```
# import library
import os
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import statsmodels.api as sm
from scipy import stats
from scipy.stats import t as t_dist
from matplotlib.lines import Line2D
from statsmodels.formula.api import ols
from IPython import get_ipython
from IPython.display import display
from statsmodels.stats.multicomp import pairwise_tukeyhsd

# Define the menu
def read_data():
    path = input("[>] Enter the Excel file path (for example: data.xlsx): ")
    try:
        data = pd.read_excel(path)

        print("[=] Data read successfully! [=]")
        return data
    except Exception as e:

        print(f"[x] Data read not successfully! [x]")
        return None

def show_data(data):
    if data is not None:
        pd.set_option('display.max_rows', None)
        pd.set_option('display.max_columns', None)
        pd.set_option('display.width', None)
        pd.set_option('display.max_colwidth', None)

        print("\n==================================")
        print("|                                |")
        print("|      DISPLAY OF INPUT DATA     |")
        print("|                                |")
        print("=================================\n")
        print(data)
    else:

        print("[x] There is no data available. Please read the data first. [x]")

def onewayANOVA(data):
    if data is not None:
        print("\n==================================")
        print("|                                |")
        print("|    DISPLAY OF ONE WAY ANOVA    |")
        print("|                                |")
        print("=================================\n")
        print("\n[>] Available columns: ", list(data.columns))
        responses = input("[1] Enter the response column: ")
        factors = input("[2] Enter a factor column: ")
        try:
            alpha = float(input("[3] Enter the significance level (example: 0.05): ")
            if not (0 < alpha < 1):
                raise ValueError("The significance level should be between 0 and 1.")
        except ValueError as e:
            print(f"[x] {e} [x]")
            return

        if responses in data.columns and factors in data.columns:
            formula = f"{responses} ~ C({factors})"
            model = ols(formula, data=data).fit()
            anovaTable = sm.stats.anova_lm(model, typ=2)
            anovaTable = anovaTable.rename(columns={'sum_sq': 'Sum of Squares', 'df':
            anovaTable["Mean Squares"] = anovaTable["Sum of Squares"] / anovaTable["d
            SSA = anovaTable.loc[f"C({factors})", "Sum of Squares"]
            SSW = anovaTable.loc["Residual", "Sum of Squares"]
            SST = SSA + SSW
            factorInfo = data[factors].value_counts().reset_index()
            factorInfo.columns = ["Factor Levels", "Counts"]
            groupStats = data.groupby(factors)[responses].agg(N="count", Mean="mean",
            groupStats["MOE"] = groupStats.apply(lambda row: t_dist.ppf(1 - alpha / 2
            groupStats["CI Lower"] = groupStats["Mean"] - groupStats["MOE"]
            groupStats["CI Upper"] = groupStats["Mean"] + groupStats["MOE"]

            # Result of One-Way ANOVA
```

File ×

..

