Anova

May 16, 2025

1 Import Necessary libraries

[1]:
```
import scipy.stats as stats
import pandas as pd
import statsmodels.api as sm
from statsmodels.formula.api import ols
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

[4]:
```
# Scores from three different teaching methods
group_A = [70, 68, 72, 65, 69]  # Traditional lectures
group_B = [78, 74, 80, 76, 77]  # Online videos
group_C = [85, 87, 90, 88, 86]  # Interactive workshops
```

2 Performing One-Way ANOVA

- Using scipy.stats.f_oneway

[5]:
```
f_statistic, p_value = stats.f_oneway(group_A, group_B, group_C)

print("F-statistic:", round(f_statistic, 2))
print("p-value:", round(p_value, 4))
```

```
F-statistic: 82.77
p-value: 0.0
```

3 Interpreting the results

[6]:
```
if p_value < 0.05:
    print("Result: Significant differences exist between the group means.")
else:
    print("Result: No significant differences between the group means.")
```

```
Result: Significant differences exist between the group means.
```

4 Generating Data for ANOVA

[15]:
```
np.random.seed(0)
# Sample data
data = {
    'Score': [70, 68, 72, 65, 69,   # Lecture-Male
              78, 74, 80, 76, 77,   # Video-Male
              85, 87, 90, 88, 86,   # Workshop-Male
              75, 73, 78, 74, 76,   # Lecture-Female
              80, 77, 82, 79, 78,   # Video-Female
              88, 86, 91, 89, 87],  # Workshop-Female

    'Method': ['Lecture']*5 + ['Video']*5 + ['Workshop']*5 +
              ['Lecture']*5 + ['Video']*5 + ['Workshop']*5,

    'Gender': ['Male']*15 + ['Female']*15
}

df = pd.DataFrame(data)
df.sample(10)
```

[15]:

	Score	Method	Gender
2	72	Lecture	Male
28	89	Workshop	Female
13	88	Workshop	Male
10	85	Workshop	Male
26	86	Workshop	Female
24	78	Video	Female
27	91	Workshop	Female
11	87	Workshop	Male
17	78	Lecture	Female
22	82	Video	Female

4.1 One way Anova For (Score ~ Method)

4.1.1 Using Scipy stats method

[33]:
```
methods = df["Method"].unique()
scores_by_methods = [df[df["Method"] == method]["Score"] for method in methods]

f_stat, p_value = stats.f_oneway(*scores_by_methods)
print(f"F-statistic: {f_stat}, P-value: {p_value}")
```

