HW0414_313707002_羅芷羚

8.6 a.

null hypothesis $H_0 : \sigma_M^2 = \sigma_F^2$

alternative hypothesis $H_1 : \sigma_M^2 \neq \sigma_F^2$

$\hat{\sigma}_M^2 = SSE_M/(n_M - k) = 97161.9174/(577 - 4) = 169.567$

test statistic: $F = \dfrac{\hat{\sigma}_M^2}{\hat{\sigma}_F^2} = \dfrac{169.567}{12.024^2} = 1.17285$

RR=$\{F^* \mid F^* > 1.19614 \ or \ F^* < 0.83804\}$

We fail to reject the null hypothesis, indicating that there is no statistically significant difference in the error variances between males and females.

b.

null hypothesis $H_0 : \sigma_{MARRIED}^2 = \sigma_{SINGLE}^2$

alternative hypothesis $H_1 : \sigma_{MARRIED}^2 > \sigma_{SINGLE}^2$

$\hat{\sigma}_{SINGLE}^2 = SSE_{SINGLE}/(400 - 5) = 56231.0382/395 = 142.357$

$\hat{\sigma}_{MARRIED}^2 = SSE_{MARRIED}/(600 - 5) = 100{,}703.0471/595 = 169.2488$

test statistic: $F = \dfrac{\hat{\sigma}_{SINGLE}^2}{\hat{\sigma}_{MARRIED}^2} = \dfrac{169.2488}{142.357} = 1.1889$

RR=$\{F^* \mid F^* > 1.1647\}$

We reject the null hypothesis, indicating that the variance of the error term is not constant and is systematically related to the explanatory variables

c.

null hypothesis $H_0 : \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$

alternative hypothesis H_1 : not all $\alpha_i = 0$ for i=2,3,4,5

$NR^2 = 59.03 > \chi^2_{0.95,4} = 9.488$

So we reject the null hypothesis of homoskedasticity in the pooled regression.

d.

original variable : *EDUC,EXPER,METRO,FEMALE*

quadratic variable : $EDUC^2$,$EXPER^2$,~~METRO~~*(indicator variable)*,~~FEMALE~~*(indicator variable)*

cross-products:4×3/2=6

There are 12 degrees of freedom.

$\chi^2_{0.95,12} = 21.026$ so we reject the null hypothesis of homoskedasticity in the pooled regression

e.

The confidence intervals for the intercept and the EDUC coefficient have expanded, while those for the remaining variables have become narrower. This outcome is not contradictory, as heteroskedasticity-robust standard errors can be either larger or smaller than the conventional OLS standard errors, which may be biased under the incorrect assumption of homoskedasticity.

f.

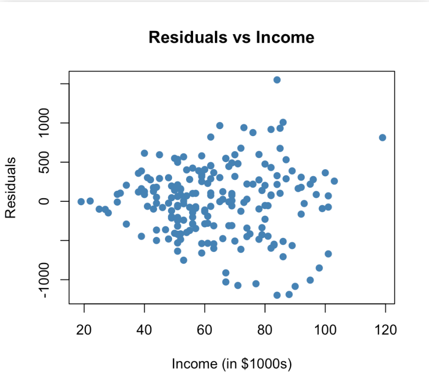
Including the dummy variable MARRIED in the model allows us to examine whether expected wages, after controlling for EDUC, EXPER, METRO, and FEMALE, differ between married and unmarried individuals. The results indicate no statistically significant difference in expected wages between the two groups. In contrast, part (b) focused on whether the variance of wages differs between married and unmarried individuals. That analysis did reveal a significant difference in variability. These two questions address distinct aspects of the model: one concerns the average outcome, and the other concerns the dispersion around that average.

