# Econometrics HomeWork

*Poongkundran Thamaraiselvan, Irene Sunny, Roshan Velpula*

**Question 1 - Compute standard descriptive statistics for the variables in this study.**

```
# Read in the bwages.dat file
bwages <- read.delim("C:/Users/kingr/Downloads/bwages.dat")
# Get a column-wise summary of the dataset
summary(bwages)
```

```
##       WAGE            LNWAGE            EDUC           EXPER
##  Min.   : 2.191   Min.   :0.7843   Min.   :1.000   Min.   : 0.00
##  1st Qu.: 8.113   1st Qu.:2.0935   1st Qu.:3.000   1st Qu.: 9.00
##  Median :10.127   Median :2.3152   Median :3.000   Median :16.50
##  Mean   :11.051   Mean   :2.3344   Mean   :3.378   Mean   :17.22
##  3rd Qu.:12.755   3rd Qu.:2.5460   3rd Qu.:4.000   3rd Qu.:24.00
##  Max.   :47.576   Max.   :3.8623   Max.   :5.000   Max.   :47.00
##     LNEXPER          LNEDUC           MALE
##  Min.   :0.000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:2.303   1st Qu.:1.099   1st Qu.:0.0000
##  Median :2.862   Median :1.099   Median :1.0000
##  Mean   :2.691   Mean   :1.137   Mean   :0.6067
##  3rd Qu.:3.219   3rd Qu.:1.386   3rd Qu.:1.0000
##  Max.   :3.871   Max.   :1.609   Max.   :1.0000
```

**Question 2 - Compute dummy variables for the educ variable: educ1=1 if primary school, 0 otherwise; educ2=1 if lower vocational training, 0 otherwise, and so forth**

```
# Create dummy variables for the EDUC variable
bwages$educ1 <- ifelse(bwages$EDUC == 1, 1, 0)
bwages$educ2 <- ifelse(bwages$EDUC == 2, 1, 0)
bwages$educ3 <- ifelse(bwages$EDUC == 3, 1, 0)
bwages$educ4 <- ifelse(bwages$EDUC == 4, 1, 0)
bwages$educ5 <- ifelse(bwages$EDUC == 5, 1, 0)
```

```
View(bwages)
```

**Question 3 - Estimate the following model by OLS: wage = β0 + βmale + β1educ1 + β2educ2 + β3educ3 + β4educ4 + β5exper + ε**

```
# Fit the OLS regression model
model <- lm(WAGE ~ MALE + educ1 + educ2 + educ3 + educ4 + EXPER, data = bwages)
```

```
summary(model)
```

```
##
## Call:
## lm(formula = WAGE ~ MALE + educ1 + educ2 + educ3 + educ4 + EXPER,
##     data = bwages)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.6859  -1.8975  -0.2973   1.5425  30.5328
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.401010   0.259946  40.012  < 2e-16 ***
## MALE         1.300905   0.194696   6.682 3.35e-11 ***
## educ1       -7.904926   0.420969 -18.778  < 2e-16 ***
## educ2       -6.089686   0.298745 -20.384  < 2e-16 ***
## educ3       -4.244575   0.261643 -16.223  < 2e-16 ***
## educ4       -2.417478   0.272185  -8.882  < 2e-16 ***
## EXPER        0.190744   0.009644  19.779  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.548 on 1465 degrees of freedom
## Multiple R-squared:  0.367,  Adjusted R-squared:  0.3644
## F-statistic: 141.5 on 6 and 1465 DF,  p-value: < 2.2e-16
```

