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.

$$\begin{cases} H_0: \sigma_M^2 = \sigma_F^2 \\ H_1: \sigma_M^2 \neq \sigma_F^2 \end{cases}$$

已知$n_M = 577$，$SSE_M = 97161.9174$，$df_M = 577 - 4 = 573$

而$n_F = 1000 - 577 = 423$，$df_F = 423 - 4 = 419$， $\hat{\sigma}_F = 12.024$ ， $\hat{\sigma}_F^2 = 12.024^2$

$\hat{\sigma}_M = \frac{SSE_M}{df_M} = \frac{97161.9174}{573} = 169.567$，在已知$\hat{\sigma}_F^2 = 12.024^2$，F=$\frac{\hat{\sigma}_M^2}{\hat{\sigma}_F^2} = \frac{169.567}{144.5766} = 1.1729$

設拒絕域(雙尾，$\alpha = 5\%$)：$F < F_{0.025,573,419} = 0.836$或$F > F_{0.975,573,419} = 1.1968$

而因 F=1.1729<$F_{0.025,573,419} = 1.1968$，因此不拒絕$H_0$

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 $\text{var}(e_i|\mathbf{x}_i, MARRIED = 0) = \sigma_{SINGLE}^2$ and $\text{var}(e_i|\mathbf{x}_i, MARRIED = 1) = \sigma_{MARRIED}^2$. Specify the null hypothesis $\sigma_{SINGLE}^2 = \sigma_{MARRIED}^2$ versus the alternative hypothesis $\sigma_{MARRIED}^2 > \sigma_{SINGLE}^2$. We add *FEMALE* to the wage equation as an explanatory variable, so that

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

$$\begin{cases} H_0: \sigma_{signe}^2 = \sigma_{married}^2 \\ H_1: \sigma_{single}^2 < \sigma_{married}^2 \end{cases}$$

已知 single：$n_s = 400$，$SSE_s = 56231.0382$， $df_F = 400 - 5 = 395$

已知 married：$n_m = 600$，$SSE_m = 100702.0471$， $df_m = 600 - 5 = 595$

$MSE_s = \frac{56231.0382}{395} = 142.3571$ ，$MSE_m = \frac{100703.0471}{595} = 169.2488$

$$F = \frac{MSE_m}{MSE_s} = \frac{169.2488}{142.3571} = 1.1889$$

設拒絕域(單尾，$\alpha = 5\%$)：或$F > F_{0.95,,595,395} = 1.22$

而因 F=1.1189<$F_{0.95,595,395} = 1.22$，因此不拒絕$H_0$

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.

$\begin{cases} H_0\text{:}homoskedasticity \\ H_1\text{:}heteroskedasticity \end{cases}$ 已知$NR^2$統計量為 59.03 ，N=100，df=4

$\chi^2_{0.94,4} = 9.488$ 而 59.03>9.488 拒絕$H_0$

此結果也支持 (b) 中，誤差變異可能因婚姻狀態（透過 FEMALE、METRO、EXPER 等）而不同

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?

$\begin{cases} H_0\text{:}homoskedasticity \\ H_1\text{:}heteroskedasticity \end{cases}$ 已知 White test statistic = 78.82

因解釋變數=4 因此輔助回歸自變數數量 =p(原)+p(平方)+$\binom{P}{2}$=4+4+6=14

而臨界值$\chi^2_{0.94,14} = 23.685$ 78.82>23.685 因此拒絕$H_0$

表示 White test 也顯示強烈的異質變異存在。

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

$$\widehat{WAGE} = -17.77 + 2.50EDUC + 0.23EXPER + 3.23METRO - 4.20FEMALE$$

| | | | | | |
|---|---|---|---|---|---|
| (se) | (2.36) | (0.14) | (0.031) | (1.05) | (0.81) |
| (robse) | (2.50) | (0.16) | (0.029) | (0.84) | (0.80) |

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

截距變寬，EDUC 變寬，EXPER 變窄，METRO 變窄，FEMALE 變窄

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.

t≈1.0（不顯著）並不衝突於誤差變異存在差異的發現。

一個變數可以對 平均值 沒有顯著影響，卻同時影響 誤差的散布。因此這兩個結果是完全相容的

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

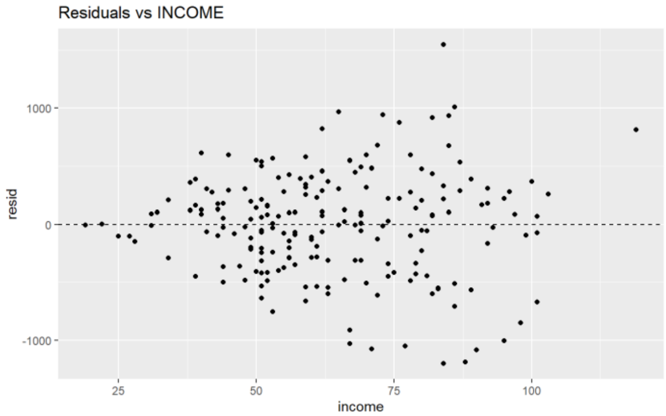
$$MILES = \beta_1 + \beta_2 INCOME + \beta_3 AGE + \beta_4 KIDS + e$$

