Student: Le Thi Phuong Thao
Student ID: 413707007

**Exercise 8.6**

8.6

$Wage_i = \beta_1 + \beta_2 Educ_i + \beta_3 Expert_i + \beta_4 Metro_i + \varepsilon_i$

a) $H_0: \sigma_M^2 = \sigma_F^2$

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

$\sigma_M^2 = SSE_M / N_M K = 97161.9174 / 573 = 169.54$

$F = \sigma_M^2 / \sigma_F^2 = 169.54 / 12.024^2 = 1.1729 \sim F(573, 419)$

Reject region $F > F(573, 419)_{0.975}$ or $F < F(573, 419)_{0.025}$

$F > 1.196781$ or $F < 0.837669$

We do not reject $H_0$

b) $H_0: \sigma_S^2 = \sigma_M^2$ $\quad n_S = 400,\ \Sigma \hat{e}_S^2 = 56231.0382$

$H_1: \sigma_S^2 < \sigma_M^2$ $\quad N_M = 600,\ \Sigma \hat{e}_M^2 = 100{,}703.0471$

$F = \sigma_S^2 / \sigma_M^2 = \dfrac{56231.0382/395}{100{,}703.0471/595} = 0.8411 < F_{(0.05, 395, 595)} = 0.8585867$

We reject $H_0$ at 5% level significance

c) $\chi^2_{(5-1, 0.95)} = 9.488 < nR^2 = 59.03$

we reject $H_0$ that there is no heteroskedasticity

$\Rightarrow$ There is heteroskedasticity

d) Degree of freedom $= 5(5+1) \times \frac{1}{2} - 1 = 14$

$\chi^2_{(14, 0.05)} = 23.68 < 78.82 \Rightarrow$ We reject $H_0$

Results are consistent with (b) and (c)

e) The interval estimates for intercept and Educ have gotten wider and other variables have gotten narrower

There is no inconsistent since the heteroskedasticity is existed

f) There is no contradiction with part (b)

In part (b) we purpose to compare $\sigma^2$ between two groups. In part (f) we expect the status of marriage effects the wage

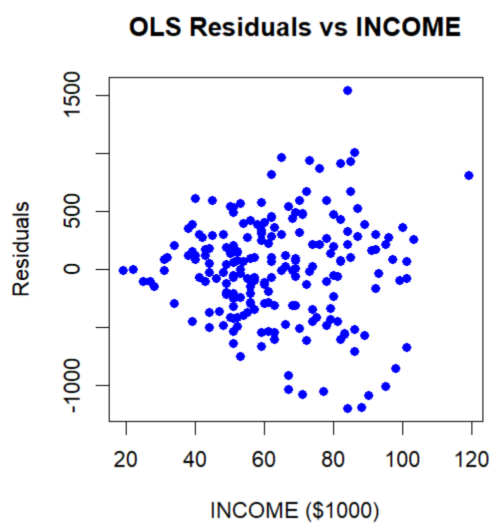
## Exercise 8.16

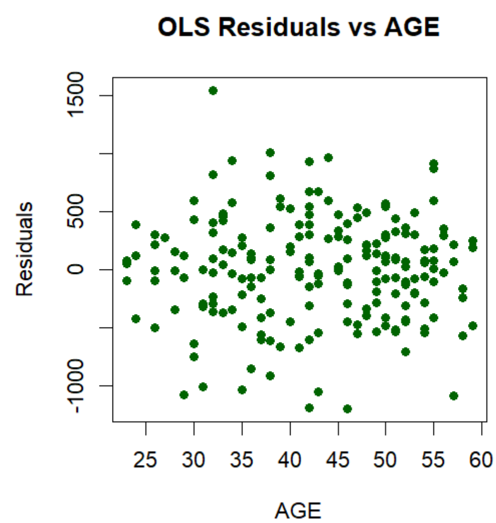
a) [-28.323, -135.330]

```
> confint(model, 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
```

```
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
```

b) Plot




Heteroskedasticity is present

c) Test result

```
> SSE_low  <- sum(resid(model_low )^2)
> SSE_high <- sum(resid(model_high)^2)
> alpha <- 0.05
> df    <- 90 - 4                              # de
> F_stat          <- (SSE_high/df) / (SSE_low/df)
> F_critical_upper <- qf(1 - alpha, df, df)
> print(F_stat)
[1] 3.104061
> print(F_critical_upper)
[1] 1.428617
```

H0: variance of errors is constant for all observations.