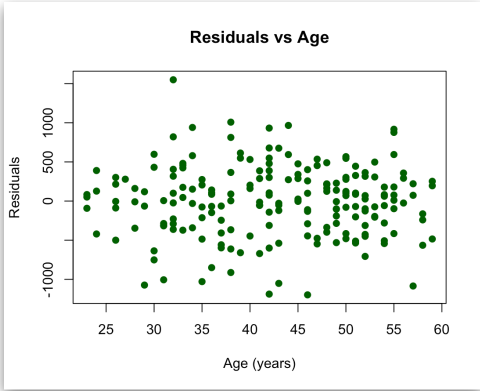
8.1

```
> summary(model_ols)

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

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
```

8.16 b.





The residual plot against income shows the variance of the residuals may increase with income. This suggests the presence of heteroskedasticity. In contrast, the residuals plotted against ageappear fairly homoscedastic, with no clear trend or pattern. These visual patterns provide preliminary evidence that error variance may depend on income.

```
> summary(model_high)

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

Residuals:
     Min       1Q   Median       3Q      Max 
-1215.44  -426.21    73.56   304.71  1602.70 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) -476.803    548.833  -0.869   0.3874   
income        15.556      5.450   2.855   0.0054 **
age           16.388      7.385   2.219   0.0291 * 
kids        -116.017     49.861  -2.327   0.0223 * 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 562 on 86 degrees of freedom
Multiple R-squared:  0.1514,	Adjusted R-squared:  0.1218 
F-statistic: 5.116 on 3 and 86 DF,  p-value: 0.002642
```

```
> summary(model_low)

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

Residuals:
    Min      1Q  Median      3Q     Max 
-684.07 -245.39    8.69  202.87  631.43 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -392.511    214.166  -1.833  0.07030 .  
income        10.960      3.770   2.907  0.00464 ** 
age           18.869      3.783   4.988 3.14e-06 ***
kids         -70.371     29.138  -2.415  0.01785 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 319 on 86 degrees of freedom
Multiple R-squared:  0.309,	Adjusted R-squared:  0.2849 
F-statistic: 12.82 on 3 and 86 DF,  p-value: 5.31e-07
```

GQ = 3.1041>1.4286=$F_{(0.95,86,86)}$

Thus, we reject null hypothesis and conclude that the error variance depends on income.

8.16 d.

```
> coefci(model_ols, vcov. = robust_se, level = 0.95)["kids", ]
      2.5 %      97.5 %
-139.32297  -24.32986
```

The robust 95% confidence interval for the effect of an additional child is slightly wider than the one based on standard OLS errors.

```
> summary(model_gls)

Call:
lm(formula = miles ~ income + age + kids, data = vacation, 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)  # GLS 信賴區間
                 2.5 %     97.5 %
(Intercept) -664.50116 -185.49119
income        11.02744   16.86718
age           10.75260   22.68240
kids        -119.89450  -33.71808
```

```
> 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
```

The GLS estimates are slightly smaller than the OLS estimates for the INCOME and KIDS variables, but slightly larger for AGE and the intercept. GLS produces smaller standard errors than robust OLS, resulting in narrower confidence intervals. Applying robust standard errors to GLS further reduces the standard errors (except for KIDS), and they remain smaller than those from robust OLS.

For example, the 95% confidence interval for KIDS using GLS is [-119.89, -33.72], and with robust GLS, it is [-121.41, -32.20], both narrower than the robust OLS interval.

	Coefficient	Std. Error	Confidence Interval
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]

8.18.a

F statistic: 1.0481

Critical region: F < 0.9451 or F > 1.0579

→We fail to reject the null hypothesis. No significant difference in error variances.

8.18.b

Using the NR^2 test with the selected variables (metro, female, and black), we obtain an R^2 of 0.0024 and a test statistic of 23.5568. Since this exceeds the 1% critical value of 11.3449 ($\chi^2(3)$), we reject the null hypothesis of homoskedasticity. When using all explanatory variables, the R^2 increases to 0.0112 and the NR^2 statistic becomes 109.4243, which also exceeds the critical value of 21.6660 ($\chi^2(9)$). Therefore, we again reject the null hypothesis, confirming strong evidence of heteroskedasticity.