**Question 4 - Interpret the coefficients.**

- The intercept (β0) represents the predicted value of the dependent variable when all the independent variables have no effect on it. Hence, 10.401 is the estimated average hourly wage for *females with university education* ('educ5' = 1) and no work experience.
- The β1,β2,β3,β4,β5,β6 are the estimated coefficient or slope of respective variables in a linear regression model.The coefficient represents the expected change in the response variable for a one-unit increase in the predictor variable, holding all other variables constant.
- The coefficient for `MALE` (β1) is 1.301, which means that on average, males earn €1.301 more per hour than females, after controlling for education and work experience.
- Negative coefficients for education dummy variables (β2 - β5) represent strongest downside to the person with the least qualification indicated by 'educ1' and the rest progressively better than the previous levels. Hence, person with most qualification have better wages compared to predecessors. For example, individuals with a primary school education (educ1) earn €7.905 less per hour on average than those with university education ( `educ5` ), all else equal.
- (Note: `educ5` is not included in the model, and the coefficients present the estimated change in the outcome variable (wage) associated with a one-unit change in each predictor variable, holding all other variables in the model constant. Since `educ5` is not included here, it is considered as the 0 reference. )
- The coefficient for `exper` (β6) is positive, which means that on average, individuals with more work experience earn more per hour than those with less experience, after controlling for gender and education. For each additional year of experience, individuals earn €0.191 more per hour on average.

- All of the coefficients are statistically significant at the 1% level, which means that we can reject the null hypothesis that the true coefficients are equal to zero.

**Question 5 - Why is there no educ5 in the model?**

Including all five education levels in the model would result in **perfect multi-collinearity**, which occurs when one predictor variable is a perfect linear combination of other predictor variables in the model. In this case, if we include `educ1` through `educ5` in the model, we would be able to perfectly predict the level of education of an individual based on which dummy variable is equal to one. Multi-Collinearity makes it difficult to interpret the model because of high standard errors and unstable coefficients. To avoid perfect multi-collinearity, it is common practice to exclude one of the categories when creating dummy variables.

**Question 6 - What is the expected salary for a man with university level education?**

```
newdata <- data.frame(MALE = 1, educ1 = 0, educ2 = 0, educ3 = 0, educ4 = 0, EXPER = 0)

expected_salary <- predict(model, newdata)

cat("The expected hourly wage of male wih university level education is: €", expected_salary,
"\n")
```

```
## The expected hourly wage of male wih university level education is: € 11.70191
```

**Question 7 - Using standard t-tests, are the explanatory variables significant?**

*Hypothesis*

- H0: Coefficients are Zero
- H1: Coefficients are not Zero

```
summary(model)$coefficients[, c(3,4) ]
```

```
##                t value      Pr(>|t|)
## (Intercept)  40.012183 3.344591e-237
## MALE          6.681719  3.345334e-11
## educ1       -18.777929  1.162822e-70
## educ2       -20.384234  1.635763e-81
## educ3       -16.222745  1.474319e-54
## educ4        -8.881756  1.883270e-18
## EXPER        19.779071  2.303428e-77
```

The output of the model includes the t-value and the corresponding p-value for each coefficient. Given that the standard p-value is 0.05. Looking at the output, all the explanatory variables have very small p-values (< 0.05), indicating that they are statistically significant. Therefore, we can reject the null hypothesis that the coefficients are zero. We conclude that there is evidence that the explanatory variables MALE, educ1, educ2, educ3, educ4, and EXPER are related to wage and statistically significant.

**Question 8 - Is there a significant constant salary difference between men and woman?**

*Hypothesis*

- H0: There is no significant difference in average salary between men and women.

- H1: There is a significant difference in average salary between men and women.

Yes, there is a significant constant salary difference between men and women. The estimated coefficient for MALE is 1.300905, which is statistically significant with a p-value of 3.35e-11. This means that on average, men earn €1.30 more per hour than women, after controlling for education and work experience. As p-value is extremely small, it shows that with a stronger evidence we can reject the null hypothesis.