H1: variance of errors differs between two subsets of the data (often increasing or decreasing in some regressor)

F value > F critical value. We conclude there are heteroskedastic errors at the 5% level

d) The point estimate is the same in both

There's heteroskedasticity in your data, the robust interval is wider than the one in (a)

```
> print(ci_classic)
[1] -135.32981  -28.32302
> print(ci_robust)
[1] -139.32297  -24.32986
```

e) GLS and GLS robust

```
Table: WLS estimates for the miles equation

|term        |   estimate|  std.error| statistic|   p.value|
|:-----------|----------:|----------:|---------:|---------:|
|(Intercept) | -424.99617| 121.444136| -3.499520| 0.0005769|
|income      |   13.94731|   1.480560|  9.420293| 0.0000000|
|age         |   16.71750|   3.024581|  5.527212| 0.0000001|
|kids        |  -76.80629|  21.848439| -3.515413| 0.0005454|
Table: WLS estimates for the miles equation with robust se

|term        |   estimate| std.error| statistic|   p.value|
|:-----------|----------:|---------:|---------:|---------:|
|(Intercept) | -424.99617| 95.803536| -4.436122| 0.0000153|
|income      |   13.94731|  1.346986| 10.354460| 0.0000000|
|age         |   16.71750|  2.797407|  5.976070| 0.0000000|
|kids        |  -76.80629| 22.618611| -3.395712| 0.0008286|
> confint(model.wls, level = 0.95)
                 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.wls, vcov. = gls_robust_se, level = 0.95)    # GLS  ( robust SE)
                 2.5 %     97.5 %
(Intercept) -613.93428 -236.05807
income        11.29086   16.60376
age           11.20062   22.23438
kids        -121.41339  -32.19919
```

95% interval estimate for the effect of one more child on miles traveled [ -119.89, -33.72] and [ -121.41, -32.19] for robust se

The point estimate is larger than OLS ( -76.8 > -81.8)

**Exercise 8.18**

a) Hypothesis:
H0: σ^2 male=σ^2 female (homoskedasticity)
H1: σ^2 male#σ^2 female (heteroskedasticity)
Test Statistic (F): 0.9489479
Critical Region (5% significance level):
F < 0.9453 or F > 1.0581
We fail to reject the null hypothesis. There is no significant evidence of heteroskedasticity between males and females

b) Results
Case 1: Using only metro, female, and black as candidate variables
Hypothesis:
H0: Error variance is constant and unrelated to metro, female, or black.
H1: Error variance depends on at least one of metro, female, or black.
Test Statistic ($NR^2$): 23.55681 > 11.34487 we reject H0
Case 2: Using all explanatory variables as candidate variables
H0: Error variance is constant and unrelated to all regressors.
H1: Error variance depends on at least one regressor.
Test Statistic ($NR^2$): 109.4243 > 21.66599. We reject the null hypothesis, concluding that the model exhibits significant heteroskedasticity.
Although the Goldfeld–Quandt test in part (a) showed no evidence of unequal variances by gender, the $NR^2$ test results in part (b) reveal that heteroskedasticity is present in the model and may be related to other variables such as metro or black.
Therefore, the evidence from this test complements and expands upon the findings in part (a)

c) White test result p value = 2.2 e^-16 < 0.01 so we conclude there is evidence of heteroskedastic

d) Result

```
> print(ci_comparison)
      Variable Conventional_SE    Robust_SE Conv_CI_Lower Conv_CI_Upper Robust_CI_Lower Robust_CI_Upper Width_Change
1  (Intercept)    3.211489e-02 3.279417e-02  1.1384302204  1.2643338265     1.137098683    1.2656653641    2.1151700
2         educ    1.758260e-03 1.905821e-03  0.0977830603  0.1046761665     0.097493811    0.1049654160    8.3924256
3        exper    1.300342e-03 1.314908e-03  0.0270727569  0.0321706349     0.027044205    0.0321991870    1.1201599
4   I(exper^2)    2.635448e-05 2.759687e-05 -0.0004974407 -0.0003941203    -0.000499876   -0.0003916849    4.7141298
5       female    9.529136e-03 9.488260e-03 -0.1841810529 -0.1468229075    -0.184100928   -0.1469030324   -0.4289553
6        black    1.694240e-02 1.609369e-02 -0.1447358548 -0.0783146449    -0.143072211   -0.0799782888   -5.0093755
7        metro    1.230675e-02 1.158215e-02  0.0948966363  0.1431441846     0.096316998    0.1417238226   -5.8878100
8        south    1.356134e-02 1.390164e-02 -0.0723384657 -0.0191724010    -0.073005508   -0.0185053588    2.5092781
9      midwest    1.410367e-02 1.372426e-02 -0.0915893895 -0.0362971859    -0.090845662   -0.0370409129   -2.6901695
10        west    1.440237e-02 1.455684e-02 -0.0348207138  0.0216425095    -0.035123519    0.0219453146    1.0725746

Coefficients with wider intervals using robust SE:
> print(wider)
[1] "(Intercept)" "educ"        "exper"       "I(exper^2)"  "south"       "west"
>
> cat("\nCoefficients with narrower intervals using robust SE:\n")

Coefficients with narrower intervals using robust SE:
> print(narrower)
[1] "female"  "black"   "metro"   "midwest"
```

