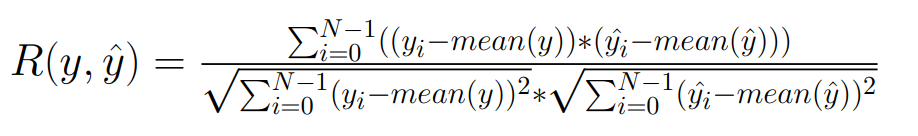
**R, R2, R2 (R2s), Adjusted R2**

Currently, there is a huge misunderstanding among frameworks around the world regarding the denoted forms of R, R2, R2s, and R2. Today, I will create this tutorial to clarify all of them.

# 1. R metric (Pearson’s Correlation Coefficient)



**R is Pearson’s Correlation Coefficient (shortname can be: R or PCC)**

Best possible score is 1.0, bigger value is better. Range = [-1, 1].

## 1.1 Microsoft (Excel software)

- Using syntax: “PEARSON(array1, array2)”

- There is no problem with this function in Excel

<https://support.microsoft.com/en-us/office/pearson-function-0c3e30fc-e5af-49c4-808a-3ef66e034c18>

## 1.2 Scikit-Learn (Python)

- Doesn’t support this metric

<https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics>

## 1.3 Matlab

- There is no version of this metric in Mathworks (They only have to function that return matrix)

<https://www.mathworks.com/help/matlab/ref/corrcoef.html>

## 1.4 Caret (R Programming Language)

- Doesn’t support this metric

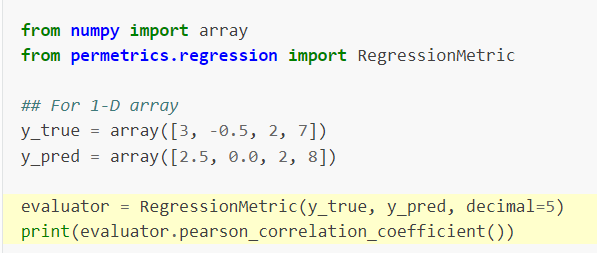
<https://rdrr.io/cran/caret/src/R/postResample.R>

## 1.5 Valmetrics (R Programming Language)

- The source code is right.

<https://rdrr.io/cran/valmetrics/src/R/r.R>

## 1.6 Permetrics (Python)

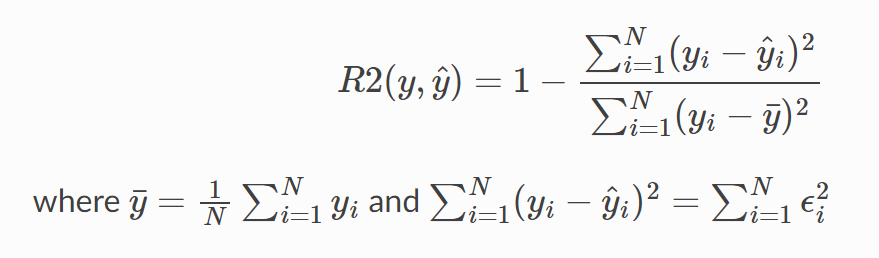


<https://permetrics.readthedocs.io/en/latest/pages/regression/R.html>

# 2. R2 Metric (Coefficient of Determination)

- Now, this metric is the real problem when considering the denotation. I don’t know who denoted the metric as R2 at first, and then all the documents on the internet started using the same notation.

- **My recommendation is that we denote this metric as R2 metric and not as R2 metric.** The problem arises because some people think this is (Pearson’s Correlation Coefficient)2 , and there are many documents on the internet that show this. This is confusing for newbies and those who have no knowledge of mathematics. I will show you why.



**-** R2 is Coefficient of Determination, we can denoted it as COD or R2

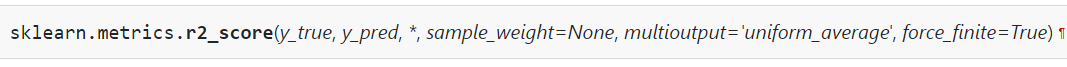
- Best possible score is 1.0, bigger value is better. Range = (-inf, 1]

## 2.1 Microsoft (Excel software)

- There is no version of this function in Microsoft Excel. However, people usually mistake it for the RSQ function, which is the square of the Pearson product moment correlation coefficient, or (Pearson’s Correlation Coefficient)2

<https://support.microsoft.com/en-us/office/rsq-function-d7161715-250d-4a01-b80d-a8364f2be08f>

## 2.2 Scikit-Learn



- The function implemented in Scikit-Learn is correct. However, they denote it as R2, which is causing confusion among people.