**Question 9 - Add squared experience as an additional variable in your model and estimate it again**

```
model2 <- lm(WAGE ~ MALE + educ1 + educ2 + educ3 + educ4 + EXPER + I(EXPER^2), data = bwages)
summary(model2)
```

```
##
## Call:
## lm(formula = WAGE ~ MALE + educ1 + educ2 + educ3 + educ4 + EXPER +
##     I(EXPER^2), data = bwages)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.8842  -1.9112  -0.2553   1.4629  30.4009
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.3199718  0.3218616  28.956  < 2e-16 ***
## MALE         1.2848381  0.1927377   6.666 3.71e-11 ***
## educ1       -7.8258289  0.4169283 -18.770  < 2e-16 ***
## educ2       -6.1480903  0.2958915 -20.778  < 2e-16 ***
## educ3       -4.3017952  0.2591850 -16.597  < 2e-16 ***
## educ4       -2.4621766  0.2695355  -9.135  < 2e-16 ***
## EXPER        0.3595458  0.0316666  11.354  < 2e-16 ***
## I(EXPER^2)  -0.0044611  0.0007979  -5.591 2.69e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.512 on 1464 degrees of freedom
## Multiple R-squared:  0.3802, Adjusted R-squared:  0.3772
## F-statistic: 128.3 on 7 and 1464 DF,  p-value: < 2.2e-16
```

**Question 10 - From the previous question, interpret the impact of experience on salary.**

In this case, the positive coefficient for experience (0.3595458) suggests that wages tend to increase with experience, but the negative coefficient for experience squared (-0.0044611) suggests that the rate of increase starts to slow down and may eventually turn negative as experience increases even further. This type of relationship is often referred to as a “curvilinear” relationship. Also there’s a multicollinearity problem with the above model as ‘experience’ and ‘experience_squared’ are highly correlated and thus leading to unreliable estimates.

Note: This is known as diminishing returns, and it is a common pattern in many fields.The effect of experience on wages becomes less pronounced as experience increases. This could be because workers with a lot of experience have already reached the highest levels of pay within their occupation, or because the skills and knowledge gained through experience become less valuable as time goes on.

**Question 11 - Is the impact of square experience on salary significant? Perform a test.**

*Hypothesis*

- H0: EXPER^2 Coefficient is Zero
- H1: EXPER^2 Coefficients is not Zero

From the coefficients table, we can see that the p-value for the coefficient of I(EXPER^2) is less than 0.05 (p-value = 2.69e-08), which means that the impact of square experience on salary is significant at a 5% significance level. Therefore, we can reject the null hypothesis that the coefficient of I(EXPER^2) is zero and conclude that there is evidence that square experience has a significant impact on salary.

**Question 12 - Until now the effect of gender was assumed constant, no matter the education or experience level. Write down a model that explicitly allows for differences between men and woman.**

To explicitly allow for differences between men and women in the model, we can include an interaction term between gender and one or more of the other independent variables. For example, we can modify the previous model by adding an interaction term between gender, education and experience:

WAGE = β0 + β1MALE + β2educ1 + β3educ2 + β4educ3 + β5educ4 + β6EXPER + β7MALEeduc1 + β8MALEeduc2 + β9MALEeduc3 + β10MALEeduc4 + β11MALEEXPER + ε

In this model, the β7-β11 coefficients represent the interaction effect between gender and the respective explanatory variables. If β7-11 is significant, it suggests that the effect of experience and education on wages differs between men and women.

( Note: We are not considering EXPER^2 in our model, because of the Multicollinearity which leads to unreliable estimates )

```
general_model <- lm(WAGE ~ MALE + educ1 + educ2 + educ3 + educ4  + EXPER +  MALE*educ1 + MALE*ed
uc2 + MALE*educ3 + MALE*educ4 + MALE*EXPER, data = bwages)
```

