practical-7

October 31, 2025

[7]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import statsmodels.api as sm
from scipy import stats
from statsmodels.formula.api import ols
import statsmodels.api as sm
from statsmodels.formula.api import ols
```

[4]:
```
df = pd.read_csv("teacher_rating.csv")
df
```

[4]:

	gender	age_group	beauty	eval
0	male	old	2.78	3.70
1	female	middle	4.66	4.40
2	male	middle	6.44	6.23
3	male	middle	4.69	3.98
4	male	middle	3.84	3.99
..	…	…	…	…
95	female	young	4.96	4.91
96	female	middle	2.10	4.40
97	female	young	3.73	4.85
98	female	young	7.05	3.46
99	male	middle	6.76	3.31

[100 rows x 4 columns]

Q1. Regression with T-test: Using the teachers rating data set, does gender affect teaching evaluation rates?

[8]:
```
# Fit a linear regression model: eval ~ gender
model = ols('eval ~ C(gender)', data=df).fit()

# Display regression summary
print(model.summary())
```

OLS Regression Results

```
==============================================================================
Dep. Variable:                   eval   R-squared:                       0.004
Model:                            OLS   Adj. R-squared:                 -0.006
Method:                 Least Squares   F-statistic:                    0.3784
Date:                Fri, 31 Oct 2025   Prob (F-statistic):              0.540
Time:                        10:10:39   Log-Likelihood:                -114.51
No. Observations:                 100   AIC:                             233.0
Df Residuals:                      98   BIC:                             238.2
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      4.0343      0.103     39.301      0.000       3.831       4.238
C(gender)[T.male]    -0.0952      0.155     -0.615      0.540      -0.402       0.212
==============================================================================
Omnibus:                        3.051   Durbin-Watson:                   1.957
Prob(Omnibus):                  0.217   Jarque-Bera (JB):                2.466
Skew:                           0.264   Prob(JB):                        0.291
Kurtosis:                       3.559   Cond. No.                         2.50
==============================================================================

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

Q2. Regression with ANOVA: Using the teachers' rating data set, does beauty score for instructors differ by age?

[9]:
```
# Assuming columns: 'beauty' and 'age_group'
model = ols('beauty ~ C(age_group)', data=df).fit()
anova_table = sm.stats.anova_lm(model, typ=2)
print(anova_table)
```

```
                 sum_sq    df         F    PR(>F)
C(age_group)    1.039123   2.0  0.216923  0.805382
Residual      232.329168  97.0       NaN       NaN
```

Q3. Correlation: Using the teachers' rating dataset, Is teaching evaluation score correlated with beauty score?

[11]:
```
# Regression model: eval as dependent variable, beauty as independent variable
model = ols('eval ~ beauty', data=df).fit()
```

```python
# Display full regression summary
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   eval   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.009
Method:                 Least Squares   F-statistic:                    0.1387
Date:                Fri, 31 Oct 2025   Prob (F-statistic):              0.710
Time:                        10:15:19   Log-Likelihood:                -114.63
No. Observations:                 100   AIC:                             233.3
Df Residuals:                      98   BIC:                             238.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      3.8979      0.265     14.700      0.000       3.372       4.424
beauty         0.0187      0.050      0.372      0.710      -0.081       0.119
==============================================================================
Omnibus:                        2.461   Durbin-Watson:                   1.935
Prob(Omnibus):                  0.292   Jarque-Bera (JB):                1.860
Skew:                           0.296   Prob(JB):                        0.394
Kurtosis:                       3.309   Cond. No.                         18.8
==============================================================================

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