Empirical Software Engineering (SE-404)

LAB A1-G2

Laboratory Manual

Department of Software Engineering

DELHI TECHNOLOGICAL UNIVERSITY(DTU)

Shahbad Daulatpur, Bawana Road, Delhi-110042

Submitted to: -

Mr. Sanjay Patidar

Submitted by:-

Name: ASHISH KUMAR

Roll number: 2K18/SE/041

INDEX

S.No.	EXPERIMENT	DATE	REMARKS
10.	Perform a comparison of the following data analysis tools. WEKA, KEEL, SPSS, MATLAB, R.	04-01-2022	
1.	Consider any empirical study of your choice (Experiments, Survey Research, Systematic Review, Postmortem analysis and case study). Identify the following components for an 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	18-01-2022	
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 of 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.	25-01-2022	
3.	Consider any prediction model of your choice. a. Analyze 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.	25-01-2022	
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).	01-02-2022	
9.	Demonstrate how Git can be used to perform version control?	01-02-2022	
11.	Validate the results obtained in experiment 3 using 10-cross validation, hold out validation or leave one out cross-validation.	15-02-2022	
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	23-02-2022	

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 hypothesos 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		

Empirical Software Engineering LAB – A1 G2
EXPERIMENT 6

- ASHISH KUMAR
- 2K18/SE/041

Experiment Objective:- 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

Introduction:- A statistical hypothesis is an assumption made by the researcher about the data of the population collected for any experiment. It is not mandatory for this assumption to be true every time. Hypothesis testing, in a way, is a formal process of validating the hypothesis made by the researcher.
In order to validate a hypothesis, it will consider the entire population into account. However, this is not possible practically. Thus, to validate a hypothesis, it will use random samples from a population. On the basis of the result from testing over the sample data, it either selects or rejects the hypothesis. Statistical Hypothesis Testing can be categorized into two types as below:

- **Null Hypothesis** – Hypothesis testing is carried out in order to test the validity of a claim or assumption that is made about the larger population. This claim that involves attributes to the trial is known as the Null Hypothesis. The null hypothesis testing is denoted by H_0.
- **Alternative Hypothesis** – An alternative hypothesis would be considered valid if the null hypothesis is fallacious. The evidence that is present in the trial is basically the data and the statistical computations that accompany it. The alternative hypothesis testing is denoted by H_1or H_a.

Procedure:- Statisticians use hypothesis testing to formally check whether the hypothesis is accepted or rejected. Hypothesis testing is conducted in the following manner:
1. **State the Hypotheses** – Stating the null and alternative hypotheses. The hypotheses are stated in such a way that they are mutually exclusive. That is, if one is true, the other must be false; and vice versa.

2. **Formulate an Analysis Plan** – The formulation of an analysis plan is a crucial step in this stage. It should specify the following elements.

- Significance level: Often, researchers choose significance levels equal to 0.01, 0.05, or 0.10; but any value between 0 and 1 can be used.

- Test method: Typically, the test method involves a test statistic and a sampling distribution. Computed from sample data, the test statistic might be a mean score, proportion, difference between means, difference between proportions, z-score, t statistic, chi-square, etc. Given a test statistic and its sampling distribution, a researcher can assess probabilities associated with the test statistic. If the test statistic probability is less than the significance level, the null hypothesis is rejected.

3. **Analyze Sample Data** – Calculation and interpretation of the test statistic, as described in the analysis plan.

- Test statistic: When the null hypothesis involves a mean or proportion, use either of the following equations to compute the test statistic.

  Test statistic = (Statistic - Parameter) / (Standard deviation of statistic)

  Test statistic = (Statistic - Parameter) / (Standard error of statistic)

  where Parameter is the value appearing in the null hypothesis, and Statistic is the point estimate of Parameter. As part of the analysis, you may need to compute the standard deviation or standard error of the statistic. Previously, we presented common formulas for the standard deviation and standard error. When the parameter in the null hypothesis involves categorical data, you may use a chi-square statistic as the test statistic. Instructions for computing a chi-square test statistic are presented in the lesson on the chi-square goodness of fit test.

- p-value: The P-value is the probability of observing a sample statistic as extreme as the test statistic, assuming the null hypothesis is true.

