

TASK:

In this assignment, a statistical analysis using analysis of variance (ANOVA) and linear models will be completed in Python. By working with a real-world dataset, students will apply ANOVA to examine differences in means across categorical variables and build linear models to predict the dependent variable. Then, assess the assumptions and limitations of ANOVA and linear models, proposing and implementing solutions to address these limitations, to improve the validity and accuracy analyses.

Task 1:

Research and find a dataset with one or more categorical variables and apply ANOVA to determine if there is a significant difference in the mean of the dependent variable across different levels of the categorical variable. Interpret the results and provide a detailed report with appropriate visualizations.

Task 2:

Develop a linear model with one or more independent variables that predicts the dependent variable of the same dataset used in Task 1 above. Interpret the coefficients and evaluate the significance of the model using the F-test. Provide a detailed report with appropriate visualizations.

Task 3:

Discuss the assumptions and limitations of ANOVA and linear models, and how these can affect the validity of the results. Propose possible solutions to address or mitigate these limitations and explain how these solutions can be implemented in Python.

Task 4:

Create a mini project that combines ANOVA and linear models to make predictions and draw conclusions from the same dataset used in Task 1 above. The project should include the following:

1. A detailed description of the problem and the research question.
2. A detailed data analysis using ANOVA and linear models to address the research question.
3. A discussion of the limitations and assumptions of the models used and possible solutions to address them.
4. A conclusion that summarizes the findings and implications of the research question