**Question 13 - Estimate this new model and interpret the coefficients.**

```
summary(general_model)
```

```
##
## Call:
## lm(formula = WAGE ~ MALE + educ1 + educ2 + educ3 + educ4 + EXPER +
##     MALE * educ1 + MALE * educ2 + MALE * educ3 + MALE * educ4 +
##     MALE * EXPER, data = bwages)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.2066  -1.8478  -0.3211   1.5834  30.1323
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.38802    0.36869  28.175  < 2e-16 ***
## MALE         1.29349    0.48523   2.666  0.00777 **
## educ1       -7.69045    0.81963  -9.383  < 2e-16 ***
## educ2       -5.78625    0.54269 -10.662  < 2e-16 ***
## educ3       -3.53112    0.41726  -8.463  < 2e-16 ***
## educ4       -1.96387    0.40112  -4.896 1.09e-06 ***
## EXPER        0.16560    0.01603  10.332  < 2e-16 ***
## MALE:educ1  -0.47504    0.95737  -0.496  0.61983
## MALE:educ2  -0.50070    0.65171  -0.768  0.44244
## MALE:educ3  -1.17034    0.53517  -2.187  0.02891 *
## MALE:educ4  -0.78444    0.54793  -1.432  0.15246
## MALE:EXPER   0.04017    0.02008   2.000  0.04564 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.543 on 1460 degrees of freedom
## Multiple R-squared:  0.3709, Adjusted R-squared:  0.3662
## F-statistic: 78.26 on 11 and 1460 DF,  p-value: < 2.2e-16
```

In this general model, we can observe

β7-β10 represent the interaction effects between gender and education, which capture the estimated differences in the effects of education between men and women, all else being equal. For example, the effect of being a male with a high school diploma (MALE:educ1) is estimated to be -0.47504, meaning that the negative effect of having a high school diploma on wages is less severe for men compared to women

β11 represents the interaction effect between gender and experience, which captures the estimated difference in the effect of experience between men and women, all else being equal. The effect of this interaction (MALE:EXPER) is estimated to be 0.04017, meaning that the positive effect of experience on wages is slightly stronger for men compared to women.

**Question 14 - Test for differences between men and woman in this general model.**

To test the differences between Men and Women, we consider the following hypothesis.

*Hypothesis*

- H0: β7-β11 Coefficients are Zero (No significant difference between Men and Women)
- H1: β7-β11 Coefficients are not Zero (Difference between Men and Women)

```
reduced_model <- lm(WAGE ~ MALE + educ1 + educ2 + educ3 + educ4 + EXPER , data = bwages)
anova(reduced_model, general_model)
```

| | Res.Df <dbl> | RSS <dbl> | Df <dbl> | Sum of Sq <dbl> | F <dbl> | Pr(>F) <dbl> |
|---|---|---|---|---|---|---|
| 1 | 1465 | 18444.32 | *NA* | *NA* | *NA* | *NA* |
| 2 | 1460 | 18328.76 | 5 | 115.5608 | 1.841027 | 0.1018842 |

2 rows

- Based on the coefficients and t-statistics of the interaction terms, we can conclude that there is a significant difference between men and women in the effect of experience on wages. The coefficient of MALE:EXPER is statistically significant with a t-value of 2.000 and a p-value of 0.04564, indicating that the relationship between experience and wages differs significantly between men and women. Specifically, on average, male workers with one additional year of experience earn an additional 0.04017 dollars compared to female workers with the same level of experience.
- However, the F-test for the overall significance of the model including the interaction terms is not statistically significant with a p-value of 0.1019. This indicates that the improvement in model fit due to the addition of the interaction terms is not statistically significant.
- Therefore, we can conclude that there is evidence of a *significant difference between men and women in the effect of experience on wages*, but the *overall difference in the model is not statistically significant*.

**Question15 - Demonstrate empirically that the sum of squared residuals (SSR) for this general model is equal to the SSR of the model explaining wage by the educ dummies and the experience variables fitted to the sample of men + the SSR of the same model fitted to the sample of woman.**

```
general_model <- lm(WAGE ~ MALE + educ1 + educ2 + educ3 + educ4 + EXPER + MALE*educ1 + MALE*educ
2 + MALE*educ3 + MALE*educ4 + MALE*EXPER,  data = bwages)
```

Splitting our data for Men and Women

```
wage_data_men <- subset(bwages, MALE == 1)
wage_data_women <- subset(bwages, MALE == 0)
```

Modeling for Men

```
model_men <- lm(WAGE ~  educ1 + educ2 + educ3 + educ4 + EXPER, data = wage_data_men)
```