Traditional OLS standard errors assume constant error variance (homoskedasticity). If this assumption is violated—for example, if wage variance changes with metro or experience—OLS standard errors can be biased, often underestimated. White's robust standard errors correct for heteroskedasticity by allowing error variance to vary. As a result, robust standard errors are usually larger, leading to wider confidence intervals. Wider intervals suggest that traditional standard errors may be overly optimistic, potentially causing overconfidence in the results

e) Result

```
> print(ci_comparison_fgls)
      Variable      FGLS_SE    Robust_SE FGLS_CI_Lower FGLS_CI_Upper Robust_CI_Lower Robust_CI_Upper Width_Change
1  (Intercept) 3.159320e-02 3.279417e-02  1.1302695254  1.2541279001     1.137098683    1.2656653641  -3.66215144
2         educ 1.764615e-03 1.905821e-03  0.0982024458  0.1051204663     0.097493811    0.1049654160  -7.40917939
3        exper 1.297517e-03 1.314908e-03  0.0275467064  0.0326335081     0.027044205    0.0321991870  -1.32261698
4   I(exper^2) 2.678918e-05 2.759687e-05 -0.0005086498 -0.0004036251    -0.000499876   -0.0003916849  -2.92675532
5       female 9.480830e-03 9.488260e-03 -0.1847977976 -0.1476290326    -0.184100928   -0.1469030324  -0.07831277
6        black 1.699247e-02 1.609369e-02 -0.1441623553 -0.0775448504    -0.143072211   -0.0799782888   5.58466271
7        metro 1.145945e-02 1.158215e-02  0.0953066354  0.1402324253     0.096316998    0.1417238226  -1.05938782
8        south 1.352230e-02 1.390164e-02 -0.0713493481 -0.0183363498    -0.073005508   -0.0185053588  -2.72870976
9      midwest 1.398389e-02 1.372426e-02 -0.0906033967 -0.0357807800    -0.090845662   -0.0370409129   1.89177946
10        west 1.437651e-02 1.455684e-02 -0.0336747637  0.0226870709    -0.035123519    0.0219453146  -1.23885273
Coefficients with wider intervals using FGLS vs. OLS robust:
> print(fgls_wider)
[1] "black"   "midwest"
>
> cat("\nCoefficients with narrower intervals using FGLS vs. OLS robust:\n")

Coefficients with narrower intervals using FGLS vs. OLS robust:
> print(fgls_narrower)
[1] "(Intercept)" "educ"        "exper"       "I(exper^2)"  "female"      "metro"       "south"       "west"
```

OLS (White robust) does not assume an error variance structure and directly adjusts standard errors to address heteroskedasticity, making results more robust but potentially less efficient (wider confidence intervals).

FGLS assumes error variance can be modeled using metro and exper. If the assumption is correct, FGLS is more efficient (narrower confidence intervals); if incorrect, FGLS may perform poorly.