.config

.ipynb_checkpoints

drive

sample_data

.ipynb_checkpoints

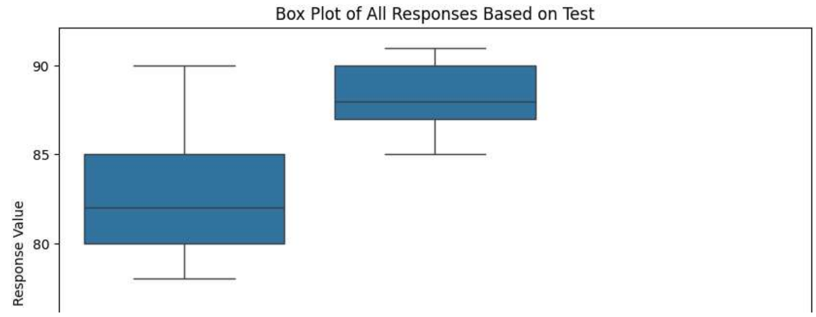
onewaydata.xlsx

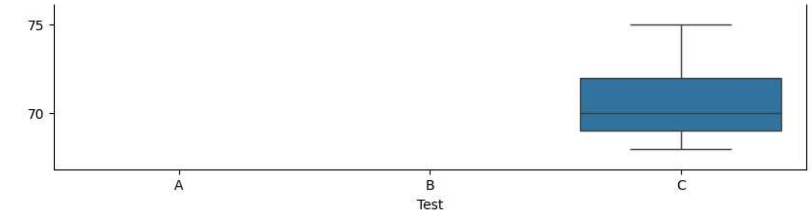
twowaydata.xlsx

```
            print("\n==================================")
            print("|                                |")
            print("|    RESULTS OF ONE WAY ANOVA    |")
            print("|                                |")
            print("=================================\n")
            print(f"Method: One-Way ANOVA")
            print(f"(H0): There is no difference in means between groups.")
            print(f"(H1): There is at least one group that is different.")
            print(f"Significance Level: {alpha}")

            print("\n[=] Factor Information: [=]")
            print(factorInfo.to_string(index=False))

            print("\n[=] One-Way ANOVA Summary Table [=]")
            print(f"Term            Sum of Squares           df            Mean Squar
            print(f"SSA             {SSA:.4f}          {anovaTable['df'].iloc[0]:.0f}
            print(f"SSW (Residual)  {SSW:.4f}          {anovaTable['df'].iloc[1]:.0f}
            print(f"SST (Total)     {SST:.4f}          {anovaTable['df'].iloc[0] + an

            print("\n[=] Means [=]")
            print(groupStats.to_string(index=False))

            # Post-Hoc Test (Tukey)
            tukey = pairwise_tukeyhsd(endog=data[responses], groups=data[factors], al
            print("\n[=] Post-Hoc Test (Tukey) [=]")
            print(tukey)

            # Conclusion
            f_test = anovaTable['F'].iloc[0]
            f_table = stats.f.ppf(1 - alpha, dfn=anovaTable['df'].iloc[0], dfd=anovaT
            print("\n[=] Decision [=]")
            if f_test > f_table:
                print(f"Reject H0: There is a significant difference between the grou
            else:
                print(f"Failure to reject H0: There is no significant difference betw
        else:

            print("[x] Invalid column. Please try again. [x]")
    else:

        print("[x] No data is available. Please load the data first. [x]")


def twowayANOVA(data):
    if data is not None:
        print("\n==================================")
        print("|                                |")
        print("|    DISPLAY OF TWO WAY ANOVA    |")
        print("|                                |")
        print("=================================\n")
        print("\n[>] Available columns: ", list(data.columns))

        responses2 = input("[1] Enter the responses column (separate with commas): ")
        responses2 = [col.strip() for col in responses2]
        factors2 = input("[2] Enter a factors column (separate with commas): ").strip
        factors2 = [col.strip() for col in factors2]

        if len(factors2) != 2:
            print("[x] Please provide exactly two factors for two-way ANOVA. [x]")
            return

        if not all(col in data.columns for col in factors2 + responses2):
            print("[x] Some columns are invalid. Please check your input. [x]")
            return
        alpha = float(input("[3] Enter the significance level (example: 0.05): "))

        # Result of Two-Way ANOVA

        print("\n==================================")
        print("|                                |")
        print("|    RESULTS OF TWO WAY ANOVA    |")
        print("|                                |")
        print("=================================\n")
        print("\n[=] Factor Information [=]")
        for factor in factors2:
            print(f"{'Factor:':<15} {factor:<20} {'Levels:':<8} {data[factor].nunique

        for response in responses2:
            print(f"[=] Analysis for response {response} [=]")
            formula = f"{response} ~ " + " + ".join([f"C({factor})" for factor in fac
