## **Correlation between Code Coverage and Test Suite Effectiveness**

***We have started our analysis with a hypothesis that, Code coverage is strongly correlated to Mutation score and increases linearly with the increase in Mutation score.*** While choosing projects, we have taken into consideration the number of test suites implemented. More no. of test suites implies better analysis of effectiveness. Table-1 shows that out of five projects, Apache Commons Collection and Apache Commons Math have the highest value of Spearman correlation coefficient between Code Coverage and Mutation Score. We have shown *Spearman Correlation Coefficients* in Table-1, however it is not calculated based on code coverage% and mutation coverage% as seen in Table-1. To calculate *Spearman Correlation Coefficient*, we have taken data range having code coverage and mutation scores for all test suites in all projects. However, it is remaining unclear that the effectiveness is affected due to test suite size or coverage of the test suite. We found that as Code Coverage increases, Mutation Score also increases in most of the test suites of projects. But in Apache Commons Collections, Mutation Score is not strongly increase with the coverage. This shows that sometimes Code coverage is moderately correlated to Mutation score.

|  |  |  |  |
| --- | --- | --- | --- |
| Project | Code Coverage | Mutation Score | Spearman Correlation Coefficient |
| Apache Commons Collections | 86% | 42% | 0.962 |
| Apache Commons Math | 92% | 79% | 0.487 |
| JFreeChart | 54% | 33% | -0.104 |
| Apache Commons Lang | 95% | 86% | -0.220 |
| Apache Commons DbUtils | 64% | 51% | -0.315 |

Table-1: Showing sorted Code coverage, Mutation score, and Spearman Correlation Coefficient.

***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