Implementing the Two-Sample t-test in Java: A Practical Approach

K. NagaVihari

Guest Faculty, Department of Statistics, Sri Venkateswara University, Tirupati.

Abstract:
This article outlines a useful method for putting the Two-Sample t-test into practice in Java. A statistical technique called the Two-Sample t-test is used to ascertain if the means of two independent samples differ significantly from one another. The processes and Java code required to carry out this test are described in detail in this paper. These include user input, sample statistics computation, t-statistic computation, and comparison with critical values. The approach is made understandable for both practitioners and students by providing an actual instance.

Keywords: Two-Sample t-test, Independent samples, t-statistic computation, Java implementation

Introduction:
A key tool in statistical analysis is the Two-Sample t-test, which compares the means of two independent samples to see if there is a significant difference. This article provides a comprehensive explanation of the complete Two-Sample t-test implementation process in Java, from input of data to conclusion. A real-world example is used to illustrate the implementation, showing how to calculate means, standard deviations, the t-statistic, and critical values.

Methodology: The methods for doing the Java Two-Sample t-test are described in the methodology section

1. User input: two sample values must be entered by the users. After parsing, these values are kept in arrays
2. Sample statistic calculation: using the given techniques, each sample mean and standard deviation are determined
3. t-Statistic Computation: the sample sizes, means and standard deviations are used to calculate the t-statistic

$$t = \left| \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right|$$

4. Degrees of Freedom and Critical value: after computing the degrees of freedom, a critical value for a certain significance level is determined.
5. Comparison: to determine if there is a significant difference between the sample means, the t-statistic is compared to the critical value.

SOURCE CODE:

```java
importjava.util.Scanner;
public class TwoSampleTTest {
public static void main(String[] args) {
     Scanner scanner = new Scanner(System.in);
     // Input for sample 1
System.out.println("Enter sample 1 values (separated by spaces):");
String[] sample1Input = scanner.nextLine().trim().split("\\s+");
double[] sample1 = new double[sample1Input.length];
for (inti = 0; i< sample1Input.length; i++) {
sample1[i] = Double.parseDouble(sample1Input[i]);
     }
     // Input for sample 2
System.out.println("Enter sample 2 values (separated by spaces):");
String[] sample2Input = scanner.nextLine().trim().split("\\s+");
double[] sample2 = new double[sample2Input.length];
for (inti = 0; i< sample2Input.length; i++) {
sample2[i] = Double.parseDouble(sample2Input[i]);
     }
     // Calculate sample statistics
double mean1 = calculateMean(sample1);
double mean2 = calculateMean(sample2);
double stdDev1 = calculateStandardDeviation(sample1, mean1);
double stdDev2 = calculateStandardDeviation(sample2, mean2);
int n1 = sample1.length;
int n2 = sample2.length;
     // Calculate t-statistic
doubletStatistic = calculateTStatistic(mean1, mean2, stdDev1, stdDev2, n1, n2);
     // Degrees of freedom
intdf = n1 + n2 - 2;
     // Significance level (e.g., 0.05 for 95% confidence level)
doublesignificanceLevel = 0.05;
     // Get critical value from t-distribution table
doublecriticalValue = getCriticalValue(df, significanceLevel);
     // Display results
System.out.println("Mean of sample 1: " + mean1);
System.out.println("Mean of sample 2: " + mean2);
System.out.println("Standard deviation of sample 1: " + stdDev1);
System.out.println("Standard deviation of sample 2: " + stdDev2);
System.out.println("Sample size of sample 1: " + n1);
System.out.println("Sample size of sample 2: " + n2);
System.out.println("Degrees of freedom: " + df);
System.out.println("T-statistic: " + tStatistic);
```

```java
System.out.println("Critical value (two-tailed test): " + criticalValue);
    // Compare t-statistic with critical value and provide conclusion
if (Math.abs(tStatistic) >criticalValue) {
System.out.println("Conclusion: There is a significant difference between the means of the two samples at the " + (significanceLevel * 100) + "% significance level.");
    } else {
System.out.println("Conclusion: There is no significant difference between the means of the two samples at the " + (significanceLevel * 100) + "% significance level.");
    }
    // Close scanner
scanner.close();
  }
  // Method to calculate mean
public static double calculateMean(double[] data) {
double sum = 0.0;
for (double num : data) {
sum += num;
    }
return sum / data.length;


  // Method to calculate sample standard deviation
public static double calculateStandardDeviation(double[] data, double mean) {
double sum = 0.0;
for (double num : data) {
sum += Math.pow(num - mean, 2);
    }
returnMath.sqrt(sum / (data.length - 1));
  }
  // Method to calculate t-statistic for two-sample t-test
public static double calculateTStatistic(double mean1, double mean2, double stdDev1, double stdDev2, int n1, int n2) {
return (mean1 - mean2) / Math.sqrt((stdDev1 * stdDev1 / n1) + (stdDev2 * stdDev2 / n2));
  }
  // Method to get critical value from t-distribution table
public static double getCriticalValue(intdf, double significanceLevel) {
    // Critical value for two-tailed test (e.g., 0.05 significance level)
    // Using t-distribution table or calculator
    // For simplicity, assuming a large enough sample size where t-distribution approximates to normal distribution
return 1.96; // For 95% confidence level, two-tailed test
  }
}
```

OUTPUT:

Enter sample 1 values (separated by spaces):

12 52 63 45 12 36 45 25 36 45 12 36 46 32 45

Enter sample 2 values (separated by spaces):

5 62 36 14 25 36 35 25 45 36 45 23 56 32 43

Mean of sample 1: 36.13333333333333

Mean of sample 2: 34.53333333333333

Standard deviation of sample 1: 15.29176560237879

Standard deviation of sample 2: 14.946890104952397

Sample size of sample 1: 15

Sample size of sample 2: 15

Degrees of freedom: 28

T-statistic: 0.28979433122552234

Critical value (two-tailed test): 1.96

Result:

There is no significant difference between the means of the two samples at the 5.0% significance level.

Conclusion:

This article provides an extensive guide for putting the Two-Sample t-test into practice in Java. Practitioners and students can efficiently determine whether there is significant variation between the means of two independent samples by following the steps provided and utilizing the code provided. This useful method improves understanding and real-world application of the Two-Sample t-test.

References:

1. Student.(1908). The Probable error of a mean. Biometrika, 6(1), 1-25.
2. Welch, B.L.(1947). The Generalization of "Student's" problem when several different population variances are involved. Biometrika, 34(1/2), 28-35.
3. Gosset, W.S.(1908). The Probable error of a mean. Biometrika, 6(1), 1-25.