f) Result

```
> print(ci_comparison_robust_fgls)
      Variable      FGLS_SE FGLS_Robust_SE OLS_Robust_SE FGLS_CI_Lower FGLS_CI_Upper FGLS_Robust_CI_Lower
1  (Intercept) 3.159320e-02   3.235961e-02  3.279417e-02  1.1302695254  1.2541279001         1.1287671947
2         educ 1.764615e-03   1.892760e-03  1.905821e-03  0.0982024458  0.1051204663         0.0979512563
3        exper 1.297517e-03   1.304616e-03  1.314908e-03  0.0275467064  0.0326335081         0.0275327897
4   I(exper^2) 2.678918e-05   2.740828e-05  2.759687e-05 -0.0005086498 -0.0004036251        -0.0005098633
5       female 9.480830e-03   9.438075e-03  9.488260e-03 -0.1847977976 -0.1476290326        -0.1847139905
6        black 1.699247e-02   1.586874e-02  1.609369e-02 -0.1441623553 -0.0775448504        -0.1419596068
7        metro 1.145945e-02   1.156288e-02  1.158215e-02  0.0953066354  0.1402324253         0.0951039024
8        south 1.352230e-02   1.383444e-02  1.390164e-02 -0.0713493481 -0.0183363498        -0.0719612018
9      midwest 1.398389e-02   1.371270e-02  1.372426e-02 -0.0906033967 -0.0357807800        -0.0900718148
10        west 1.437651e-02   1.450875e-02  1.455684e-02 -0.0336747637  0.0226870709        -0.0339339956
   FGLS_Robust_CI_Upper OLS_Robust_CI_Lower OLS_Robust_CI_Upper
1          1.2556302309         1.137098683        1.2656653641
2          0.1053716558         0.097493811        0.1049654160
3          0.0326474248         0.027044205        0.0321991870
4         -0.0004024116        -0.000499876       -0.0003916849
5         -0.1477128397        -0.184100928       -0.1469030324
6         -0.0797475989        -0.143072211       -0.0799782888
7          0.1404351583         0.096316998        0.1417238226
8         -0.0177244961        -0.073005508       -0.0185053588
9         -0.0363123619        -0.090845662       -0.0370409129
10         0.0229463028        -0.035123519        0.0219453146
```

```
> # 4) Print summaries
> cat("\n--- Robust FGLS vs. FGLS ---\n")

--- Robust FGLS vs. FGLS ---
> cat("Wider under Robust FGLS:\n");    print(wider_rf_vs_f)
Wider under Robust FGLS:
[1] "(Intercept)" "educ"        "exper"       "I(exper^2)"  "metro"       "south"       "west"
> cat("Narrower under Robust FGLS:\n"); print(narrower_rf_vs_f)
Narrower under Robust FGLS:
[1] "female"  "black"   "midwest"
>
> cat("\n--- Robust FGLS vs. Robust OLS ---\n")

--- Robust FGLS vs. Robust OLS ---
> cat("Wider under Robust FGLS:\n");    print(wider_rf_vs_o)
Wider under Robust FGLS:
character(0)
> cat("Narrower under Robust FGLS:\n"); print(narrower_rf_vs_o)
Narrower under Robust FGLS:
 [1] "(Intercept)" "educ"        "exper"       "I(exper^2)"  "female"      "black"       "metro"       "south"
 [9] "midwest"     "west"
```

For part (d), OLS (White robust) does not assume any variance structure and directly adjusts standard errors to handle heteroskedasticity, typically resulting in wider but more reliable confidence intervals.

FGLS (robust standard errors) may have narrower intervals than OLS (White robust) if the heteroskedasticity model is partially correct, but wider intervals than FGLS in part (e) since robust standard errors account for model errors.

g) Result

```
> # Print the summary table
> print(summary_wide)
             Method (Intercept)_Coef (Intercept)_SE educ_Coef     educ_SE exper_Coef    exper_SE I(exper^2)_Coef
1 OLS Conventional          1.201382     0.03211489 0.1012296 0.001758260 0.02962170 0.001300342   -0.0004457805
2       OLS Robust          1.201382     0.03279417 0.1012296 0.001905821 0.02962170 0.001314908   -0.0004457805
3 FGLS Conventional         1.192199     0.03159320 0.1016615 0.001764615 0.03009011 0.001297517   -0.0004561375
4      FGLS Robust          1.192199     0.03235961 0.1016615 0.001892760 0.03009011 0.001304616   -0.0004561375
  I(exper^2)_SE female_Coef   female_SE black_Coef   black_SE metro_Coef   metro_SE  south_Coef   south_SE
1  2.635448e-05  -0.1655020 0.009529136 -0.1115252 0.01694240  0.1190204 0.01230675 -0.04575543 0.01356134
2  2.759687e-05  -0.1655020 0.009488260 -0.1115252 0.01609369  0.1190204 0.01158215 -0.04575543 0.01390164
3  2.678918e-05  -0.1662134 0.009480830 -0.1108536 0.01699247  0.1177695 0.01145945 -0.04484285 0.01352230
4  2.740828e-05  -0.1662134 0.009438075 -0.1108536 0.01586874  0.1177695 0.01156288 -0.04484285 0.01383444
  midwest_Coef midwest_SE    west_Coef    west_SE
1  -0.06394329 0.01410367 -0.006589102 0.01440237
2  -0.06394329 0.01372426 -0.006589102 0.01455684
3  -0.06319209 0.01398389 -0.005493846 0.01437651
4  -0.06319209 0.01371270 -0.005493846 0.01450875
```

Choosing FGLS (White robust):

FGLS (White robust) has slightly smaller standard errors than OLS (White robust) leading to narrower confidence intervals and slightly higher efficiency.

Additionally, using White robust standard errors ensures that even if the heteroskedasticity model (metro and exper) is misspecified, the standard errors and confidence intervals remain reliable, avoiding the risk of FGLS (traditional standard errors) underestimating uncertainty