<https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html>

## 2.3 Matlab

- The official Matlab gives the correct equation. However, since we cannot see the source code, we do not know if their implemented version is correct or not.

<https://www.mathworks.com/help/stats/coefficient-of-determination-r-squared.html>

- However, there is a report that discusses the difference between R squared and the 'Squared of Correlation Coefficient'. This report shows that they are not the same, so I think the implemented version in Matlab is correct.

## 2.4 Caret (R Programming Language)

- Now, this package is widely used by tons of people, and the equation they show appears to be correct. <https://rdrr.io/cran/caret/man/postResample.html>



- However, they implemented with the wrong code. <https://rdrr.io/cran/caret/src/R/postResample.R>

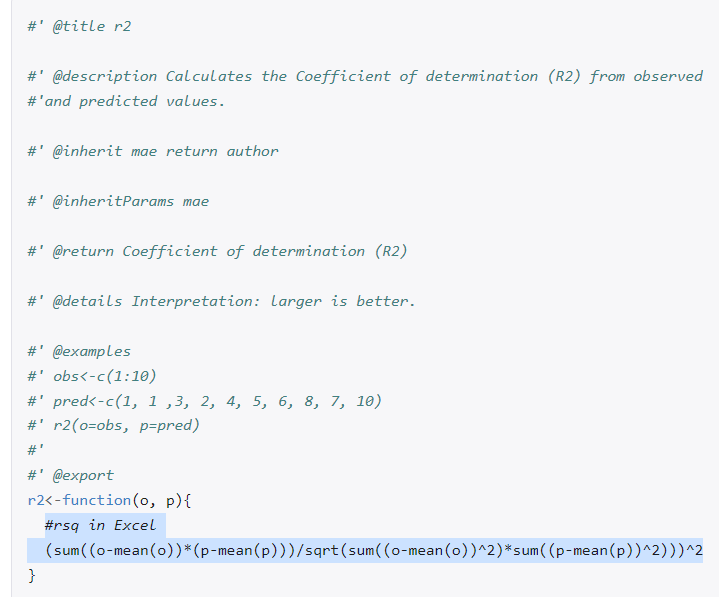


However, what they return is (Pearson's Correlation Coefficient)2, which is not the same as their equation. It is important to note that many researchers use this package, yet no one has discovered this problem.

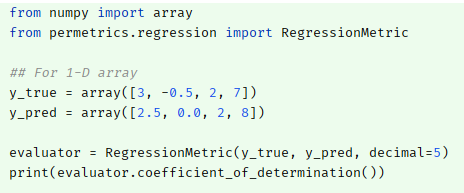
## 2.5 Valmetrics (R Programming Language)

- The same problem with this library in R language. <https://rdrr.io/cran/valmetrics/src/R/r2.R>

- They do not provide the equation, but they state that this is the RSQ function in Excel. So you can see that they are not technically wrong. However, it can be misleading for users because when they use this library, they may assume that it is the Coefficient of Determination, which it is not. Instead, it is RSQ or (Pearson's Correlation Coefficient)2.



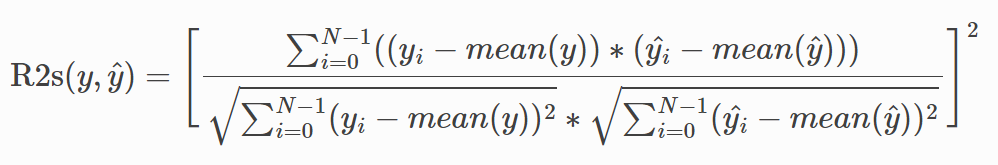
## 2.6 Permetrics (Python)



<https://permetrics.readthedocs.io/en/latest/pages/regression/R2.html>

# 3. R2 Metric (R-squared or (Pearson’s Correlation Coefficient)2 )

Basically, the Scikit-Learn library needs to change the notation of this metric so that users do not get confused when considering this metric and the Coefficient of Determination.



