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 $\hat{\sigma}_F^2 = 12.024$. 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.

Ans.

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

$\alpha = 0.05 \qquad df_M = n_M - k = 577 - 4 = 573 \qquad df_F = n_F - k = (1000 - 577) - 4 = 419$

$$\hat{\sigma}_M^2 = \frac{SSE_M}{df_M} = \frac{97161.9174}{573} = 169.57$$

$$F = \frac{\hat{\sigma}_M^2}{\hat{\sigma}_F^2} = \frac{169.57}{12.024^2} = 1.17$$

```
> alpha <- 0.05
> df_m <- 573  # 男性樣本數
> df_f <- 419  # 女性樣本數
> f_lower <- qf(alpha/2, df_m, df_f) # F左尾臨界值
> f_upper <- qf(1-alpha/2, df_m, df_f)  # F右尾臨界值
> cat(f_lower, f_upper)
0.837669 1.196781
```

由於計算得到的 F 統計量為 1.17，沒有落在拒絕域（0.837669 < 1.17 < 1.196781），因此我們無法拒絕虛無假設 $H_0: \sigma_M^2 = \sigma_F^2$。

結論：在 5% 顯著水平下，沒有足夠的證據表明男性和女性的薪資變異存在顯著差異。也就是說，控制教育、經驗和居住地（METRO）後，男性和女性的薪資波動程度基本相同。

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 $\text{var}(e_i \mid \mathbf{x}_i, MARRIED = 0) = \sigma_{SINGLE}^2$ and $\text{var}(e_i \mid \mathbf{x}_i, MARRIED = 1) = \sigma_{MARRIED}^2$. 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

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

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.

Ans.

$H_0: \sigma_{SINGLE}^2 = \sigma_{MARRIED}^2 \qquad H_1: \sigma_{SINGLE}^2 < \sigma_{MARRIED}^2$

$\alpha = 0.05 \qquad df_{SINGLE} = n_{SINGLE} - k = 400 - 5 = 395 \qquad df_{MARRIED} = n_{MARRIED} - k = 600 - 5 = 595$

$$\hat{\sigma}_{SINGLE}^2 = \frac{SSE_{SINGLE}}{df_{SINGLE}} = \frac{56231.0382}{395} = 142.36$$

$$\hat{\sigma}_{MARRIED}^2 = \frac{SSE_{MARRIED}}{df_{MARRIED}} = \frac{100703.0471}{595} = 169.25$$

$$F = \frac{\hat{\sigma}_{SINGLE}^2}{\hat{\sigma}_{MARRIED}^2} = \frac{142.36}{169.25} = 0.84$$

```
> alpha <- 0.05
> df_single <- 395  # 單身樣本數
> df_married <- 595  # 結婚樣本數
> f <- qf(alpha, df_single, df_married) # F左尾臨界值
> cat(f)
0.8585867
```

由於 F 統計量 (0.84) > 臨界值 (0.8585867)，落在拒絕域，我們拒絕虛無假設 $H_0: \sigma_{SINGLE}^2 = \sigma_{MARRIED}^2$。

結論：在 5% 顯著水平下，有足夠的證據表明已婚個體的薪資變異性確實高於未婚個體。這支持了研究假設：已婚個體由於有配偶支持，可以尋找更多樣化的就業類型，因此薪資的變異性更大。

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.

Ans.

$H_0: \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ (homoskedasticity) H_1 : not all the α_i in H_0 are zero (hetroskedasticity)

$\alpha = 0.05 \quad S = 5$

$NR^2 = 59.03 \quad \geq \chi^2_{0.95,\,5-1} = 9.49$

```
> qchisq(1-0.05, df = 5-1)
[1] 9.487729
```

由於檢定統計量 59.03 遠大於臨界值 9.49，落在拒絕域，我們強烈拒絕虛無假設。

結論：在 5% 顯著性水平下，數據中存在顯著的異質變異性。NR^2 檢驗表明誤差變異與至少一個解釋變量（EDUC, EXPER, METRO, FEMALE）相關，但無法具體指出是哪個變量。

8.6c 確認存在異質變異，但未直接檢驗 MARRIED，因此不是直接證據。然而，8.6c 的結果（存在異質變異）與 8.6b 的結論（誤差變異隨 MARRIED 變化）間接相容，因為異質變異可能受到多個變量的共同影響，包括 MARRIED 的間接作用。

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?

