

DELHI TECHNOLOGICAL UNIVERSITY

Empirical Software Engineering Experiment list

1. Consider any empirical study of your choice (Experiments, Survey Research, Systematic Review, Postmortem analysis and case study). Identify the following components for a empirical study.
 - a. Identify parametric and nonparametric tests
 - b. Identify Independent, dependent and confounding variables
 - c. Is it Within-company and cross-company analysis?
 - d. What type of dataset is used? Proprietary and open source software

2. Defect detection activities like reviews and testing help in identifying the defects in the artifacts (deliverables). These defects must be classified into various buckets before carrying out the root cause analysis. Following are some the defect categories: Logical, User interface, Maintainability, and Standards.
In the context of the above defect categories, classify the following statements under the defect categories.
 - a. Divide by Zero Error is not guarded
 - b. Usage of 3.14 in the statement $\text{Circle_Area} = 3.14 * \text{Radius} * \text{Radius};$
 - c. 3500 lines of code in a single function
 - d. A pointer is declared but not initialized. It is used in the program for storing a value.
 - e. A program designed to handle 1000 simultaneous users, crashed when 1001 the user logged in.
 - f. A “while” loop never exits
 - g. User interface displays “MALFUNCTION 54” when something goes wrong in the back-end
 - h. No documentation (comments) for the source code
 - i. Hungarian Notation not followed while coding, even though the coding guidelines mandate to use Hungarian Notation
 - j. Pressing of “Tab” key moves the cursor in different fields of a web form randomly.

3. Consider any prediction model of your choice.
 - a. Analze the dataset that is given as a input to the prediction model
 - b. Find out the quartiles for the used dataset
 - c. Analyze the performance of a model using various performance metrics.

4. Consider defect dataset and perform following feature reduction techniques using Weka tool. Validate the dataset using 10-cross validation.
 - a. Correlation based feature evaluation
 - b. Relief Attribute feature evaluation
 - c. Information gain feature evaluation
 - d. Principle Component

5. Online loan system has two modules for the two basic services, namely Car loan service and House loan service. The two modules have been named as Car_Loan_Module and House_Loan_Module. Car_Loan_Module has 2000 lines of uncommented source code. House_Loan_Module has 3000 lines of uncommented source code. Car_Loan_Module was completely implemented by Mike. House_Loan_Module was completely implemented by John. Mike took 100 person hours to implement Car_Loan_Module. John took 200 person hours to implement House_Loan_Module. Mike's module had 5 defects. John's module had 6 defects. With respect to the context given, which among the following is an INCORRECT statement? Identify the null and alternate hypotheses for the followings options.

Justify and Choose one:

- a. John's Quality is better than Mike's Quality
 - b. John's Productivity is more than Mike's Productivity
 - c. John introduced more defects than Mike
 - d. John's Effort is more than Mike's Effort.
6. Statistical Hypothesis Testing in R- Statisticians use hypothesis testing to formally check whether the hypothesis is accepted or rejected. Consider an example or data of your choice and identify the following:
- a. State the Hypotheses
 - b. Formulate an Analysis Plan
 - c. Analyze Sample Data
 - d. Interpret Results
 - e. Estimate type-I and type-II error
7. Consider defect dataset and implement following statistical test using SPSS tool.
- a. t-test
 - b. Chi-Square Test
 - c. Wilcoxon Signed Test
 - d. Friedman Test
 - e. Kruskal Wallis Test
8. Why is version control important? How many types of version control systems are there? Demonstrate how version control is used in a proper sequence (stepwise).
9. Demonstrate how Git can be used to perform version control?
10. Perform comparison of following data analysis tools.
- a. WEKA
 - b. KEEL
 - c. SPSS
 - d. MATLAB
 - e. R
11. Validate the results obtained in experiment 3 using 10-cross validation, hold out validation or leave one out cross-validation.