THE TASK

You are assigned a dataset from the Wooldridge package based on the last digit of your student ID number. Submitting a report for a wrong dataset results in the mark of zero for this assignment. This computer assignment corresponds to 10% of your final grade. The details about individual dataset and models are provided on the pages below.

Your answers have to be reported in the following table.

STUDENT ID NUMBER: (*insert here*)

Question 1	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\hat{\sigma}[x_1]$	$\hat{\sigma}[x_2]$	$\hat{\sigma}[x_3]$
Question 2	F	p-value	Test decision	LM	p-value	Test decision
Question 3	Size F	Size LM	Power F D1	Power LM D1	Power F D2	Power LM D2

You have to submit a single pdf file which starts with the above table. Fill in your answers in the table. For all the numbers report the first 4 digits after the decimal point: 3.567 instead of 3.5. You then have to include your R or MATLAB code. *Hint: in RStudio go to* `File - Knit Document` *and include the generated report. The code will compile if and only if it is written without mistakes. If your code does not compile you still have to include it (just copy and paste as a text after the table).*

The submission deadline for this assignment is **12.00hrs Greenwich Mean Time January 4, 2021**. The report has to be submitted via Turnitin on Blackboard. The assignment will be marked in accordance with the general *SoSS PG Marking Criteria* (available on Blackboard). Please make sure you are familiar with the University's rules and regulations regarding plagiarism.

In general your task is to produce an analysis of commonly used tests for heteroscedasticity.

1. Estimate a linear model of the form (individual for every dataset, details below)

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i \tag{1}$$

   Report the sample mean and sample standard deviation of each explanatory variable. Insert your answers in the table.

2. Conduct a test for heteroscedasticity (individual for every dataset, details below)

   - Compute the F-test version of the test together with the corresponding p-value. Do you reject the null hypothesis at the 5% level?
   - Compute the LM-test version of the test together with the corresponding p-value. Do you reject the null hypothesis at the 5% level?
   - Insert your answers in the table.

3. Examine the power and size properties of the given heteroscedasticity test in a simulation study with the fixed regressors design over $MC = 1000$ simulation draws. For every simulation iteration keep explanatory variables x_1, x_2, x_3 fixed. Use the sample size n from the assigned dataset. Set the coefficients for the data generating process to the OLS estimates from (1): $\delta_j = \hat{\beta}_j^{OLS}$.

   - In order to investigate the size of the test generate $u \sim \mathcal{N}(0,1)$ and $y_i = \delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2} + \delta_3 x_{i3} + u_i$ for every simulation iteration. Regressors x_{ij} remain fixed. Report the size of the F-version of the test and LM-version of the test separately.

- In order to investigate the power properties of the test install the `MASS` package. Use the function `mvrnorm` to simulate the heteroscedastic error term: `mvrnorm(mc, rep(0,n),Sigma)`.
- Power Design 1 (D1 in the table). Generate $u \sim \mathcal{N}(0, \Sigma)$, with $\Sigma_{ii} = x_{i1}$ and $\Sigma_{ij} = 0 \; \forall i \neq j$ and $y_i = \delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2} + \delta_3 x_{i3} + u_i$ for every simulation iteration. Report the power of the F-version of the test and LM-version of the test separately.
- Power Design 2 (D2 in the table). Generate $u \sim \mathcal{N}(0, \Sigma)$, with $\Sigma_{ii} = x_{i1} + x_{i2}$ and $\Sigma_{ij} = 0 \; \forall i \neq j$ and $y_i = \delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2} + \delta_3 x_{i3} + u_i$ for every simulation iteration. Report the power of the F-version of the test and LM-version of the test separately.
- Insert your answers in the table.

STUDENT ID NUMBERS ENDING WITH 0 OR 1

- ID Example: 123450 or 123451.
- Use the Wooldridge dataset `wage2` to analyse the **Breusch-Pagan** test for heteroscedasticity, i.e. use the misspecified auxiliary regression of the form

$$u^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \xi.$$

- Use `help(wage2)` to get the variable description.
- Follow the task step details from page 1.

- For the regression analysis use $y = wage$, $x_1 = educ$, $x_2 = exper$, $x_3 = IQ$

STUDENT ID NUMBERS ENDING WITH 2 OR 3

- ID Example: 123452 or 123453.
- Use the Wooldridge dataset `wage2` to analyse the **White** test for heteroscedasticity in its general form (not in terms of $\hat{y}$), i.e. use the misspecified auxiliary regression of the form

$$u^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_1 \cdot x_2 + \alpha_5 x_1 \cdot x_3 + \alpha_6 x_2 \cdot x_3 + \alpha_7 x_1^2 + \alpha_8 x_2^2 + \alpha_9 x_3^2 + \xi.$$

- Use `help(wage2)` to get the variable description.
- Follow the task step details from page 1.
- For the regression analysis use $y = wage$, $x_1 = educ$, $x_2 = exper$, $x_3 = IQ$

STUDENT ID NUMBERS ENDING WITH 4 OR 5

- ID Example: 123454 or 123455.
- Use the Wooldridge dataset `beauty` to analyse the **Breusch-Pagan** test for heteroscedasticity, i.e. use the misspecified auxiliary regression of the form

$$u^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \xi.$$

- Use `help(beauty)` to get the variable description.
- Follow the task step details from page 1.
- For the regression analysis use $y = lwage$, $x_1 = looks$, $x_2 = exper$, $x_3 = educ$

STUDENT ID NUMBERS ENDING WITH 6 OR 7

- ID Example: 123456 or 123457.
- Use the Wooldridge dataset `beauty` to analyse the **White** test for heteroscedasticity in its general form (not in terms of $\hat{y}$), i.e. use the misspecified auxiliary regression of the form

$$u^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_1 \cdot x_2 + \alpha_5 x_1 \cdot x_3 + \alpha_6 x_2 \cdot x_3 + \alpha_7 x_1^2 + \alpha_8 x_2^2 + \alpha_9 x_3^2 + \xi.$$

- Use `help(beauty)` to get the variable description.
- Follow the task step details from page 1.
- For the regression analysis use $y = lwage$, $x_1 = looks$, $x_2 = exper$, $x_3 = educ$

STUDENT ID NUMBERS ENDING WITH 8 OR 9

- ID Example: 123458 or 123459.
- Use the Wooldridge dataset `campus` to analyse the **White** test for heteroscedasticity in its general form (not in terms of $\hat{y}$), i.e. use the misspecified auxiliary regression of the form

$$u^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_1 \cdot x_2 + \alpha_5 x_1 \cdot x_3 + \alpha_6 x_2 \cdot x_3 + \alpha_7 x_1^2 + \alpha_8 x_2^2 + \alpha_9 x_3^2 + \xi.$$

- Use `help(campus)` to get the variable description.
- Follow the task step details from page 1.
- For the regression analysis use $y = crime$, $x_1 = enroll$, $x_2 = police$, $x_3 = priv$