                      " * ".join([f"C({factor})" for factor in factors2])
            model = ols(formula, data=data).fit()
            anovaTable = sm.stats.anova_lm(model, typ=2)
            anovaTable = anovaTable.rename(columns={'sum_sq': 'Sum of Squares', 'df':
            SSA = anovaTable.loc['C(' + factors2[0] + ')', 'Sum of Squares']
            SSB = anovaTable.loc['C(' + factors2[1] + ')', 'Sum of Squares']
            SSAB = anovaTable.loc['C(' + factors2[0] + '):C(' + factors2[1] + ')', 'S
            SSW = anovaTable.loc['Residual', 'Sum of Squares']
            SST = SSA + SSB + SSAB + SSW

            a = data[factors2[0]].nunique()
            b = data[factors2[1]].nunique()
            n = len(data) / (a * b)

            df_ab = (a - 1) * (b - 1)
            df_error = a * b * (n - 1)
            df_total = a * b * n - 1

            print("\n[=] Two-Way ANOVA Summary Table [=]")
            print(f"{'Source':<20} {'DF':<10} {'Adj SS':<10} {'Adj MS':<10} {'F-Value
            print(f"{'SSA':<20} {a - 1:<10} {SSA:<10.4f} {SSA/(a - 1):<10.4f} {SSA/(a
            print(f"{'SSB':<20} {b - 1:<10} {SSB:<10.4f} {SSB/(b - 1):<10.4f} {SSB/(b
            print(f"{'Interaction AB':<20} {df_ab:<10} {SSAB:<10.4f} {SSAB/df_ab:<10.
            print(f"{'SSW (Residuals)':<20} {df_error:<10} {SSW:<10.4f} {SSW/df_error
            print(f"{'SST (Total)':<20} {df_total:<10} {SST:<10.4f} {SST/df_total:<10

            ssTotal = anovaTable["Sum of Squares"].sum()

            mse = anovaTable.loc["Residual", "Sum of Squares"] / anovaTable.loc["Resi
            ssModal = ssTotal - anovaTable.loc["Residual", "Sum of Squares"]
            s = mse**0.5
            ssModal = ssTotal - anovaTable.loc["Residual", "Sum of Squares"]
            rSq = ssModal / ssTotal
            rSq_adj = 1 - (1 - rSq) * ((len(data) - 1) / (len(data) - len(model.param


            print("\n[=] Model Summary [=]")
            print(f"{'Standard Error (S):':<25} {s:.4f}")
            print(f"{'R-squared:':<25} {rSq:.4f}")
            print(f"{'Adjusted R-squared:':<25} {rSq_adj:.4f}")

            print("\n[=] Coefficients [=]")
            print(f"{'Term':<25} {'Coef':<10} {'SE Coef':<10} {'T-Value':<10} {'P-Val
            for term, coef, se_coef, t_val, p_val in zip(model.params.index, model.pa
                                                         model.bse.values, model.tval
                                                         model.pvalues.values):
                vif = 1 / (1 - model.rsquared)
                print(f"{term:<25} {coef:<10.4f} {se_coef:<10.4f} {t_val:<10.4f} {p_v

            print("\n[=] Decision [=]")
            for idx, row in anovaTable.iterrows():
                f_val = row["F"]
                p_val = row["P-value"]
                if p_val < alpha:
                    print(f"Reject H0 for {idx}: There is a significant difference (F
                else:
                    print(f"Fail to reject H0 for {idx}: No significant difference (F
    else:

        print("[x] No data is available. Please load the data first. [x]")


def summary_data(data):
    if data is not None:
        print("\n=================================")
        print("|                                |")
        print("|          SUMMARY DATA          |")
        print("|                                |")
        print("=================================\n")
        print("[>] Number of Rows: ", data.shape[0])
        print("[>] Number of Columns: ", data.shape[1])
        print("[>] Columns: ", list(data.columns))
        print("\n[=] Statistical Description [=]")
        print(data.describe())
    else:

        print("[x] No data is available. Please load the data first. [x]")


def plot_data(data):
    if data is not None:
