

TASK:

In this assignment, you will perform hypothesis testing, calculate correlation coefficients, build linear regression models, and diagnose potential issues in the models using Python.

Tasks:

1. **Hypothesis:** Locate a dataset containing the heights, weights, and ages of at least 500 individual males and females. Conduct a hypothesis test to determine whether there is a significant difference in the mean weight between males and females. Write a report discussing your findings, including relevant statistics, visualizations, and interpretations. Use Python to perform the analysis and include the relevant code in a Jupyter notebook.
2. **Correlation Coefficient:** Using the same dataset as in Task 1, calculate the correlation coefficient between height and weight. Interpret the coefficient and visualize the relationship between the two variables using a scatter plot. Write a report discussing your findings, including relevant statistics, visualizations, and interpretations. Use Python to perform the analysis and include the relevant code in a Jupyter notebook.
3. **Linear Regression Model:** Using the same dataset as in Task 1, build a linear regression model to predict weight based on height. Perform model verification to determine potential issues, such as heteroscedasticity or multicollinearity, and address any identified issues. Write a report discussing your findings, including relevant statistics, visualizations, and interpretations. Use Python to perform the analysis and include the relevant code in a Jupyter notebook.
4. **Multiple Regression Model:** Continuing with the dataset from Task 3, build a multiple regression model to predict weight based on height and age. Perform model verification to determine potential issues, such as heteroscedasticity or multicollinearity, and address any identified issues. Write a report discussing your findings, including relevant statistics, visualizations, and interpretations. Use Python to perform the analysis and include the relevant code in a Jupyter notebook.
5. **Solutions:** Identify and discuss two common assumptions of linear regression models. Using a dataset of your choice, build a linear regression model that violates one of these assumptions. Perform model verification to determine any violation and propose a solution to address the issue. Write a report discussing your findings, including relevant statistics, visualizations, and interpretations. Use Python to perform the analysis and include the relevant code in a Jupyter notebook.
6. **Nonlinear Model:** Using the same dataset as in Task 5, propose a non-linear model to predict the response variable. Compare the performance of the non-linear model to that of the linear regression model built in Task 5, using appropriate metrics. Write a report discussing your findings, including relevant statistics, visualizations, and interpretations. Use Python to perform the analysis and include the relevant code in a Jupyter notebook.