**CODE :**

rm(list=ls())  
library(data.table)  
library(DBI)  
library(tidyverse)  
library(RSQLite)  
library(plm)  
library(gdata)

setwd("C:/Users/Vikram Arikath/Desktop/Applie Econ Projects")

mydata = read.csv("war\_countries.csv")

pdata <- mydata %>% pdata.frame(index=c('Country','Year'))  
head(pdata)

femodel1<- plm(Life.expectancy~War+Alcohol+Adult.Mortality+infant.deaths+percentage.expenditure+Hepatitis.B+Measles+BMI+under.five.deaths+Polio+Total.expenditure+Diphtheria+HIV.AIDS+GDP+Population+thinness..1.19.years+thinness.5.9.years+Income.composition.of.resources+Schooling, model="within", data=pdata)  
summary(femodel1)

femodel2 <- plm(Life.expectancy~War+Alcohol+Adult.Mortality+percentage.expenditure+Hepatitis.B+Measles+BMI+under.five.deaths+Polio+Total.expenditure+Diphtheria+HIV.AIDS+GDP+Population+thinness.5.9.years+Income.composition.of.resources+Schooling, model="within", data=pdata)  
summary(femodel2)

femodel3 <- plm(Life.expectancy~War+Alcohol+Adult.Mortality+Hepatitis.B+Measles+BMI+under.five.deaths+Polio+Total.expenditure+Diphtheria+HIV.AIDS+GDP+Population+thinness.5.9.years+Income.composition.of.resources+Schooling, model="within", data=pdata)  
summary(femodel3)

num\_columns <- mydata[, c(5:23)]  
res <- cor(num\_columns, use = "complete.obs")  
round(res, 2)

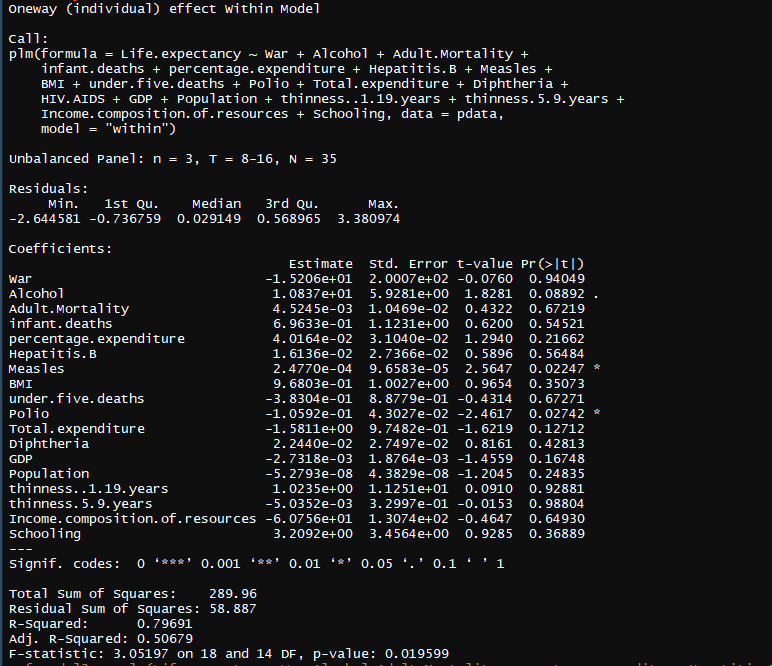
Checking for **multicollinearity** we see that :

The threshold is set to 0.85

Correlated variables:   
under.five.deaths & Infant.deaths - 0.999  
Total.expenditure & Infant.deaths - 0.895  
Total.expenditure & under.five.deaths - 0.899  
GDP & Percentage.expenditure - 0.929  
Thinness.5.9.years & thinness.1.19 years - 0.918

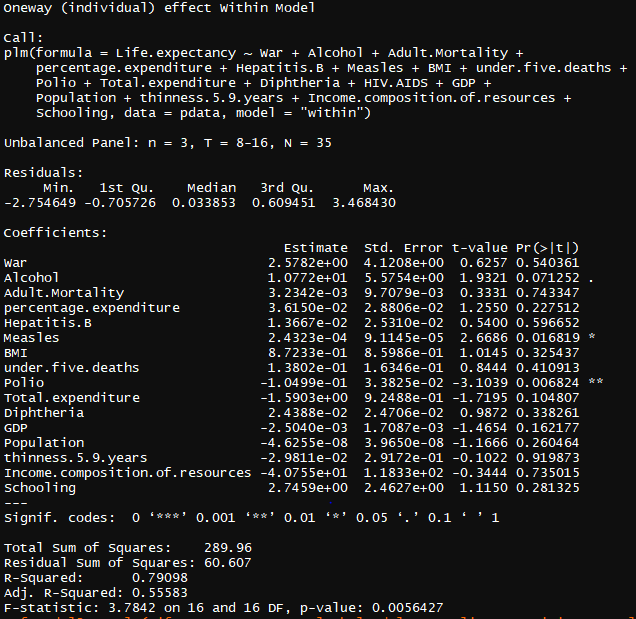
Under.five.deaths & infiant.deaths have high correlation since both have a similar menaning.

