## **Correlation between Code Coverage and Post-Release Defect Density**

***We have started our analysis with a hypothesis that, with higher maintainability index we achieve less number of post-release defects.*** Maintainability index reflects the ease of maintaining a project. That means with higher maintainability index, the cost and effort needed to make changes or fix bugs is reduced. Having said that, having a better/higher maintainability index ensures the less number of bugs in the system. For computing the Correlation between Maintainability Index and post release defect density, We have calculated the spearman correlation coefficient for that as Shown in the Table-1. Here we have considered the 6 versions of each projects for calculation of the Maintainability Index and Post- Release Defect Density for better analysis and Statistics.

***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.

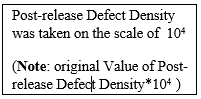
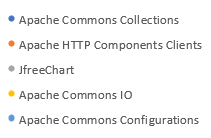
Table-1 gives the values of the No of defects found , SLOC , Post-Release Defect Density, Maintainability Index , as well as Spearman Correlation Coefficient for the particular version for each project.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Project Name** | **Versions** | **No of Post Release Defects** | **SLOC** | **Post-Release Defect Density** | **Maintainability Index** | **Spearman Correlation Coefficient** |
| **Apache**  **Commons Collections** | 1 | 1 | 27083 | 0.000036923 | 63.82 | 0.143 |
| 2 | 4 | 13563 | 0.00029492 | 59.40 |
| 3 | 15 | 45670 | 0.000328443 | 69.03 |
| 4 | 36 | 51608 | 0.000697566 | 66.52 |
| 4.1 | 13 | 60717 | 0.000214108 | 65.01 |
| 4.2 | 3 | 62645 | 0.00004789 | 76.30 |
| **Apache HTTP Components Clients** | 4.4.1 | 1 | 74293 | 1.34602E-05 | 71.42 | -0.46 |
| 4.5.2 | 2 | 75021 | 2.66592E-05 | 68.91 |
| 4.5.3 | 2 | 75427 | 2.65157E-05 | 69.12 |
| 4.5.5 | 1 | 76095 | 1.31415E-05 | 69.27 |
| 4.5.6 | 1 | 76130 | 1.31354E-05 | 69.26 |
| 4.5.7 | 2 | 76128 | 2.62715E-05 | 79.10 |
| **JfreeChart** | 1.0.12 | 9 | 144406 | 6.23243E-05 | 52.29 | -0.49 |
| 1.0.13 | 7 | 145897 | 4.79791E-05 | 55.23 |
| 1.0.14 | 23 | 145030 | 0.000158588 | 56.86 |
| 1.0.15 | 13 | 144690 | 8.98473E-05 | 72.24 |
| 1.0.16 | 1 | 144192 | 6.9352E-06 | 75.52 |
| 1.0.17 | 3 | 144340 | 2.07843E-05 | 81.29 |
| **Apache Commons IO** | 2 | 8 | 21434 | 0.000373239 | 67.61 | 0.26 |
| 2.1 | 12 | 22129 | 0.000542275 | 67.07 |
| 2.2 | 6 | 23789 | 0.000252217 | 66.28 |
| 2.3 | 10 | 23983 | 0.000416962 | 66.50 |
| 2.4 | 64 | 24384 | 0.002624672 | 66.30 |
| 2.5 | 24 | 29152 | 0.000823271 | 74.94 |
| **Apache Commons Configurations** | 1.1 | 39 | 15181 | 0.002569001 | 54.97 | -0.60 |
| 1.2 | 12 | 18489 | 0.000649035 | 53.34 |
| 1.3 | 19 | 27819 | 0.000682986 | 56.04 |
| 2.1 | 12 | 65558 | 0.000183044 | 68.35 |
| 2.2 | 8 | 66697 | 0.000119945 | 65.09 |
| 2.3 | 6 | 67096 | 0.00008942 | 64.15 |

Table-1: Showing Post-Release Defect Density, Maintainability Index with Spearman Correlation Coefficient for each Project.

By analysis of the result as seen in Table-1 , we have studied 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 also some 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 taken into account and the time required for the calculation of the post release defect density calculation.

**Scatterplots for Graphically representing the correlation**

For scatterplots, we have used Excel. The given Scatterplots depicts the correlation between the two metrics where we have found 2 outliers. This might be because the maintainability index depends on other factors like Halstead volume, SLOC and cyclomatic complexity which changes with operators and operands. This increase does not mean there should be increase in cyclomatic or the other way around.