**Correlation between Statement and Branch coverage with Cyclomatic complexity**

Correlation was started with a hypothesis that the project with higher complexity will likely have less code coverage.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Project Name** | **M1- Statement Coverage** | **M2-Branch**  **Coverage** | **M4-Average Cyclomatic Complexity** | **Spearman Correlation (M1 & M4)** | **Spearman Correlation (M2 & M4)** |
| Apache Commons Collections | 51% | 82% | 12.3 | 0.42417 | -0.16242 |
| Apache Commons Configuration | 89% | 83% | 14.7 | 0.04871 | -0.23174 |
| Apache Commons DbUtils | 64% | 77% | 7.1 | -0.66689 | -0.5 |

Table-1: Showing Statement Coverage, Branch Coverage, and Cyclomatic complexity.

Negative correlation between branch coverage and cyclomatic complexity is observed.

Correlation between statement coverage and cyclomatic complexity is not strong. Positive values for Collections and Configuration are observed while DbUtils has negative correlation.

Correlation between Statement and Branch Coverage with McCabe Complexity is observed to be not strongly related. We found that there might be other factors affecting the complexity.

***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.

***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 spearmen correlation coefficient by considering the statement coverage and branch coverage values as X- values and theCyclomatic Complexity as Y- values.