        print("\n=================================")
        print("|                                |")
        print("|          SUMMARY DATA          |")
        print("|                                |")
        print("=================================\n")
        print(                                  \n")
        print("[>] Columns: ", list(data.columns))

        factors = input("[1] Enter the factor column for the X axis : ")
        if factors not in data.columns:
            print(f"[x] The column ‘{factors}’ does not exist in the data. [x]")
            return
        numericCols = data.select_dtypes(include=[np.number]).columns.tolist()
        if factors in numericCols:
            numericCols.remove(factors)
        meltedData = data.melt(id_vars=[factors], value_vars=numericCols, var_name='R
        plt.figure(figsize=(10, 6))
        ax = sns.boxplot(x=factors, y='Value', hue='Response Variable', data=meltedDa
        plt.title(f'Box Plot of All Responses Based on {factors}')
        plt.xlabel(factors)
        plt.ylabel('Response Value')

        for line in ax.artists:
            avgLine = line.get_paths()[0]
            ax.plot(avgLine.vertices[:, 0], avgLine.vertices[:, 1], linestyle='--', c
        plt.show()

    else:
        print("[x] No data is available. Please load the data first. [x]")

def main():
    data = None
    while True:
        print("1. Read Data")
        print("2. Display Data")
        print("3. Data Summary")
        print("4. One-Way ANOVA")
        print("5. Two-Way ANOVA")
        print("6. Data Plot")
        print("7. Exit")
        choice = input("[>] Enter Choice (1-7): ")

        if choice == "1":
            data = read_data()
        elif choice == "2":
            show_data(data)
        elif choice == "3":
            summary_data(data)
        elif choice == "4":
            onewayANOVA(data)
        elif choice == "5":
            twowayANOVA(data)
        elif choice == "6":
            plot_data(data)
        elif choice == "7":
            print("[!] Thank you for using this program [!]")
            break
        else:
            print("[x] Invalid selection. [x]")

if __name__ == "__main__":
    main()
```

```
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): 1
[>] Enter the Excel file path (for example: data.xlsx): /content/sample_data/one
[=] Data read successfully! [=]
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): 2

==================================
|                                |
|      DISPLAY OF INPUT DATA     |
|                                |
=================================

   Test  Score
0     A     80
1     A     85
2     A     90
3     A     78
4     A     82
5     B     88
6     B     90
7     B     85
8     B     87
9     B     91
10    C     70
11    C     72
12    C     68
13    C     75
14    C     69
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): 3

==================================
|                                |
|          SUMMARY DATA          |
|                                |
=================================

[>] Number of Rows:  15
[>] Number of Columns:  2
[>] Columns:  ['Test', 'Score']

[=] Statistical Description [=]
           Score
count  15.000000
mean   80.666667
std     8.191168
min    68.000000
25%    73.500000
50%    82.000000
75%    87.500000
max    91.000000
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): 4

==================================
|                                |
|    DISPLAY OF ONE WAY ANOVA    |
|                                |
=================================


[>] Available columns:  ['Test', 'Score']
[1] Enter the response column: Score
[2] Enter a factor column: Test
[3] Enter the significance level (example: 0.05): 0.05
```

