```python
# import required libraries
import pandas as pd
import statsmodels.api as sm
from statsmodels.formula.api import ols
from scipy import stats

import pandas as pd
import numpy as np

# new random seed for different numeric values
np.random.seed(2025)

n = 100  # number of records

# gender
gender = np.random.choice(['Male', 'Female'], size=n)

# age group
age_group = np.random.choice(['Young', 'Middle-aged', 'Old'], size=n)

# beauty scores (scale 1–10)
beauty = np.random.normal(
    loc=[7.5 if a == 'Young' else 6.3 if a == 'Middle-aged' else 5.2
for a in age_group],
    scale=1.1,
    size=n
)
beauty = np.clip(beauty, 1, 10)

# teaching evaluation (scale 1–10)
eval_score = (
    0.6 * beauty +                       # slightly stronger
relationship
    np.where(gender == 'Female', 0.5, -0.1) +  # small gender bias
    np.random.normal(0, 0.9, size=n)    # less random noise
)
eval_score = np.clip(eval_score, 1, 10)

# build DataFrame
data = pd.DataFrame({
    'gender': gender,
    'age_group': age_group,
    'beauty': beauty.round(2),
    'eval': eval_score.round(2)
})

# save to CSV
data.to_csv("teachers_rating.csv", index=False)

data.head()
```

```
  gender age_group  beauty  eval
0   Male       Old    5.59  1.78
1   Male       Old    7.71  3.96
2   Male     Young    7.13  4.13
3   Male       Old    4.31  3.23
4   Male     Young    7.47  4.23
```

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

```
# Q1. Regression with T-test: Does gender affect teaching evaluation
rates?

# assuming columns are named 'eval' (teaching evaluation) and 'gender'
(male/female)
model_ttest = ols('eval ~ gender', data=data).fit()
ttest_result = model_ttest.t_test([0, 1])  # tests coefficient of
gender variable

print("\nQ1. Regression with T-test: Does gender affect teaching
evaluation rates?")
print(model_ttest.summary())
```

```
Q1. Regression with T-test: Does gender affect teaching evaluation
rates?
                            OLS Regression Results

==============================================================================
========
Dep. Variable:                    eval   R-squared:
0.071
Model:                             OLS   Adj. R-squared:
0.061
Method:                  Least Squares   F-statistic:
7.448
Date:                 Mon, 27 Oct 2025   Prob (F-statistic):
0.00753
Time:                         23:34:23   Log-Likelihood:
-160.24
No. Observations:                 100   AIC:
324.5
Df Residuals:                      98   BIC:
329.7
Df Model:                           1

Covariance Type:             nonrobust

==============================================================================
```

```
============
                     coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------
------------
Intercept          4.5308      0.175     25.866      0.000       4.183
4.878
gender[T.Male]    -0.6629      0.243     -2.729      0.008      -1.145
-0.181
========================================================================
========
Omnibus:                        1.599   Durbin-Watson:
2.041
Prob(Omnibus):                  0.450   Jarque-Bera (JB):
1.330
Skew:                           0.282   Prob(JB):
0.514
Kurtosis:                       3.025   Cond. No.
2.67
========================================================================
========

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?

```
# Q2. Regression with ANOVA: Does beauty score differ by age group?

model_anova = ols('beauty ~ C(age_group)', data=data).fit()
anova_table = sm.stats.anova_lm(model_anova, typ=2)

print("\nQ2. Regression with ANOVA: Does beauty score for instructors
differ by age?")
print(anova_table)


Q2. Regression with ANOVA: Does beauty score for instructors differ by
age?
                   sum_sq    df          F        PR(>F)
C(age_group)   56.844902   2.0  19.91131  5.685528e-08
Residual      138.462898  97.0       NaN           NaN
```

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

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

import statsmodels.api as sm
from statsmodels.formula.api import ols

# perform OLS regression: eval vs beauty
model = ols('eval ~ beauty', data=data).fit()

# display summary
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   eval   R-squared:                       0.579
Model:                            OLS   Adj. R-squared:                  0.574
Method:                 Least Squares   F-statistic:                     134.6
Date:                Mon, 27 Oct 2025   Prob (F-statistic):           4.28e-20
Time:                        23:36:11   Log-Likelihood:                -120.70
No. Observations:                 100   AIC:                             245.4
Df Residuals:                      98   BIC:                             250.6
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.2880      0.394     -0.730      0.467      -1.070       0.494
beauty         0.6783      0.058     11.600      0.000       0.562       0.794
==============================================================================
Omnibus:                        0.576   Durbin-Watson:                   2.020
Prob(Omnibus):                  0.750   Jarque-Bera (JB):                0.714
```

```
Skew:                          -0.103   Prob(JB):
0.700
Kurtosis:                       2.641   Cond. No.
33.2
==============================================================================
========

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