4. Steps for analyzing the data

It is necessary to do correlation analysis after collecting the data for these 6 metrics. The correlation analysis was adopted by the Pearson correlation coefficient and Spearman correlation coefficient. (If there is no introduction in the ‘Introduction’ part, the formulas and references for both should be included)

The Steps of data analysis are as follows:

(1) Determining which two metrics are used for correlation comparative analysis, and determining which level of data they are ( such as, package level, class level ). Extracting the metric data of specific project from the collected data.

(2) Importing the collected metric data into ‘Matlab’ for both Pearson correlation coefficient analysis and Spearman correlation coefficient analysis. Collecting the correlation coefficient and generating the distribution map of the data points.

(3) Comparing the results of the specific metric correlation coefficients of the five projects and drawing the most general conclusion.

5.Results analysis

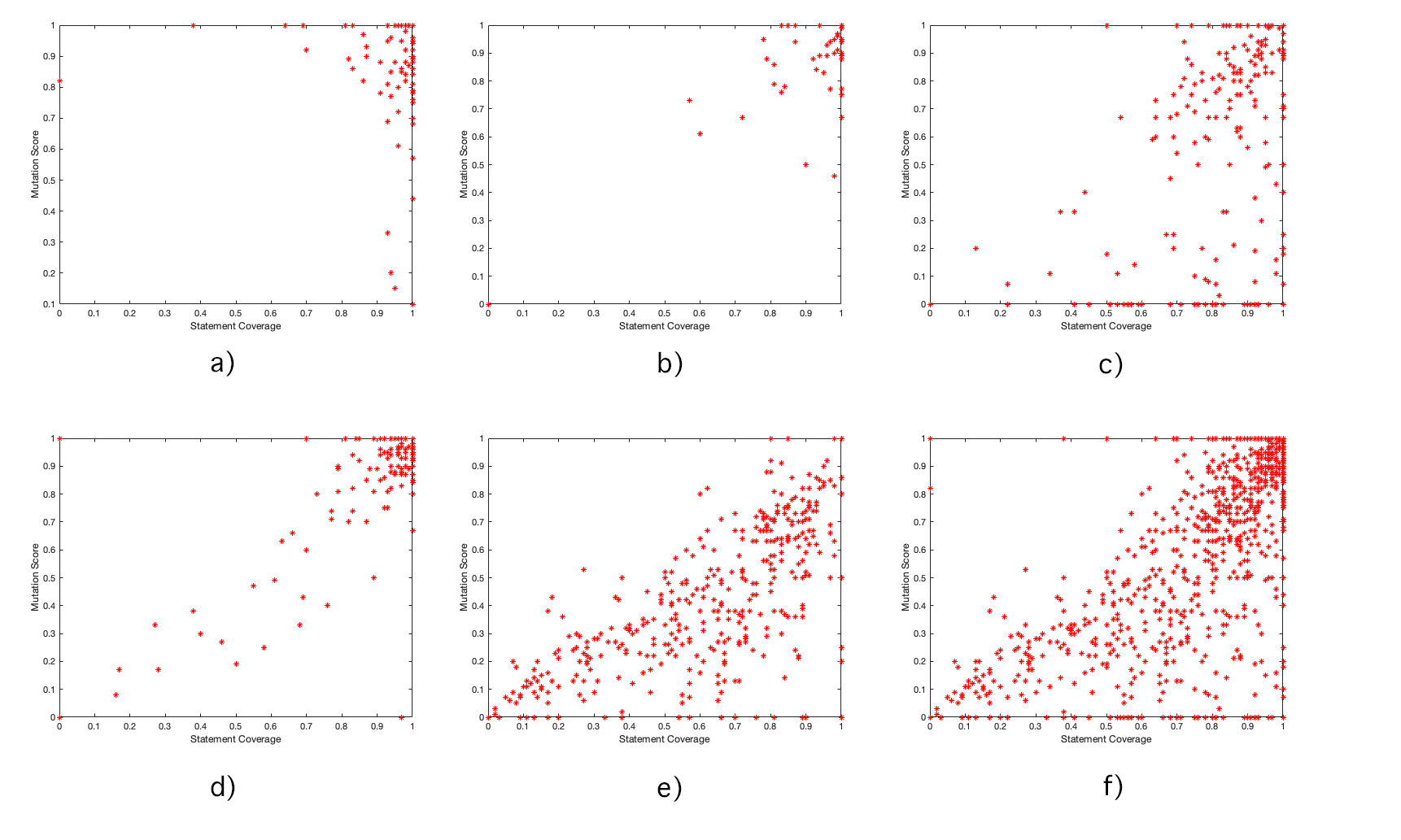
5.1 Correlation between Metric 1 & 2 & 3.

|  |  |  |
| --- | --- | --- |
| Project | Sets of data (Class level) | R (Pearson) of metric 1&3 |
| Total 5 project | 1063 | 0.7476 |
| Apache commons Lang | 89 | -0.0564 |
| Apache commons codec | 52 | 0.8027 |
| Apache commons collections | 264 | 0.4510 |
| Apache commons configuration | 177 | 0.8266 |
| JFreeChart | 481 | 0.7996 |
| Apache commons Lang | 11 sets Package-Level data | 0.3152 |