Modeling for Women

```
model_women <- lm(WAGE ~  educ1 + educ2 + educ3 + educ4 + EXPER, data = wage_data_women)
```

*Sum of Squared Residuals for General Model*

```
ssr_general <- sum(general_model$residuals^2)
print(ssr_general)
```

```
## [1] 18328.76
```

*Sum of Squared Residuals for Men's Model*

```
ssr_men <- sum(model_men$residuals^2)
print(ssr_men)
```

```
## [1] 12727.48
```

*Sum of Squared Residuals for Women's Model*

```
ssr_women <- sum(model_women$residuals^2)
print(ssr_women)
```

```
## [1] 5601.283
```

*Checking if SSR MEN + SSR WOMEN = SSR General Model*

```
ssr_sum <- ssr_men + ssr_women
```

```
round(ssr_general,2) == round(ssr_sum, 2)
```

```
## [1] TRUE
```

```
ssr_men_women <- ssr_men + ssr_women
print(all.equal(ssr_general, ssr_men_women))
```

```
## [1] TRUE
```

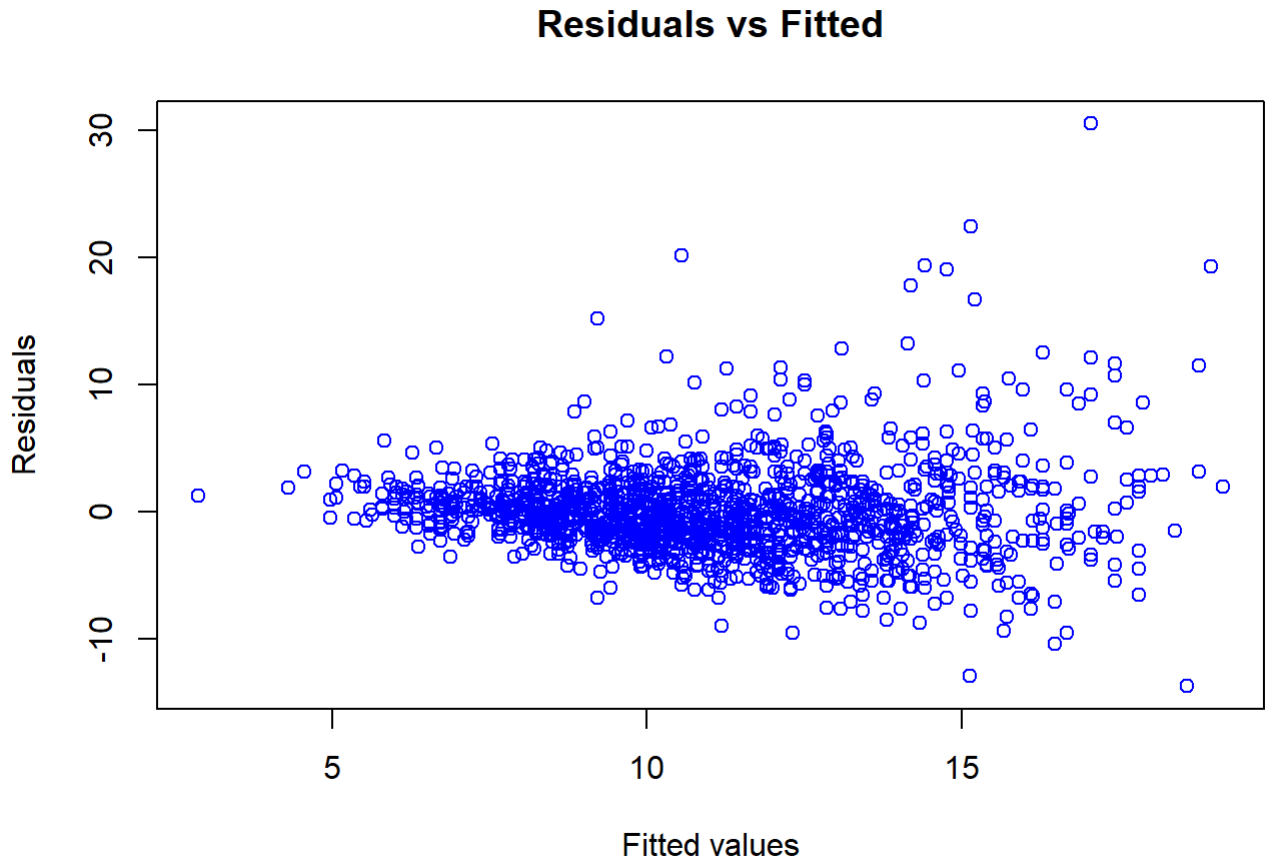
This indicates that the sum of squared residuals (SSR) for the general model with gender interaction term is equal to the sum of SSR for the model with education and experience fitted separately to men and women. This confirms the equivalence between the two models in terms of fitting the data.