```
F-statistic: 75.84618834080717, P-value: 8.317244651215984e-12
```

4.1.2 Using OLS Method and statsmodels library

[ ]:
```
model = ols("Score ~ C(Method)", data=df).fit()
sm.stats.anova_lm(model, typ=2)
```

[ ]:

	sum_sq	df	F	PR(>F)
C(Method)	1252.866667	2.0	75.846188	8.317245e-12
Residual	223.000000	27.0	NaN	NaN

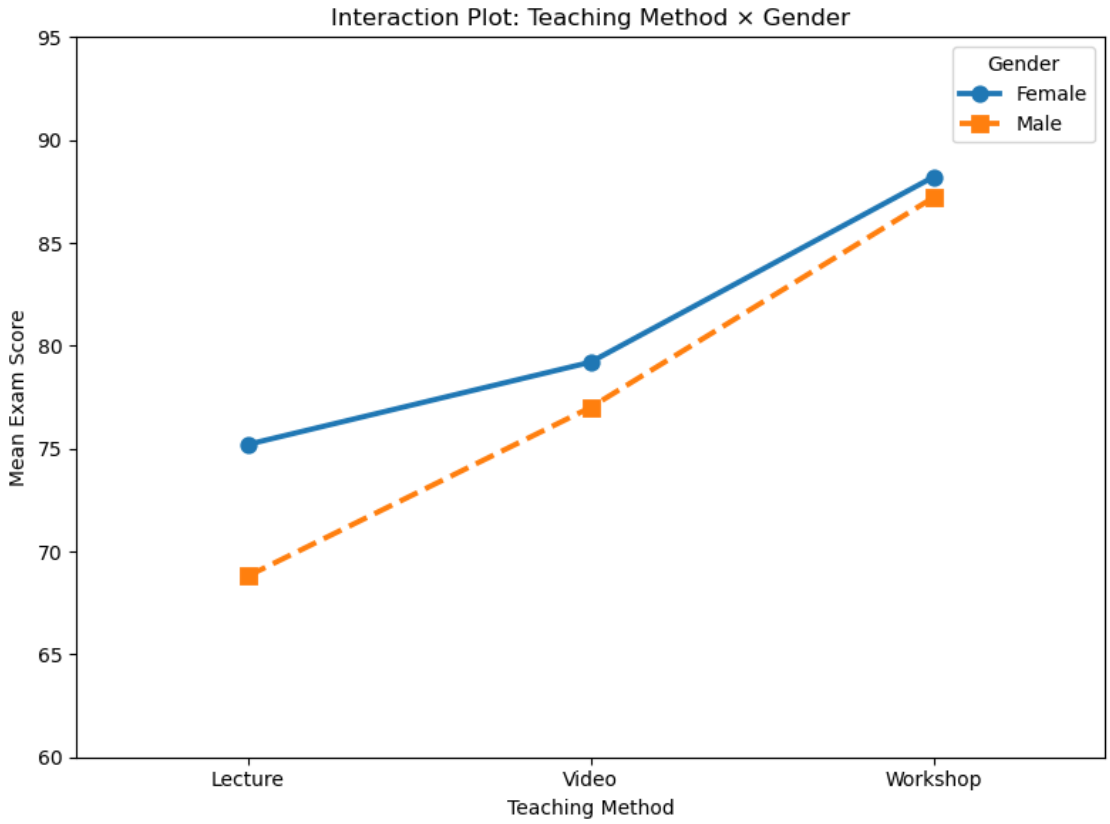
4.2 2 way Anova

1. **Dependent Variable:** Score
2. **Independent Variable / Features:** Method and Gender

4.3 Visualizing interaction effect between (Gender and Method)

[36]:
```
# Teaching Method VS Gender
mean_scores = df.groupby(["Method", "Gender"])["Score"].mean().reset_index()
plt.figure(figsize=(8, 6))
sns.pointplot(
    data=mean_scores,
    x="Method",
    y="Score",
    hue="Gender",
    markers=["o", "s"],
    linestyles=["-", "--"],
)

# Customize plot
plt.title("Interaction Plot: Teaching Method × Gender")
plt.ylabel("Mean Exam Score")
plt.xlabel("Teaching Method")
plt.ylim(60, 95)
plt.legend(title="Gender")
plt.tight_layout()
plt.show()
```



[39]:
```
# Build the model for two-way ANOVA
model = ols('Score ~ C(Method) + C(Gender) + C(Method):C(Gender)', data=df).
  ↪fit() #Score ~ C(Method) * C(Gender)
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Score   R-squared:                       0.928
Model:                            OLS   Adj. R-squared:                  0.913
Method:                 Least Squares   F-statistic:                     62.03
Date:                Thu, 15 May 2025   Prob (F-statistic):           6.16e-13
Time:                        11:47:40   Log-Likelihood:                -61.502
No. Observations:                  30   AIC:                             135.0
Df Residuals:                      24   BIC:                             143.4
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                                                 coef    std err          t
P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
```

```
---------------------------
Intercept                                   75.2000      0.940     80.012
0.000      73.260      77.140
C(Method)[T.Video]                           4.0000      1.329      3.009
0.006       1.257       6.743
C(Method)[T.Workshop]                       13.0000      1.329      9.781
0.000      10.257      15.743
C(Gender)[T.Male]                           -6.4000      1.329     -4.815
0.000      -9.143      -3.657
C(Method)[T.Video]:C(Gender)[T.Male]         4.2000      1.880      2.234
0.035       0.320       8.080
C(Method)[T.Workshop]:C(Gender)[T.Male]      5.4000      1.880      2.873
0.008       1.520       9.280
==============================================================================
Omnibus:                        0.957   Durbin-Watson:                   2.948
Prob(Omnibus):                  0.620   Jarque-Bera (JB):                0.819
Skew:                           0.098   Prob(JB):                        0.664
Kurtosis:                       2.214   Cond. No.                         9.77
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[40]:
```
# Two-way ANOVA (Implemented same as one way except for model fitting in ols)
anova_table = sm.stats.anova_lm(model, typ=2)
anova_table
```

[40]:

	sum_sq	df	F	PR(>F)
C(Method)	1252.866667	2.0	141.833962	5.076311e-14
C(Gender)	76.800000	1.0	17.388679	3.426366e-04
C(Method):C(Gender)	40.200000	2.0	4.550943	2.110095e-02
Residual	106.000000	24.0	NaN	NaN

Please feel free to ask any questions