8.18.c

Regressing the squared residuals on all regressors, their squares, and selected interactions gives an R^2 of 0.0198 and an NR2 statistic of 194.4447. With 44 regressors, the 5% critical value from the χ2(44) distribution is 60.4809. Since the test statistic exceeds this threshold, we reject the null hypothesis of homoskedasticity.

	OLS_SE	OLS_Robust_SE	SE_Change	OLS_Width	OLS_Robust_Width	CI_Change
(Intercept)	3.211489e-02	3.277743e-02	變大	0.1259036061	0.1285010626	變寬
educ	1.758260e-03	1.904848e-03	變大	0.0068931063	0.0074677917	變寬
exper	1.300342e-03	1.314237e-03	變大	0.0050978780	0.0051523513	變寬
I(exper^2)	2.635448e-05	2.758278e-05	變大	0.0001033205	0.0001081359	變寬
female	9.529136e-03	9.483417e-03	變小	0.0373581454	0.0371789103	變窄
black	1.694240e-02	1.608548e-02	變小	0.0664212100	0.0630617199	變窄
metro	1.230675e-02	1.157624e-02	變小	0.0482475482	0.0453836492	變窄
south	1.356134e-02	1.389454e-02	變大	0.0531660647	0.0544723330	變寬
midwest	1.410367e-02	1.371725e-02	變小	0.0552922035	0.0537772884	變窄
west	1.440237e-02	1.454941e-02	變大	0.0564632233	0.0570397063	變寬

	OLS_Robust_SE	FGLS_SE	SE_Change	OLS_Robust_Width	FGLS_Width	CI_Change
(Intercept)	3.277743e-02	3.184437e-02	變小	0.1285010626	0.124843079	變窄
educ	1.904848e-03	1.761461e-03	變小	0.0074677917	0.006905656	變窄
exper	1.314237e-03	1.298873e-03	變小	0.0051523513	0.005092118	變窄
I(exper^2)	2.758278e-05	2.657195e-05	變小	0.0001081359	0.000104173	變窄
female	9.483417e-03	9.505454e-03	變大	0.0371789103	0.037265303	變寬
black	1.608548e-02	1.696582e-02	變大	0.0630617199	0.066513034	變寬
metro	1.157624e-02	1.186360e-02	變大	0.0453836492	0.046510222	變寬
south	1.389454e-02	1.354227e-02	變小	0.0544723330	0.053091297	變窄
midwest	1.371725e-02	1.404549e-02	變大	0.0537772884	0.055064111	變寬
west	1.454941e-02	1.438967e-02	變小	0.0570397063	0.056413445	變窄

	FGLS_SE	FGLS_Robust_SE	SE_Change	FGLS_Width	FGLS_Robust_Width	CI_Change
(Intercept)	3.184437e-02	3.250910e-02	變大	0.124843079	0.1274490902	變寬
educ	1.761461e-03	1.895323e-03	變大	0.006905656	0.0074304492	變寬
exper	1.298873e-03	1.307055e-03	變大	0.005092118	0.0051241957	變寬
I(exper^2)	2.657195e-05	2.744395e-05	變大	0.000104173	0.0001075916	變寬
female	9.505454e-03	9.445177e-03	變小	0.037265303	0.0370289927	變窄
black	1.696582e-02	1.595853e-02	變小	0.066513034	0.0625640260	變窄
metro	1.186360e-02	1.155933e-02	變小	0.046510222	0.0453173516	變窄
south	1.354227e-02	1.384176e-02	變大	0.053091297	0.0542654167	變寬
midwest	1.404549e-02	1.369010e-02	變小	0.055064111	0.0536708228	變窄
west	1.438967e-02	1.450663e-02	變大	0.056413445	0.0568719873	變寬

8.18.g

Given the evidence of heteroskedasticity established through various diagnostic tests, I would recommend using the robust OLS results. These estimates are reliable under heteroskedasticity and, in this example, all the coefficients—except for WEST—remain statistically significant at the 0.001 level or better. Moreover, using FGLS does not offer any meaningful advantage in this case, as the results are nearly identical and the performance gains from using it are negligible.