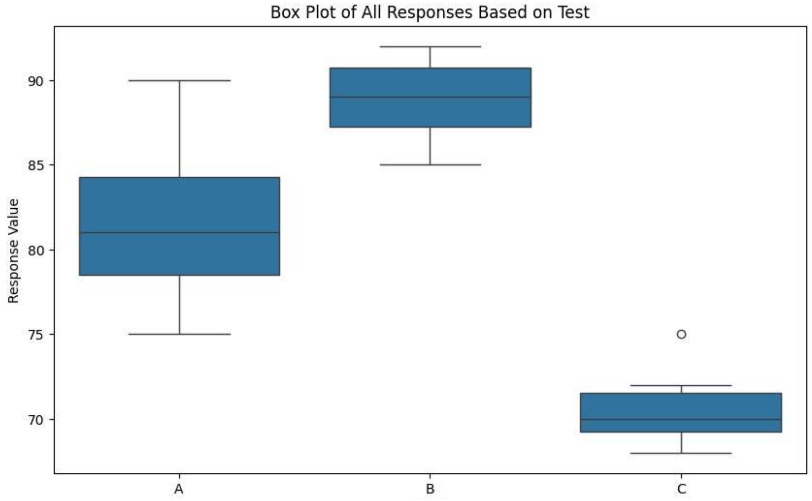
```
==================================
|                                |
|    RESULTS OF ONE WAY ANOVA    |
|                                |
=================================

Method: One-Way ANOVA
(H0): There is no difference in means between groups.
(H1): There is at least one group that is different.
Significance Level: 0.05

[=] Factor Information: [=]
Factor Levels  Counts
            A       5
            B       5
            C       5

[=] One-Way ANOVA Summary Table [=]
Term            Sum of Squares           df            Mean Squares
SSA             797.7333          2          398.8667     33.8023     0.0000
SSW (Residual)  141.6000          12          11.8000
SST (Total)     939.3333          14

[=] Means [=]
Test  N  Mean    StDev      MOE  CI Lower  CI Upper
   A  5  83.0 4.690416 5.823920 77.176080 88.823920
   B  5  88.2 2.387467 2.964432 85.235568 91.164432
   C  5  70.8 2.774887 3.445478 67.354522 74.245478

[=] Post-Hoc Test (Tukey) [=]
 Multiple Comparison of Means - Tukey HSD, FWER=0.05
======================================================
group1 group2 meandiff p-adj   lower    upper   reject
------------------------------------------------------
     A      B      5.2 0.0805  -0.5961  10.9961  False
     A      C    -12.2 0.0003 -17.9961  -6.4039   True
     B      C    -17.4    0.0 -23.1961 -11.6039   True
------------------------------------------------------

[=] Decision [=]
Reject H0: There is a significant difference between the groups (F_test = 33.80
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): 6

==================================
|                                |
|          SUMMARY DATA          |
|                                |
=================================

[>] Columns:  ['Test', 'Score']
[1] Enter the factor column for the X axis : test
[x] The column ‘test’ does not exist in the data. [x]
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): 6

==================================
|                                |
|          SUMMARY DATA          |
|                                |
=================================

[>] Columns:  ['Test', 'Score']
[1] Enter the factor column for the X axis : Test
```





```
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): 1
[>] Enter the Excel file path (for example: data.xlsx): /content/sample_data/tw
[=] Data read successfully! [=]
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): 2

==================================
|                                |
|      DISPLAY OF INPUT DATA     |
|                                |
=================================

   Test  Gender  Score
0     A    Male     80
1     A    Male     85
2     A    Male     90
3     A  Female     75
4     A  Female     78
5     A  Female     82
6     B    Male     88
7     B    Male     92
8     B    Male     91
9     B  Female     85
10    B  Female     87
11    B  Female     90
12    C    Male     70
13    C    Male     72
14    C    Male     75
15    C  Female     68
16    C  Female     69
17    C  Female     70
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): 3

==================================
|                                |
|          SUMMARY DATA          |
|                                |
=================================

[>] Number of Rows:  18
[>] Number of Columns:  3
[>] Columns:  ['Test', 'Gender', 'Score']

[=] Statistical Description [=]
           Score
count  18.000000
mean   80.388889
std     8.444552
min    68.000000
25%    72.750000
50%    81.000000
75%    87.750000
max    92.000000
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): 5
```

```
==================================
|                                |
|    DISPLAY OF TWO WAY ANOVA    |
|                                |
=================================


[>] Available columns:  ['Test', 'Gender', 'Score']
[1] Enter the responses column (separate with commas): Score
[2] Enter a factors column (separate with commas): Test, Gender
[3] Enter the significance level (example: 0.05): 0.05

==================================
|                                |
|    RESULTS OF TWO WAY ANOVA    |
|                                |
=================================


[=] Factor Information [=]
Factor:         Test                Levels:  3    Values:  A, B, C
Factor:         Gender              Levels:  2    Values:  Male, Female
[=] Analysis for response Score [=]

[=] Two-Way ANOVA Summary Table [=]
Source               DF         Adj SS     Adj MS     F-Value    P-Value
SSA                  2          1004.7778  502.3889   54.4759    0.0000
SSB                  1          84.5000    84.5000    9.1627     0.0105
Interaction AB       2          12.3333    6.1667     0.6687     0.5305
SSW (Residuals)      12.0       110.6667   9.2222     -
SST (Total)          17.0       1212.2778  71.3105    -

[=] Model Summary [=]
Standard Error (S):       3.0368
R-squared:                0.9087
Adjusted R-squared:       0.8707

[=] Coefficients [=]
Term                      Coef       SE Coef    T-Value    P-Value    VIF
Intercept                 78.3333    1.7533     44.6776    0.0000     10.9543
C(Test)[T.B]              9.0000     2.4795     3.6297     0.0035     10.9543
C(Test)[T.C]              -9.3333    2.4795     -3.7641    0.0027     10.9543
C(Gender)[T.Male]         6.6667     2.4795     2.6887     0.0197     10.9543
C(Test)[T.B]:C(Gender)[T.Male] -3.6667    3.5066     -1.0456    0.3163     10.9
C(Test)[T.C]:C(Gender)[T.Male] -3.3333    3.5066     -0.9506    0.3606     10.9

[=] Decision [=]
Reject H0 for C(Test): There is a significant difference (F = 54.476, P = 0.000
Reject H0 for C(Gender): There is a significant difference (F = 9.163, P = 0.01
Fail to reject H0 for C(Test):C(Gender): No significant difference (F = 0.669, 
Fail to reject H0 for Residual: No significant difference (F = nan, P = nan)
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): 6

==================================
|                                |
|          SUMMARY DATA          |
|                                |
=================================

[>] Columns:  ['Test', 'Gender', 'Score']
[1] Enter the factor column for the X axis : Test
```



Test

```
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): 6

==================================
|                                |
|          SUMMARY DATA          |
|                                |
=================================

[>] Columns:  ['Test', 'Gender', 'Score']
[1] Enter the factor column for the X axis : 6
[x] The column ‘6’ does not exist in the data. [x]
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): Gender
[x] Invalid selection. [x]
1. Read Data
2. Display Data
3. Data Summary
4. One-Way ANOVA
5. Two-Way ANOVA
6. Data Plot
7. Exit
[>] Enter Choice (1-7): 7
[!] Thank you for using this program [!]
```