# In-Class Lab 9

*ECON 4223 (Prof. Tyler Ransom, U of Oklahoma)*

*September 25, 2018*

The purpose of this in-class lab is to use R to practice testing for the presence of heteroskedasticity in regression models. We will do this using the Breusch-Pagan and White tests. We will also use the Lagrange Multiplier (LM) test. The lab should be completed in your group. To get credit, upload your .R script to the appropriate place on Canvas.

## For starters

First, install the `lmtest` package.

Open up a new R script (named `ICL9_XYZ.R`, where `XYZ` are your initials) and add the usual "preamble" to the top:

```r
# Add names of group members HERE
library(tidyverse)
library(broom)
library(wooldridge)
library(car)
library(lmtest)
library(magrittr)
```

## Load the data

We'll use a data set on college GPA, called `gpa3`. The data set contains a sample of 732 students.

```r
df <- as_tibble(gpa3)
```

Check out what's in the data by typing

```r
glimpse(df)
```

The main variables we're interested in are: SAT, high school percentile, credit hours, gender and race. We also only want to look at observations that are in the Spring semester.

### Restrict to observations in spring semester

Use a `filter()` statement to drop observations not in the Spring semester. (I won't show you the code; refer to a previous lab if you can't remember how to do it.)

### Get rid of variables you won't use

Use a `select()` statement to keep only the variables that will be used:

```r
df %<>% select(cumgpa,sat,hsperc,tothrs,female,black,white)
```

Look at the data to make sure the code worked as expected. You should now have 366 observations and 7 variables.

## Testing for Heteroskedasticity

Estimate the following regression model. (I won't show you the code; refer to a previous lab if you can't remember how to do it.)

$$cumpga = \beta_0 + \beta_1 sat + \beta_2 hsperc + \beta_3 tothrs + \beta_4 female + \beta_5 black + \beta_6 white + u$$

### Breusch-Pagan and White tests for Heteroskedasticity

To conduct the Breusch-Pagan test for heteroskedasticity, we use the `bptest()` function from the `lmtest` package:

```
bptest(est)
```

To do the White test, we simply modify the arguments in the `bptest()` function:

```
bptest(est, ~ fitted(est) + I(fitted(est)^2) )
```

1. Based on the results of each test, can you reject the null hypothesis of homoskedasticity?

### Inference with Heteroskedasticity-Robust Standard Errors

Now let's obtain standard errors from the above regression that are robust to heteroskedasticity. To do so, we use the `coeftest()` function from the `lmtest` package, and pass it the `hccm` option, which corresponds to the refined version of White's robust standard errors.

```
tidy(coeftest(est, vcov=hccm))
```

2. Compare your new estimates with the original ones. Are any of the default hypothesis test results overturned?

Now look at the robust version of the overall F-test. Is its conclusion changed?

```
glance(est)
linearHypothesis(est, c('sat','hsperc','tothrs','female','black','white'), vcov=hccm)
```

### The LM test

The LM test is an alternative to the overall F test that is reported in `glimpse(est)`. To perform the LM test, we need to do the following:

- Estimate the restricted model (in the current case, this is an intercept-only model) and then obtain the residuals from that.
- Regress the residuals from (a) on the regressors in the full model.
- the LM statistic is equal to $N * R^2$ from the regression in the second bullet.

3. Conduct an LM test following the steps above (also on p.159 of Wooldridge (2015))

```
# Restricted model
restr <- lm(cumgpa ~ 1, data=df)
LMreg <- lm(resid(restr) ~ sat + hsperc + tothrs + female + black + white, data=df)
LM    <- nobs(LMreg)*glance(LMreg)$r.squared
pval  <- 1-pchisq(LM,6)
```

4. Compare the p-value from the LM test with the p-value for the overall F test (with and without heteroskedasticity-robust correction).

# References

Wooldridge, Jeffrey M. 2015. *Introductory Econometrics: A Modern Approach.* 6th ed. Cengage Learning.