**Question 16 - Test for heteroskedasticity of the residuals in (1) that might arise from the fact that men and woman have difference residual variance.**

```
model <- lm(WAGE ~ MALE + educ1 + educ2 + educ3 + educ4 + EXPER, data = bwages)
residuals <- residuals(model)
```

*Plotting Residuals vs Fitted*

```
plot(fitted(model), residuals, main="Residuals vs Fitted", xlab="Fitted values", ylab="Residual
s", col="blue")
```




The above graph suggest heteroskadasticity.

*Testing*

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
# Breusch-Pagan test for heteroskedasticity
bp_test <- bptest(model)
```

```
bp_test
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model
## BP = 99.164, df = 6, p-value < 2.2e-16
```

*Hypothesis*

- H0: Homoskedacity Assumption is True
- H1: Homoskedacity Assumption is Not True

In Breusch-Pagan test, Null hypothesis is Homoskedasticity. And since the p-value is less than 0.05, we can reject the null hypothesis with a strong confidence.

This concludes that there is an evidence for heteroskedasticity.

**Question 17 - If in the previous question the test rejects, then perform generalized least squares.**

In the residuals vs fitted graph,we can clearly see the heteroskedasticity. So OLS is not a 'BLUE' Estimator here because it fails the 5th condition of Gauss-Markov that is homoskedastic assumption.

We use a generalized least squares method to approach this problem. We need to devise a method that accounts more importance to the residuals with low variability than the residuals with high variability.

One such method is called WLS - Weighted Least Squares. This is form of Generalized Least Squares.

```
# Define weights
weights <- 1 / residuals^2

# Fit weighted least squares model
model_wls <- lm(WAGE ~ MALE + educ1 + educ2 + educ3 + educ4 + EXPER, data = bwages, weights = we
ights)

summary(model_wls)
```

```
##
## Call:
## lm(formula = WAGE ~ MALE + educ1 + educ2 + educ3 + educ4 + EXPER,
##     data = bwages, weights = weights)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3179 -1.0023 -0.9939  0.9976  1.2170
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.3943319  0.0067498  1539.9   <2e-16 ***
## MALE         1.2959950  0.0084796   152.8   <2e-16 ***
## educ1       -7.8890349  0.0419919  -187.9   <2e-16 ***
## educ2       -6.0847736  0.0141202  -430.9   <2e-16 ***
## educ3       -4.2303335  0.0101160  -418.2   <2e-16 ***
## educ4       -2.4017911  0.0098674  -243.4   <2e-16 ***
## EXPER        0.1908636  0.0005901   323.4   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.001 on 1465 degrees of freedom
## Multiple R-squared:  0.9946, Adjusted R-squared:  0.9946
## F-statistic: 4.524e+04 on 6 and 1465 DF,  p-value: < 2.2e-16
```

The coefficients for both OLS and WLS are very similar, but there are some slight differences.

- The estimates for the intercept and the coefficient for MALE are almost identical.
- However, the coefficients for the education variables and EXPER differ slightly between the two models.
- This is because WLS puts more weight on observations with smaller residuals, which means that it is able to give more influence to the observations that are more informative and less influence to the observations that are outliers or have higher variance.

As a result, the standard errors for the coefficients in the WLS model are smaller, which means that the t-values are larger and the p-values are smaller. Overall, the WLS model has a better fit and is more reliable for inference than the OLS model.

**Question 18 - Compute White standard errors for (1) and compare them with the OLS standard errors found in question 2.**

OLS standard errors assume that errors are homoskedastic and uncorrelated across observations. On the other hand, White standard errors are robust to heteroscedasticity and serial correlation of errors, which are violations of the assumptions of OLS.

White standard errors take into account the estimated residual variances and covariances and adjust the standard errors accordingly. In other words, White standard errors provide more accurate and reliable inference when the OLS assumptions are violated

