Module 2: Data Analytics with Python - Statistics

Assignment 2.2: Advanced Statistics

Objective:

ANOVA (Analysis of Variance) F-test

Task 1: Understanding ANOVA (Analysis of Variance)

Instructions:

- ANOVA (Analysis of Variance) is a statistical inference test that lets you compare multiple groups at the same time. It works well when dealing with more than two groups and is used to compare their means based on their mean similarity and f-score.
- The F-distribution does not have any negative values because between and within-group variability are always positive due to squaring each deviation.

```
%matplotlib inline
import pandas as pd
import statsmodels.api as sm
from statsmodels.formula.api import ols

# Load the data
url =
"https://raw.githubusercontent.com/bluedataconsulting/AIMasteryProg
ram/main/Lab_Exercises/Module2/PlantGrowth.csv"
data = pd.read_csv(url)

# Set up the model using ols
mod = ols('weight ~ group', data=data).fit()

# Perform ANOVA
aov_table = sm.stats.anova_lm(mod, type=2)
print(aov table)
```

Expected Output:

- The ANOVA table displaying the degrees of freedom, sum of squares, mean squares, F-value, and p-value.

Task 2: Interpreting the ANOVA Results

Instructions:

- **p > 0.05** implies that the groups are similar.
- **p < 0.05** implies that the groups are not similar.
- Based on the output from the ANOVA table, interpret whether the groups are similar or not.

Expected Output:

- A conclusion on the similarity of the groups based on the p-value obtained from the ANOVA table.