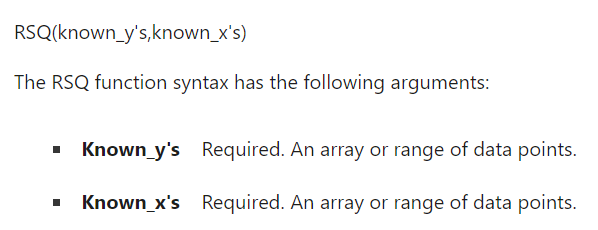
- It is (Pearson’s Correlation Index)2 = R2 = R2s (R squared)

- Best possible score is 1.0, bigger value is better. Range = [0, 1]

## 3.1 Microsoft (Excel software)

- The function is RSQ.

<https://support.microsoft.com/en-us/office/rsq-function-d7161715-250d-4a01-b80d-a8364f2be08f>



## 3.2 Scikit-Learn

- Doesn’t support this metric

<https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics>

## 3.3 Matlab

- Mathworks doesn’t support this metric. There is an R-squared metric, but it is the Coefficient of Determination, and its equation is correct. The only issue is that they give it the same name as Scikit-Learn, which can be confusing for users.

<https://www.mathworks.com/help/stats/coefficient-of-determination-r-squared.html>

## 3.4 Caret (R Programming Language)

While their equation is incorrect for this metric, their code implementation is correct. They have implemented it as the RSQ function in Excel.



## 3.5 Valmetrics (R Programming Language)

This library doesn’t support this metric

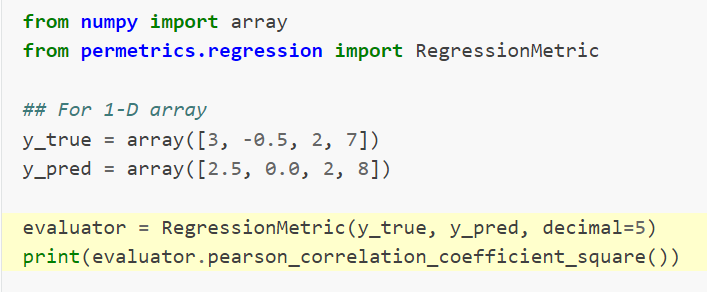
<https://rdrr.io/cran/valmetrics/man/>

## 3.6 Permetrics (Python)

You can calculate this metric by difference names such as :

R2s = r2s = pearson\_correlation\_coefficient\_square

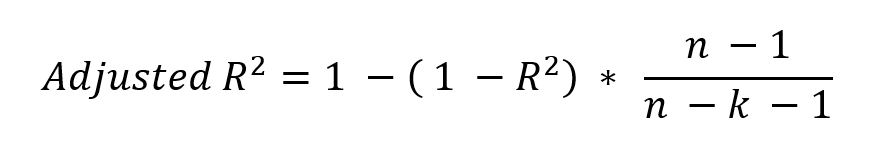
Please read the document from here: <https://permetrics.readthedocs.io/en/latest/pages/regression/R2s.html>



# 4. Adjusted R2 (AR2)

- Adjusted Coefficient of Determination (ACOD or AR2)

- Best possible score is 1.0, bigger value is better. Range = (-inf, 1]



In which: n is the total number of data points, k is the number of input features of the model. For example, in the equation (model) y = ax1 + bx2 + c, k = 2 (x1 and x2)."

## 4.1 Microsoft (Excel software)

- Excel doesn’t support this function, you need to calculate by hand.

<https://www.statology.org/adjusted-r-squared-excel/>

## 4.2 Scikit-Learn

- Doesn’t support this metric

## 4.3 Matlab

- Their equation is right. You can call the function.

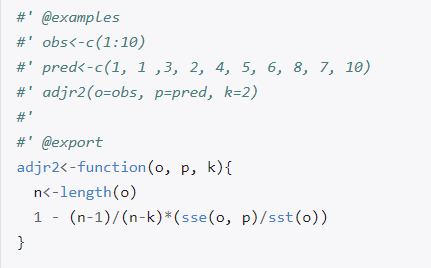
<https://www.mathworks.com/help/stats/coefficient-of-determination-r-squared.html>

## 4.4 Caret (R Programming Language)

- This library doesn’t support this function, you need to calculate by hand.

## 4.5 Valmetrics (R Programming Language)

- This library doesn’t provide the equation. However, based on their source code. It is correted implementation. <https://rdrr.io/cran/valmetrics/man/adjr2.html>



## 4.6 Permetrics (Python)

Check the document: <https://permetrics.readthedocs.io/en/latest/pages/regression/AR2.html>