```r
# load required library
library(sandwich)

# fit OLS model
model_ols <- lm(WAGE ~ MALE + educ1 + educ2 + educ3 + educ4 + EXPER, data = bwages)

# compute White standard errors
vcov_white <- sandwich(model_ols, type = "HC0")

# compare standard errors
summary(model_ols)$coefficients[, "Std. Error"]
```

```
## (Intercept)        MALE       educ1       educ2       educ3       educ4
## 0.259946086 0.194696086 0.420968990 0.298744901 0.261643435 0.272184669
##       EXPER
## 0.009643716
```

```r
se_white <- sqrt(diag(vcov_white))
print(se_white)
```

```
## (Intercept)        MALE       educ1       educ2       educ3       educ4
##  0.28397537  0.18823308  0.40777013  0.35258093  0.32883652  0.32319594
##       EXPER
##  0.01106643
```

*Test to compare both standar errors*

*Hypothesis*

- H0 : There is no significance difference between both the standar errors.
- H1: There is a difference between the standard errors.

```r
coeftest(model_ols, vcov_white)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 10.401010   0.283975  36.6265 < 2.2e-16 ***
## MALE         1.300905   0.188233   6.9111 7.151e-12 ***
## educ1       -7.904926   0.407770 -19.3857 < 2.2e-16 ***
## educ2       -6.089686   0.352581 -17.2717 < 2.2e-16 ***
## educ3       -4.244575   0.328837 -12.9079 < 2.2e-16 ***
## educ4       -2.417478   0.323196  -7.4799 1.276e-13 ***
## EXPER        0.190744   0.011066  17.2363 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