4. **Interpret Results** – Application of the decision rule described in the analysis plan. Hypothesis testing ultimately uses a p-value to weigh the strength of the evidence or in other words what the data are about the population. The p-value ranges between 0 and 1. It can be interpreted in the following way:

- A small p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, so you reject it.
- A large p-value (> 0.05) indicates weak evidence against the null hypothesis, so you fail to reject it.
- A p-value very close to the cutoff (0.05) is considered to be marginal and could go either way.

Decision Errors in R

The two types of error that can occur from the hypothesis testing:

- Type I Error – Type I error occurs when the researcher rejects a null hypothesis when it is true. The term significance level is used to express the probability of Type I error while testing the hypothesis. The significance level is represented by the symbol *α (alpha).*

- Type II Error – Accepting a false null hypothesis H0 is referred to as the Type II error. The term power of the test is used to express the probability of Type II error while testing hypothesis. The power of the test is represented by the symbol *β (beta).*

Problem Taken: Within a school district, students were randomly assigned to one of two Math teachers - Mrs. Smith and Mrs. Jones. After the assignment, Mrs. Smith had 30 students, and Mrs. Jones had 25 students.

At the end of the year, each class took the same standardized test. Mrs. Smith's students had an average test score of 78, with a standard deviation of 10; and Mrs. Jones' students had an average test score of 85, with a standard deviation of 15.

Test the hypothesis that Mrs. Smith and Mrs. Jones are equally effective teachers. Use a 0.10 level of significance. (Assume that student performance is approximately normal.)

Result:-

1. **State the hypotheses.** The first step is to state the null hypothesis and an alternative hypothesis.

   Null hypothesis: $\mu_1 - \mu_2 = 0$

   Alternative hypothesis: $\mu_1 - \mu_2 \neq 0$

   Note that these hypotheses constitute a two-tailed test. The null hypothesis will be rejected if the difference between sample means is too big or if it is too small.

2. **Formulate an analysis plan**. For this analysis, the significance level is 0.10. Using sample data, we will conduct a two-sample t-test of the null hypothesis.

3. **Analyze sample data**. Using sample data, we compute the standard error (SE), degrees of freedom (DF), and the t statistic test statistic (t).

$$SE = sqrt[(s_1^2/n_1) + (s_2^2/n_2)]$$

$$SE = sqrt[(10^2/30) + (15^2/25] = sqrt(3.33 + 9)$$

$$SE = sqrt(12.33) = 3.51$$

$$DF = (s_1^2/n_1 + s_2^2/n_2)^2 / \{ [(s_1^2 / n_1)^2 / (n_1 - 1)] + [(s_2^2 / n_2)^2 / (n_2 - 1)] \}$$

$$DF = (10^2/30 + 15^2/25)^2 / \{ [(10^2 / 30)^2 / (29)] + [(15^2 / 25)^2 / (24)] \}$$

$$DF = (3.33 + 9)^2 / \{ [(3.33)^2 / (29)] + [(9)^2 / (24)] \} = 152.03 / (0.382 + 3.375)$$
$$= 152.03/3.757 = 40.47$$

$$t = [(x_1 - x_2) - d] / SE = [(78 - 85) - 0] / 3.51 = -7/3.51 = -1.99$$

where s_1 is the standard deviation of sample 1, s_2 is the standard deviation of sample 2, n_1 is the size of sample 1, n_2 is the size of sample 2, x_1 is the mean of sample 1, x_2 is the mean of sample 2, d is the hypothesized difference between the population means, and SE is the standard error.

Since we have a two-tailed test, the P-value is the probability that a t statistic having 40 degrees of freedom is more extreme than -1.99; that is, less than -1.99 or greater than 1.99.

We use the t Distribution Calculator to find $P(t < -1.99) = 0.027$, and $P(t > 1.99) = 0.027$. Thus, the P-value = 0.027 + 0.027 = 0.054.

Interpret results. Since the P-value (0.054) is less than the significance level (0.10), we cannot accept the null hypothesis.

Type 1 error. Significance level in this case is 0.1

Specifically, the approach is appropriate because the sampling method was simple random sampling, the samples were independent, the sample size was much smaller than the population size, and the samples were drawn from a normal population.

<u>Learning from experiment:</u>- This experiment gives insights into the use of hypothetical testing on real-life examples. In this experiment we learnt how to validate whether the proposed hypothesis is correct or not using statistical hypothesis.