Obaid Masih
Assignment #7
Heteroskedasticity

<u>Questions 8.4 parts a and c (p.322)</u>

a)

Residual plot of income shows heteroskedasticity.
Because as income grows then residual (error) grows exponentially. This shows that variance will be wider as income increases.

Residual plot of Age shows homoscedasticity. Because as Age grows the residual variance does not seem to change.

c)

(i) In all 3 regression models Income had a positive impact on the Miles travelled. In least squares regression increase of $1k is estimated to increase travelled miles by 14.2.
In all 3 regressions models AGE have a positive impact. In least squares regression model an increase of 1 year in AGE is estimated to increase travelled miles by 15.74Miles.
In all 3 regressions models KIDS have a negative impact. In least squares regression model an increase of 1 unit in KIDS is estimated to decrease travelled miles by 81.23 miles.

In all 3 regressions all 3 estimates are significant at 5% level

(ii) White test errors are not that far off from least square standard error. But they do give more precise t value & p values which in return give better confidence intervals and better hypothesis results.

(iii) In this case generalized least square estimates resulted in standard error of estimators to be smaller compared to ones obtained using White standard errors and least square estimates. But this is not evidence enough to suggest generalized least square estimates are better estimates. We do not know if our assumption about the form of the variance is correct. This means that standard error and estimates both could be wrong

<u>Question 8.12 (p.326)</u>

a)

WE expect heteroskedasticity because we know that as GDP per capita will increase then variance in Exp per capita will widen. As some countries like Austria or Denmark will have high GDP per capita but very low Exp per capita
but larger countries like US, CHINA, Russia will also have high GPD per capita but very large EE per capita as they are larger economies.

b)

Below is the equation estimated using least square
y = -.12457 + .0731732*x

where y is per capita spending on education and x is the GDP per capita

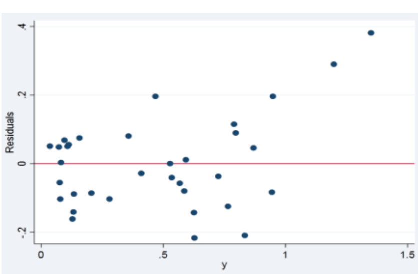
Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-0.124573	0.048523	-2.567	0.0151	*
xcol	0.073173	0.005179	14.128	2.65e-15	***

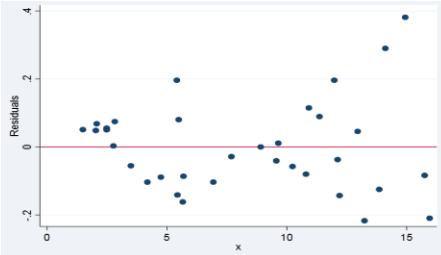
R studio Equation:

plot_ly(context,x= ~xcol, y= ~ycol)

Below is the plot of least square function and residual which shows that higher values of y or higher expenditure on education also have greater variance



Below is the plot of residuals with respect to x which suggests that residuals on the expenditure on education increase as the GDP of country increases which suggests there is heteroscedasticity



(c)
Test for heteroskedastic using White Test

Test was performed in car with library (lmtest) and function white.test()

HO : all alpha value are 0 meaning that model is Homoscedastic

H1: At least one value of alpha is not 0. And model has Heteroskedasticity

Chi square = 9.96 and P value = 0.0069

HO can be rejected and we accept that Model has Heteroskedasticity

(part d)

We use least square estimates to correct the standard error and below is the R code and output which calculates alternative standard errors using White's formulae. The least square estimate remains the same but standard error is corrected.

R studio White correction
coeftest(model, vcov. = vcovHC)

t test of coefficients:

```
             Estimate Std. Error t value  Pr(>|t|)
(Intercept) -0.1245728  0.0424400 -2.9353  0.006123 **
xcol         0.0731732  0.0066033 11.0813 1.735e-12 ***
```

tc <- 2.037

confidence interval without White correction
high = 0.073173 + (tc*0.005179)   # 0.083722
low =   0.073173 - (tc*0.005179)   # 0.062623

[0.062623, 0.0836]

 # confidence interval with heter correction
high = 0.073173 + (tc*0.0066033) # 0.0866
low =   0.073173 - (tc*0.0066033) # 0.0597

[0.0597,0.0866]

The confidence interval that ignores heteroskedastic errors is incorrect while the point estimate remains the same. Use of White test results in confidence interval that is wider than the one without white test.

(part e)

We assume the functional form of variance of error is $\operatorname{var}(e_i) = \sigma^2 x_i$

Since the form of variance of error is known, we apply the weighted least squares method with

Weight = $1/x_i$

Code for R is as follows:

```
w= 1/context$xcol

modelw = lm(ycol ~ xcol , data =context,weights = w)

summary(modelw)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.092921   0.028904  -3.215  0.00298 **
xcol         0.069321   0.004412  15.713  < 2e-16 ***
```

B2 95% interval is [0.06034,0.07841]

The error interval has reduced resulting in narrow confidence interval and reduced standard error.

********************************RSTUDIO CODE FOR above question/s ******************

```
###########################################
## QUESTION 8.12
rm(list=ls(all=TRUE))


library(multcomp)
library(data.table)
library(dplyr)
library(plotly)
library(lmtest)
library(sandwich)
library(car)

context = fread('pubexp.csv')

context = mutate(context, ycol = ee/p)
context = mutate(context, xcol = gdp/p)
## WE expect heteroskadasticity because we know that as Gdp per capita will increase then
## variance in Exp per capita will widen. As some countries like austria or denmark will have
## high gdp per capita but very low Exp per capita
## but larger countries like US, CHINA , rUSSIA will also have high GPD per capita but very
## large EE per capita

model = lm(ycol ~ xcol , data =context)
summary(model)

#Coefficients:
#  Estimate Std. Error t value Pr(>|t|)
#(Intercept) -0.124573   0.048523  -2.567   0.0151 *
#  xcol        0.073173   0.005179  14.128 2.65e-15 ***
#context =  mutate(context,residual= (context$ycol-predict(model))^2)

plot_ly(context,x= ~xcol, y= ~ycol)

# part b
## test for hetero with white test

ncvTest(model)
# Chisquare = 9.92055    Df = 1     p = 0.001634435
# p VALUE IS LOWER THAN 0.05 meaning that HO ( homoskaticity) is rejecetd
# This is not the White test.



#t test of coefficients:

#  Estimate Std. Error t value  Pr(>|t|)
```

```
#(Intercept) -0.1245728  0.0424400 -2.9353  0.006123 **
#  xcol        0.0731732  0.0066033 11.0813 1.735e-12 ***

## part c
coeftest(model, vcov. = vcovHC)
tc <- 2.037

# confidence interval without heter correction
high = 0.073173  + (tc*0.005179)   # 0.083722
low =   0.073173  - (tc*0.005179)   # 0.062623

# confidence interval without heter correction
high = 0.073173  + (tc*0.0066033)  # 0.0866
low =   0.073173  - (tc*0.0066033)  # 0.0597




## interval with correction is wider

# part d
###### Linear modelling with weighted
w= 1/context$xcol
modelw = lm(ycol ~ xcol , data =context,weights = w)
summary(modelw)


#  Estimate Std. Error t value Pr(>|t|)
#(Intercept) -0.092921   0.028904  -3.215  0.00298 **
 # xcol        0.069321   0.004412  15.713  < 2e-16 ***
```