CI變寬。

	OLS_SE	OLS_Robust_SE
(Intercept)	3.211489e-02	3.277743e-02
educ	1.758260e-03	1.904848e-03
exper	1.300342e-03	1.314237e-03
I(exper^2)	2.635448e-05	2.758278e-05
female	9.529136e-03	9.483417e-03
black	1.694240e-02	1.608548e-02
metro	1.230675e-02	1.157624e-02
south	1.356134e-02	1.389454e-02
midwest	1.410367e-02	1.371725e-02
west	1.440237e-02	1.454941e-02

(e)與Robust_SE比較，female、black、metro以及midwest的FGLS_SE變大，CI變寬，其他變數則呈現相反的趨勢。

	OLS_Robust_SE	FGLS_SE
(Intercept)	3.277743e-02	3.184437e-02
educ	1.904848e-03	1.761461e-03
exper	1.314237e-03	1.298873e-03
I(exper^2)	2.758278e-05	2.657195e-05
female	9.483417e-03	9.505454e-03
black	1.608548e-02	1.696582e-02
metro	1.157624e-02	1.186360e-02
south	1.389454e-02	1.354227e-02
midwest	1.371725e-02	1.404549e-02
west	1.454941e-02	1.438967e-02

(f)比較OLS與FGLS的Robust_SE，所有變數的FGLS_ROBUST_SE都變小了

若比較FGLS_SE與FDLS_ROBUST_SE則只有female、black、metro以及midwest變小

Variable	OLS_Robust_SE	FGLS_Robust_SE
(Intercept)	3.277743e-02	3.250910e-02
educ	1.904848e-03	1.895323e-03
exper	1.314237e-03	1.307055e-03
I(exper^2)	2.758278e-05	2.744395e-05
female	9.483417e-03	9.445177e-03
black	1.608548e-02	1.595853e-02
metro	1.157624e-02	1.155933e-02
south	1.389454e-02	1.384176e-02
midwest	1.371725e-02	1.369010e-02
west	1.454941e-02	1.450663e-02

	FGLS_SE	FGLS_Robust_SE
(Intercept)	3.184437e-02	3.250910e-02
educ	1.761461e-03	1.895323e-03
exper	1.298873e-03	1.307055e-03
I(exper^2)	2.657195e-05	2.744395e-05
female	9.505454e-03	9.445177e-03
black	1.696582e-02	1.595853e-02
metro	1.186360e-02	1.155933e-02
south	1.354227e-02	1.384176e-02
midwest	1.404549e-02	1.369010e-02
west	1.438967e-02	1.450663e-02

(g)

應選擇FGLS_ROBUST，因SE變小，代表可能修正了原本過高的估計值