

Scenario 1:

A company wants to analyze whether there is a significant difference in the average sales across four regions using One-Way ANOVA.

Hint: (A retail company has four regions: **North, South, East, and West**. The company wants to check if the **average sales** are different across these regions.)

Ans:

Step 1: Collect the sales data and preprocess them by handling missing value.

Step 2: Group the sales data according to the regions.

Step 3: Calculate the mean of each region group to check the average sales.

Step 4: Perform f_oneway test to check the variance of mean between the groups and checks atleast one group mean has significant difference.

Step 5: if the p_value is less than 0.05, then it rejects the null hypothesis and there is a significant differences. If not, it fails to reject the null hypothesis and there is no significant differences.

Scenario 2:

A company wants to analyze whether the employee performance scores are affected by the type of training program attended and the employee's experience level. How can Data Science concepts be applied using a Two-Way ANOVA?

Ans:

Step 1: Collect the employee performance score data with program details and preprocess the data by handling the missing value and categorical data.

Step 2: Use hypothesis formula such as null and alternative hypotheses to check the difference between the variable.

Step 3: Perform two way Anova test to calculate F-statistics and p_value.

Step 4: if the p_value is less than 0.05, then it rejects the null hypothesis and there is a significant improvement in the employee performance. If not, it fails to reject the null hypothesis and there is no improvement.

Scenario 3:

A real estate company wants to build a model to predict house prices using features like area, number of bedrooms, number of bathrooms, and overall house size. How can Data Science concepts be applied to check for multicollinearity using VIF?

Ans:

Step 1: Collect the data and preprocess them by handling missing value and categorical data by encoding.

Step 2: To check multicollinearity, use Variance inflation factor to calculate how much inflation had happen in variance between the variables.

Step 3: If the score is 1 - no correlated, 1-5 - medium correlated or more than 5 is highly correlated.

Step 5. if its is highly correlated use feature selection technique to reduce multicollinearity.

Scenario 4:

A school tested two teaching methods: one traditional and one interactive. After the course, both groups took the same exam. The school wants to know if there's a significant difference in their average scores.

Ans:

Step 1: Collect the student group of two teaching methods.

Step 2: Formulate the hypothesis using null and alternative hypothesis to check the significant differences.

Step 3: Perform T-test. And Check the p_value is less than 0.05. if yes, then there is a significant differences in their average score. If no, then there is no significant differences in their score.

Scenario 5:

A fitness center wants to understand if increasing workout time leads to more calorie burn. How can data science concepts be applied to analyze this relationship?

Ans:

Step 1: Collect the data which has workout time increased as group 1 and normal timing as group 2 with their calorie burn details.

Step 2: Formulate the hypothesis formulae, H_0 - null hypothesis, H_1 - Alternative hypothesis.

Step 3: Use t-test to check the p_value is less than 0.05. if yes, there is a significant difference between the two groups. If no, there is no significant differences between the two groups.