Ans.

White 檢驗的解釋變量：

- 原始變量：EDUC, EXPER, METRO, FEMALE，共 4 個。
- 平方項：$EDUC^2$, $EXPER^2$, $METRO^2$, $FEMALE^2$，共 4 個。但 METRO 和 FEMALE 是二元變量（0 或 1），因此：$METRO^2$ = METRO、$FEMALE^2$ = FEMALE（因為 $0^2 = 0$, $1^2 = 1$），這些平方項與原始變量相同，會被排除，以避免多重共線性。因此，獨立的平方項只有：$EDUC^2$, $EXPER^2$，共 2 個。
- 交叉項：計算所有變量間的兩兩交叉項：EDUC×EXPER, EDUC×METRO, EDUC×FEMALE, EXPER×METRO, EXPER×FEMALE, METRO×FEMALE 共 6 個交叉項。

總自由度：原始變量（4 個） + 獨立平方項（2 個） + 交叉項（6 個） = 4 + 2 + 6 = 12。

$\alpha = 0.05$ 臨界值：$\chi^2_{0.95,\,12} = 21.026$

$NR^2 = 78.82 \quad \geq \chi^2_{0.95,\,12} = 21.026$

```
> qchisq(1-0.05, df =12)
[1] 21.02607
```

由於檢定統計量 78.82 遠大於臨界值 21.026，落在拒絕域，我們強烈拒絕虛無假設。

結論：在 5% 顯著性水平下，數據中存在顯著的異質變異性。這表明回歸模型中存在異質變異，即誤差變異與解釋變量（EDUC, EXPER, METRO, FEMALE）及其平方項、交叉項相關。

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

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

(se) (2.36) (0.14) (0.031) (1.05) (0.81)

(robse) (2.50) (0.16) (0.029) (0.84) (0.80)

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

Ans.

變數	普通標準誤(se)	穩健標準誤(robse)	變化
常數項	2.36	2.50	變寬
EDUC	0.14	0.16	變寬
EXPER	0.031	0.029	變窄
METRO	1.05	0.84	變窄
FEMALE	0.81	0.80	變窄

EXPER, METRO, 和 FEMALE 的標準誤在使用穩健標準誤時變窄了→信賴區間變窄。

截距和 EDUC 的標準誤在使用穩健標準誤時變寬了→信賴區間變寬。

結果不存在不一致。異質變異不影響 OLS 係數估計的一致性，係數估計值仍然一致。標準誤差的差異是因為異質變異的存在，普通標準誤不再正確，而穩健標準誤提供了更可靠的估計。

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.

Ans.

自由度很大（樣本數 $N = 1000$，自由度約為 1000 − 6 = 994），t 分布近似於標準正態分布，5% 顯著性水平的雙尾臨界值約為 ±1.96。$|t| = 1.0 < 1.96$，因此 MARRIED 的係數在 5% 顯著性水平下不顯著。這表明 MARRIED 對工資的直接影響（在控制其他變量後）不顯著，MARRIED 的影響可能已被其他變量解釋。

8.6b、8.6f 兩個結論並不衝突，因為它們檢驗的是不同的事情，結論不矛盾：

- 8.6b 檢驗的是誤差變異是否與 MARRIED 相關（異質變異檢驗）。
- 8.6f 檢驗的是 MARRIED 是否直接影響工資（係數顯著性檢驗）。

8.6b、8.6f 兩個結論是相容的，結果可以同時成立。