

SECTION 4.2

DATA MINING

SAMPLING

STRUCTURAL METRICS
MINING

- Select project sample
- Retrieve source code for each project revision and conduct static code analysis to mine structural metrics

SECTION 4.3

SAMPLE ANALYSIS

TEAM SIZE

STRUCTURAL METRICS

- Analyse sample to ensure it is representative of forge
- Analyse team size trends
- Analyse structural metric distributions and correlations

SECTION 4.4

UNIVARIATE TEAM
SIZE ANALYSIS

TEAM SIZE

STRUCTURAL METRICS

- Bucket project-level metrics by committer count
- Carry out bucket comparison analysis (Mann-Whitney tests)
- Analyse relationship between the team size and metric values

SECTION 4.5

MULTIVARIATE TEAM
SIZE ANALYSIS

TEAM SIZE

CONFOUNDING FACTORS

STRUCTURAL METRICS

- Analyse code volume and revision counts as confounding factors
- Bucket class-level metrics by team size and revision count to mitigate impact of functional complexity
- Carry out bucket comparison analysis

SECTION 4.6

PROJECT-LEVEL
ANALYSIS

PROJECT SELECTION

PROJECT QUALITATIVE AND
QUANTITATIVE ANALYSIS

MODEL VALIDATION

- Data set dimensionality reduction and visualisation to support project selection
- Quantitatively compare metrics. Qualitatively compare committer activity and source code
- Perform linear regression on individual projects