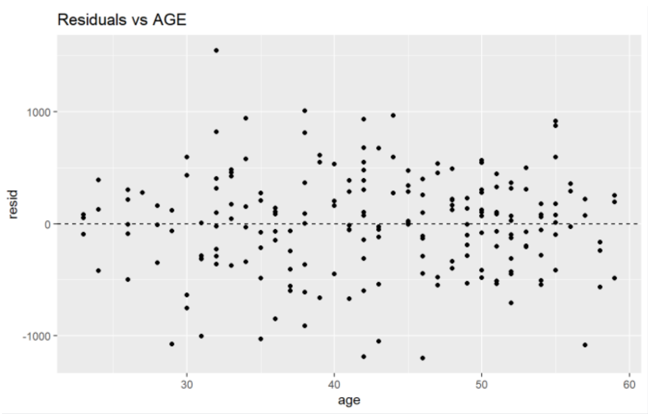
*MILES* is miles driven per year, *INCOME* is measured in \$1000 units, *AGE* is the average age of the adult members of the household, and *KIDS* is the number of children.

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.

```
[1] "=== (a) 常規 95% CI for kids ==="
> print(ci_a)
         2.5 %    97.5 %
kids -135.3298 -28.32302
```

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




從 INCOME 的殘差可以看出呈現漏斗形呈漏斗形（離群隨 X 增大散布加寬），則暗示異質變異。

而 AGE 的殘差沒有

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.

$$\begin{cases} H_0: \sigma^2_{low} = \sigma^2_{high} \\ H_1: \sigma^2_{low} < \sigma^2_{high} \end{cases}$$

```
F_stat = 3.1041  (df1=86, df2=86)
> cat(sprintf("F_crit(0.95, %d, %
t))
F_crit(0.95, 86, 86) = 1.4286
> if(F_stat > F_crit) {
+   cat("Result: Reject H0 → 存在異
+ } else {
+   cat("Result: Fail to reject H(
+ }
Result: Reject H0 → 存在異質變異
```

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)?

```
=== (d) kids 的 95% robust-SE 信賴區間 ===
> print(CI_d)
   Lower95     Upper95
-139.32297  -24.32986
```

```
[1] "=== (a) 常規 95% CI for kids ==="
> print(ci_a)
          2.5 %     97.5 %
kids -135.3298 -28.32302
```

可以知道多一個孩子的效果，在考慮異質變異後，比常規 OLS 顯得更不精確（區間變寬）

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)?

```
=== (a) 常規 OLS 95% CI for kids ===
> print(ci_a)          # 來自前面 (a)
          2.5 %     97.5 %
kids -135.3298 -28.32302
> cat("\n=== (d) 95% Robust-SE CI for kids ===\n")

=== (d) 95% Robust-SE CI for kids ===
> print(CI_d)          # 來自前面 (d)
   Lower95     Upper95
-139.32297  -24.32986
> cat("\n=== (e) GLS 常規 SE 95% CI for kids ===\n")

=== (e) GLS 常規 SE 95% CI for kids ===
> print(ci_gls_conv)
          2.5 %     97.5 %
kids -119.8945 -33.71808
> cat("\n=== (e) GLS Robust-SE 95% CI for kids ===\n")

=== (e) GLS Robust-SE 95% CI for kids ===
> print(ci_gls_robust)
   Lower95     Upper95
-121.41339  -32.19919
```

a 的 OLS 常規的 CI:低估變異數，不夠保守

d 的 OLS Robust-SE 的 CI:因考慮整組變質變異，最保守、最寬

e 的 GLS 常規的 CI:正確加權後的常規 SE 通常介於 a 跟 d 之間，反映模型以部分移除異質變異影響

區間比 a,d 窄

d 的 GLS Robust-SE 的 CI:在正確加權的基礎上再做 robust 調整，與 GLS 常規的 CI 接近，稍微寬一點

但還是比 OLS Robust-SE 的 CI 窄

8.18

Consider the wage equation,

$$\ln(WAGE_i) = \beta_1 + \beta_2 EDUC_i + \beta_3 EXPER_i + \beta_4 EXPER_i^2 + \beta_5 FEMALE_i + \beta_6 BLACK + \beta_7 METRO_i + \beta_8 SOUTH_i + \beta_9 MIDWEST_i + \beta_{10} WEST + e_i$$

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. Use the data file *cps5* for the exercise.

