**LAB 1 - Simple Linear Regression**

In this lab, we will perform simple linear regression in Python. To do this, we will need the .csv provided in the LAB1 folder.

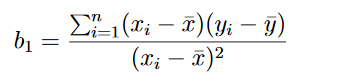
The .csv file contains information related to an imaginary soccer team (it is the same .csv file we examined in LAB0). We will investigate the “age” and “experience” columns and see how player age affects player experience. Simple linear regression is going to be the method of choice for this task. Please read the following instructions:

* (10 pts) Read the .csv file, extract the “age” and “experience” values into two different arrays. The age list is going to be our “x”, and experience list is going to be our “y”, in the upcoming instructions. When extracting the lists, don’t use standard lists. We will use numpy arrays since it is much easier to do element-wise operations with them, unlike standard lists.
* (10 pts) Implement two different methods. In the script (or main method), read the .csv file, extract the lists and call the corresponding methods in order. The script and the methods should be on the same file.
* (50 pts) The first method, named “simlin\_coef()”,computes and returns the regression coefficients. The eventual aim is to find two coefficients (b0, b1) such that:

ŷ = b1x + b0,

where the line “ŷ” refers to our resulting model, and x is our “age” data. This line is later going to be plotted.

The calculation of b0 and b1 consists of simple mathematical equations, where:





,

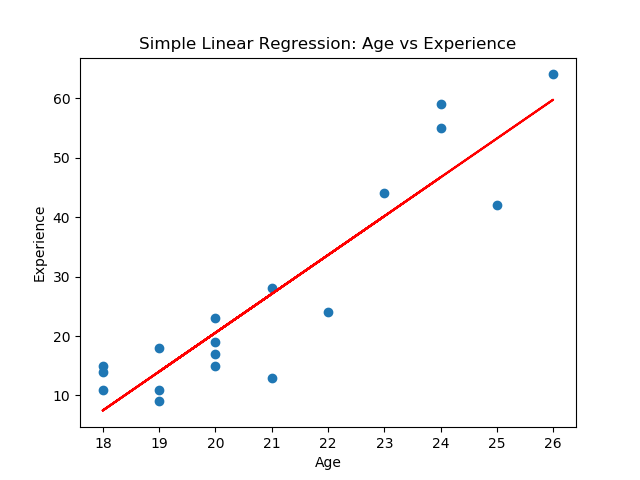
refers to the *average* values of the *x* and *y* datasets, respectively.



* (30 pts) The second method named “simlin\_plot()” will plot the results.
  + First, do a “scatter” plot using the actual data points: Use “age” as your x-axis and “experience” as your y-axis for the graph. Use the blue color for this plot.
  + Next, draw the regression line on the same window: Use “age” as your x-axis and the “ŷ” (the line you have already computed) as your y-axis. Use the red color for this line.
  + Last, call the show() function to display your window.

Any other packages besides “csv”, “matplotlib” and “numpy” are strictly forbidden for this lab. In the future, when the mathematical models get more complicated, we will use packages which make our lives easier.

The result of your code should look like this:



Now, some insight regarding this lab session:

In this lab, we took our data, and produced the best possible linear model for it. This alone is not very useful though! In standard machine learning tasks, we need some data to *predict*.

For example, imagine that there is a new batch of data, but that new batch has its “experience” column empty. Now, looking at the “age” column of this new batch, we can predict what the “experience” values would be. If our linear model is accurate enough, the “experience” values should be very close to that “line” we drew earlier.

In the next lab, we will see more about this distinction.