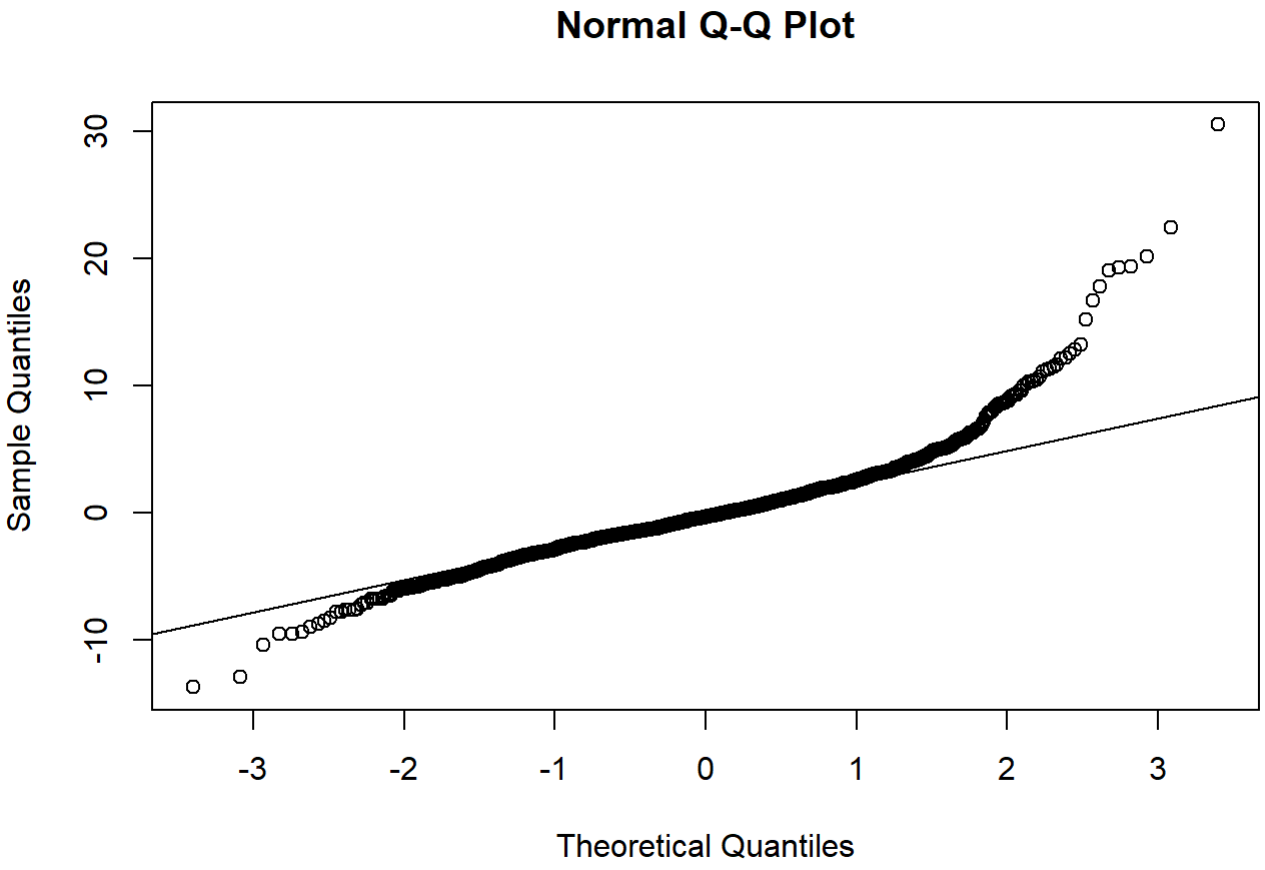
Since all the p-values are very small (less than 0.05), we can reject the null hypothesis that the corresponding coefficients are equal to zero, and conclude that they are statistically significant.

By comparing the OLS standard errors and the White standard errors, we can see that the White standard errors are larger than the OLS standard errors for all coefficients, indicating that the OLS standard errors underestimated the true standard errors. This suggests that the assumption of homoscedasticity (equal variance of errors) may not hold in the data, and the use of White standard errors, which account for heteroscedasticity, is more appropriate.

**Question 19 - Test if the distribution of ε in (1) is Gaussian**

```
# extract residuals
resid <- residuals(model_ols)

# create normal probability plot
qqnorm(resid)
qqline(resid)
```




```
shapiro.test(model_ols$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  model_ols$residuals
## W = 0.90163, p-value < 2.2e-16
```

The test resulted in a p-value < 2.2e-16, which is very small. This means that there is strong evidence to reject the null hypothesis that the residuals follow a normal distribution. Therefore, it can be concluded that the distribution of the errors is not Gaussian.

**Question 20 - If the previous question rejects, what are the consequences for the tests performed in this study?**

This can have several consequences for the tests performed in the study:

- The estimated coefficients of the model may be biased and inconsistent. This means that the parameter estimates may not reflect the true population values, and they may have large standard errors.
- The t-tests and p-values associated with the estimated coefficients may be unreliable. Since the t-statistics are based on the assumption of normally distributed errors, if this assumption is violated, the tests may produce incorrect results.
- The confidence intervals and prediction intervals may not be accurate. Since they are based on the assumption of normality, they may not capture the true uncertainty in the estimates.
- In such cases, it may be necessary to use alternative methods that do not rely on normality assumptions or to transform the data to make the error term more normally distributed. Alternatively, it may be necessary to collect more data or re-evaluate the model specification to better capture the underlying data-generating process.