a. Estimate the model by OLS. Carry out the $NR^2$ test using the right-hand-side variables *METRO*, *FEMALE*, *BLACK* as candidates related to the heteroskedasticity. What do we conclude about heteroskedasticity, at the 1% level? Do these results support your conclusions in (a)? Repeat the test using all model explanatory variables as candidates related to the heteroskedasticity.

```
#-----------------------------------------------
# (a) F 檢定：male vs. female 的誤差變異是否相同？
#  H0: σ²_M = σ²_F    H1: σ²_M ≠ σ²_F
#-----------------------------------------------
(a) H0: σ²_M = σ²_F vs. H1: σ²_M ≠ σ²_F
cat(sprintf("     F = %.4f, rejection if F < %.4f or F > %.4f\n\n",
            F_a, Fcrit_a_low, Fcrit_a_high))
   F = 0.6915, rejection if F < 0.9453 or F > 1.0581
```

b. Estimate the model by OLS. Carry out the $NR^2$ test using the right-hand-side variables *METRO*, *FEMALE*, *BLACK* as candidates related to the heteroskedasticity. What do we conclude about heteroskedasticity, at the 1% level? Do these results support your conclusions in (a)? Repeat the test using all model explanatory variables as candidates related to the heteroskedasticity.

```
> # (b) NR² Test (LM test)
> #    H0: homoskedasticity   H1: heteroskedasticity

 (b1) NR² (metro,female,black): 5.6964  vs. χ²0.99,df=3 = 11.3449
> cat("      Conclusion:",
+     if(NR2_1>crit1) "Reject H0\n\n" else "Fail to reject H0\n\n")
      Conclusion: Fail to reject H0


(b2) NR² (all vars): 24.5347  vs. χ²0.99,df=9 = 21.666
 cat("      Conclusion:",
     if(NR2_2>crit2) "Reject H0\n\n" else "Fail to reject H0\n\n")
     Conclusion: Reject H0
```

c. Carry out the White test for heteroskedasticity. What is the 5% critical value for the test? What do you conclude?

```
#-----------------------------------------------
# (c) White Test
#    H0: homoskedasticity   H1: heteroskedasticity
#-----------------------------------------------
```

參數數量: (原)+(平方)+(交乘項) =4+4+6=14

```
(c) White test statistic = 46.23 , df = 54 , χ²0.95 = 72.15
cat("      Conclusion:",
    if(White>critW) "Reject H0\n\n" else "Fail to reject H
    Conclusion: Fail to reject H0
```

d. Estimate the model by OLS with White heteroskedasticity robust standard errors. Compared to OLS with conventional standard errors, for which coefficients have interval estimates gotten narrower? For which coefficients have interval estimates gotten wider? Is there an inconsistency in the results?

```
(Intercept): conv=3.8028, robust=3.7854 → narrower
educ       : conv=0.2082, robust=0.2444 → wider
exper      : conv=0.1540, robust=0.1348 → narrower
I(exper^2) : conv=0.0031, robust=0.0027 → narrower
female     : conv=1.1284, robust=1.1195 → narrower
black      : conv=2.0062, robust=1.5945 → narrower
metro      : conv=1.4573, robust=1.0937 → narrower
south      : conv=1.6058, robust=1.5985 → narrower
midwest    : conv=1.6701, robust=1.6911 → wider
west       : conv=1.7054, robust=1.8185 → wider
```

e. Obtain FGLS estimates using candidate variables *METRO* and *EXPER*. How do the interval estimates compare to OLS with robust standard errors, from part (d)?

```
=== (e) FGLS conventional 95% CI ===
> print(ci_e_conv)
          lower      est.     upper
exper 0.5556566 0.6326549 0.7096532
metro 2.4831456 3.2118362 3.9405268
```

跟 d 比較可以看到 e 更窄，反映模型正確加權後的效率提升

f. Obtain FGLS estimates with robust standard errors using candidate variables *METRO* and *EXPER*. How do the interval estimates compare to those in part (e) and OLS with robust standard errors, from part (d)?

```
=== (f) FGLS + robust-SE 95% CI (使用 lm + weights) ===
> print(ci_fgls2_rob)
        Lower95   Upper95
exper 0.5643251 0.6991606
metro 2.6687969 3.7624798
```

與 e 的 FGLS 常規 CI 相近，但稍微寬一點，可是還是比 PLS RobustCI 還要窄，因為同時考慮加權&robust 調整

g. If reporting the results of this model in a research paper which one set of estimates would you present? Explain your choice.

(g) 建議報告：OLS + White heteroskedasticity-robust SE

建議呈現上面的形式，因為 OLS+White hetroskedasticity-robusr 不需要假設誤差異質變異的具體函數形式&提供對平均效應最保守以及可靠的標準誤估計&在實務上也是最常見以及最具說服力的作法