# RStudio Lab Week 6 – Non-Parametrics and Simple Linear Regression

*Instructions*

1. Make sure that you have downloaded the dataset 'calls.xlsx', "Cars_data.xlsx", and "Goals_data.xlsx".
2. Copy the datasets into your student drive.
3. Make sure you have opened RStudio.
4. Set your working directory to your student drive by using Session – Set working directory – Choose directory and navigate to your student drive (or wherever you have saved the dataset for this quiz).
5. Insert a new R Script, using File – New File – R Script OR by clicking the white page with the green + button and selecting R Script.
6. Import the three datasets. You may need to use the function **read_excel()** from the **readxl** package.

*Non-parametrics*

This section deals with the 'Goals_data.xlsx' and 'Cars_data.xslx' datasets.

Consider the same Liverpool fan from last week. After discovering that he is not delusional about Darwin Nunez, he wants to perform another test on the players of his football club. He knows that Liverpool have 5 main attackers (as can be seen in the fictional data 'Goals_data.xlsx'). This time he wants to see if some attackers tend to score more than others? He gathered this data as well as the goals they scored in each of their kits (home, away, alternate) for the season.

7. Conduct the appropriate non-parametric test at the 1% significance level.

A car enthusiast, Zindile, often hears really loud cars in the early hours in the morning right outside her home. This has lead her to wondering if the louder a car is, the higher its maximum speed? In order to gather data she manages to stop many of the drivers in their loud cars and asks them to tell her the maximum speed their car has ever reached, they also rev their car for her and she judges how loud it is on a scale from 0 to 100. Her results are captured in the (fictional) Cars_data.xlsx file.

8. Conduct the appropriate non-parametrics test to determine whether Noise level and maximum speed are positively related. Test at the 5% significance level.

*Simple Linear Regression*

This section deals with the 'calls.xlsx' data.

A brokerage house would like to be able to predict the number of trade executions per day and has decided to use the number of incoming phone calls as an explanatory variable. Data were collected over a period of 35 days and saved in the dataset calls.xlsx. The task at hand is to perform a regression analysis on the data in order to develop a prediction equation for the number of trade executions as a function of the number of calls per day. (Note: All answers must be entered rounded to TWO decimal places).

9. Find the mean, variance and standard deviation of each of the variables in the dataset.

10. Apply the **cor()** function to the data where the independent variable *X* is Calls and the dependent variable *Y* is Executions to estimate the correlation coefficient.

11. Use the **plot()** function to create a scatterplot with Executions on the y axis and Calls on the x axis.

12. Perform a simple linear regression using the **lm()** function and save results as an object called fit. Hint: *formula = dependent variable ~ explanatory variable*

13. Use **summary(), anova()** and **confint()** on the fit object to generate the summary of regression results, the Anova table and 95% confidence intervals for the coefficients.