## **Correlation between Maintainability index and Post-Release Defect Density**

## Correlation was started with a hypothesis that with high maintainability index we will have low post release defect density value.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Projects** | **Versions** | **Maintainability Index** | **Defect density** | **Spearman coefficient** |
| **Apache Commons DbUtils** | 1.1 1.2 1.4 1.6  1.7 | 72.63  70.57  78.43  78.79  79.03 | 0.00173 0.00134  0.000514  0.00074  0.000874 | *rs* = -0.5 |
| **Apache Commons Configuration** | 1.8  2.0  2.2  2.4  2.6 | 70.41  68.92  72.6  73.92  75.16 | 0.000379  0.000169  0.000105  0.0000297  0.0000592 | *rs* = -0.8 |
| **Apache Commons Collections** | 2.0 3.0 3.2 4.1 4.4 | 79.62  78.96  75.24  60.56  61.06 | 0.000293  0.000326  0.000992 0.000263  0.0000789 | *rs* = 0.5 |

By analysis of the correlation, we have observed that each project has correlation between weak to medium. So that, we can say that having a better/higher maintainability index ensures less software maintenance costs. However, it might not always ensure a bug free system. In addition, there are might be other factors that affects the value of the post release defect density like the experience of the developers and testers, the type of the defects considered and the time required for the calculation of the post release defect density calculation.

***Calculating Spearman Correlation***

Spearman’s Rank correlation coefficient is one of the most-prominent technique which can be used to find out the strength and correlation between two variables. We have calculated the spearman correlation coefficient using the following method for one project and for the rest of the projects we have used the online tool combining all the versions of a project

***Method used to calculate the Spearman correlation***

* Create a table from your data and get the ordered pairs of two variables.
* Rank the two data sets. Ranking is achieved by giving the ranking '1' to the biggest number in a column, '2' to the second biggest value and so on. The smallest value in the column will get the lowest ranking. This should be done for both sets of measurements or the variables used to find the correlation for.
* Tied scores are given the mean (average) rank.
* Find the difference in the ranks (d).
* Square the differences (d²) To remove negative values and then sum them
* Calculate the coefficient (***Rs***) using the formula mentioned below.

When written in mathematical notation the Spearman Rank formula looks like this:



Here,

ρ= Spearman rank correlation

di= the difference between the ranks of corresponding variables

n= number of observations

We have used the formula of spearman correlation coefficient by considering the Post-Release Defect Density values as X- values and the Maintainability index values as Y- values.