**RESULTS :**



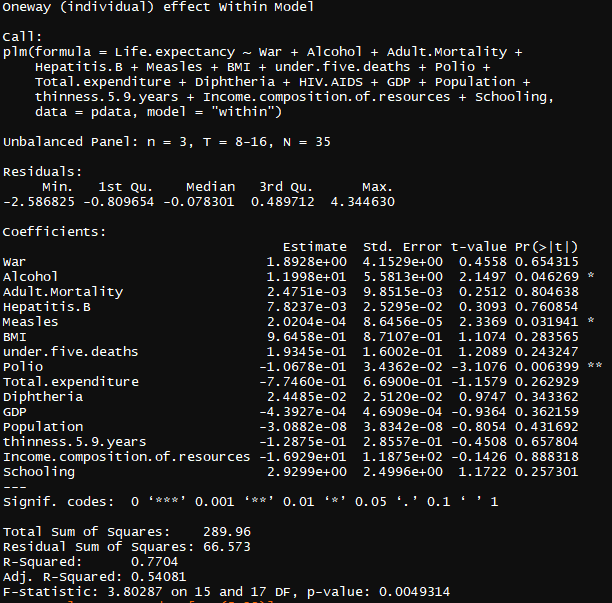
From the first model, we see that War does have a negative coefficient so the presence of it has an adverse effect on life expectancy, however from the p-value we see that it highly insignificant. Only significant variables are Measles, Polio and Alcohol. In this model we have not included year since it used as an index.

Checking for multicollinearity we drop infant.deaths and thinness 1.19 years.

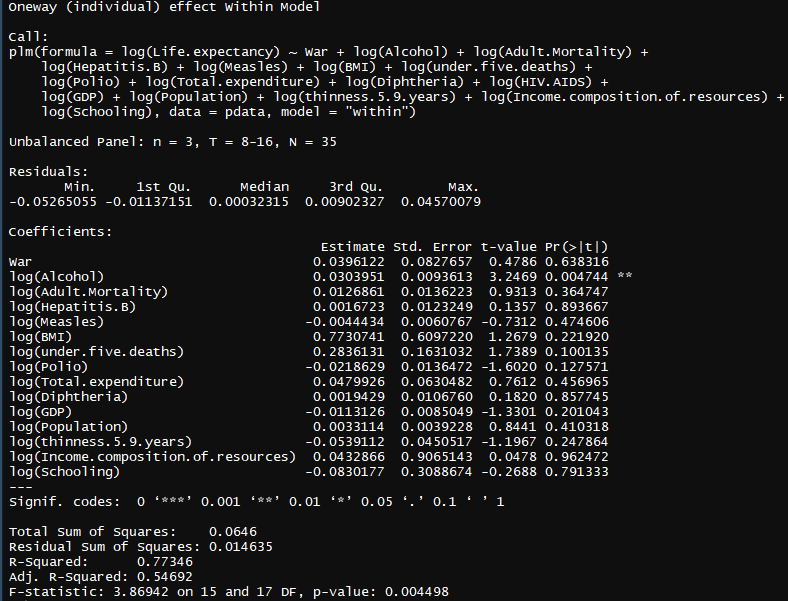


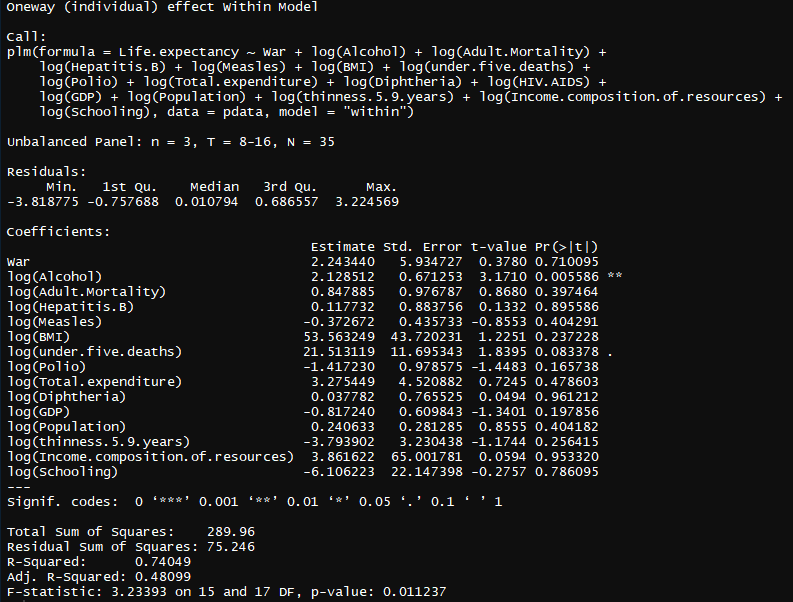
Second model we get the highest R squared value but here war has a positive coefficient with life expectancy, seems a little contradictory. Again Measles, Polio and Alcohol are significant variables here. Also what seems off here is that the total.expenditure also has a negative effect on life expectancy. Population and thinness 5-9 have a negative impact which makes sense.

Checking for multicollinearity we drop total.exepnditure and see the results.

**c**

Here, the model accuracy drops and similar trend from model 2 is seen. Polio, Measles and Alcohol still being the most significant variables for life expectancy. We see that significance of alcohol has increased.





**Analysis :**