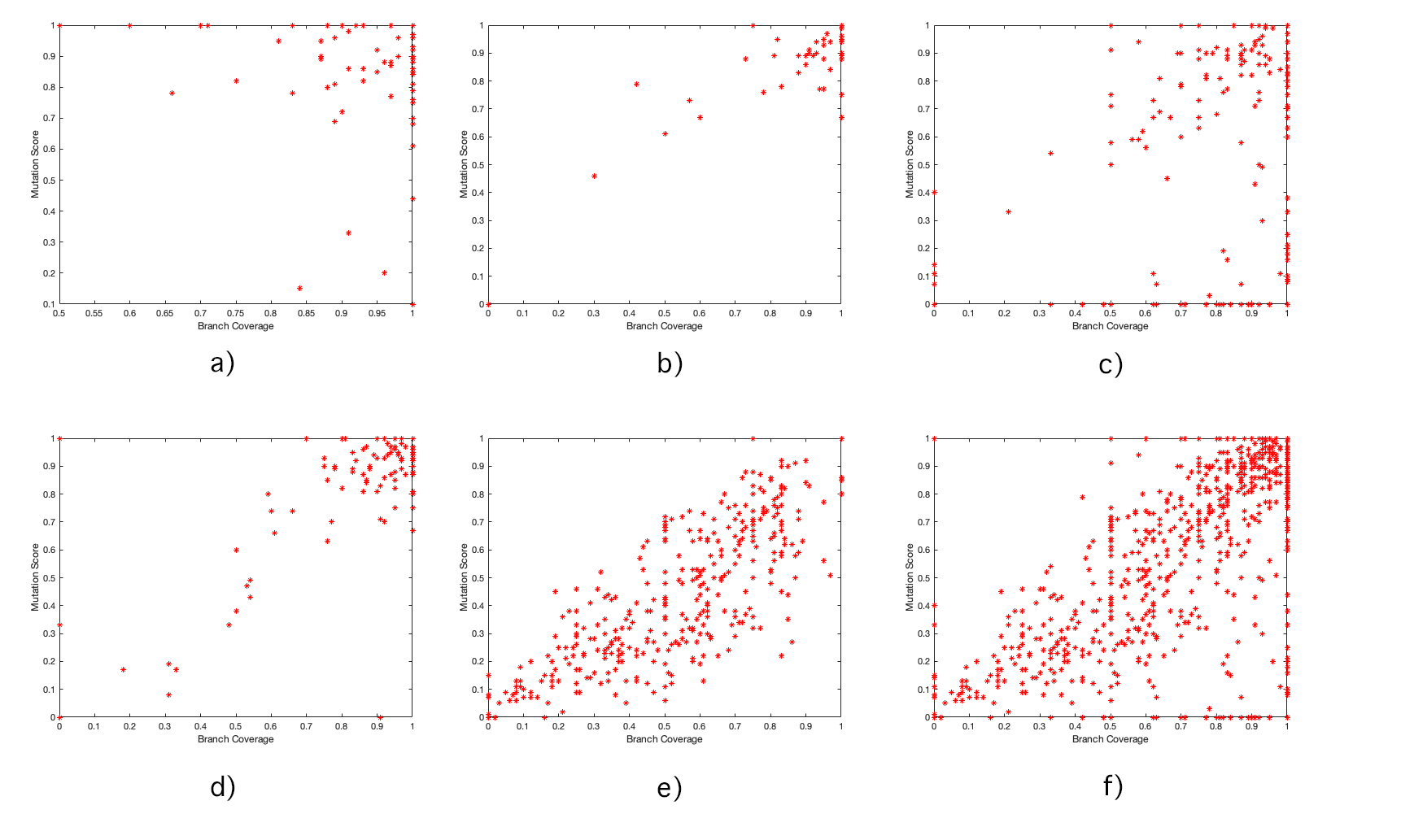
Table1. The Pearson correlation coefficient between Metric 1 & 3

|  |  |  |
| --- | --- | --- |
| Project | Sets of data(Class level) | R (Pearson) of metric 2&3 |
| Total 5 project | 899 | 0.7707 |
| Apache commons Lang | 75 | -0.0847 |
| Apache commons codec | 47 | 0.8674 |
| Apache commons collections | 206 | 0.3714 |
| Apache commons configuration | 143 | 0.753 |
| JFreeChart | 428 | 0.7996 |
| Apache commons Lang | 11 sets Package-Level data | 0.8627 |

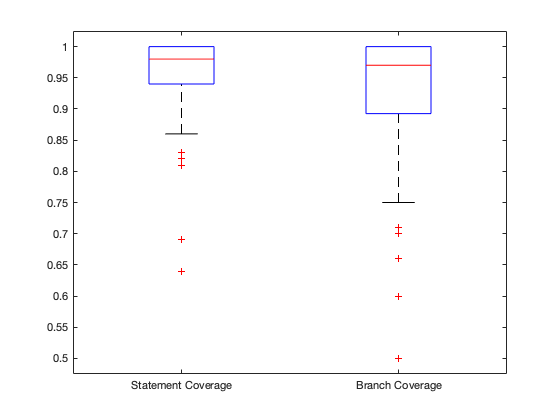
Table2. The Pearson correlation coefficient between Metric 2 & 3



**Figure1** Class-Level data distribution diagram of Class-Level between Metric 1 & 3 a) Apache commons Lang b) Apache commons codec c) Apache commons collections d) Apache commons configuration e) JFreeChart f）Total five project class level data



**Figure2** Class-Level data distribution diagram of Class-Level between Metric 2 & 3 a) Apache commons Lang b) Apache commons codec c) Apache commons collections d) Apache commons configuration e) JFreeChart f）Total five project class level data



**Figure3** Apache commons Lang metric 1 & 2 boxplot