

📖 시험점수 예측 OLS 분석 - 문제, 코드, 해설, 실행결과 포함

📋 문제

다음은 학생들의 공부 시간, 수면 시간, 커피 섭취량과 시험 점수에 대한 데이터이다. 선형회귀를 통해 시험 점수에 영향을 주는 요인을 분석하시오.

📁 변수 설명:

- Study_Hours: 하루 평균 공부 시간 (시간)
- Sleep_Hours: 하루 평균 수면 시간 (시간)
- Coffee_Intake: 하루 평균 커피 섭취량 (잔)
- Exam_Score: 시험 점수 (종속변수)

🧠 문제:

1. Exam_Score 를 종속변수로 하여 선형회귀 모델을 학습하시오.
2. 2. 각 독립변수의 회귀계수와 유의미한 변수(p-value 기준)를 해석하시오.
3. 3. R-squared 값을 해석하고, 전체 모델 설명력을 평가하시오.

📖 데이터셋 내용 (40 건)

Study_Hours	Sleep_Hours	Coffee_Intake	Exam_Score
2.8	6.2	3.0	30.8
7.0	5.3	2.0	58.0
5.6	6.6	0.0	55.3
2.0	7.6	2.0	26.3
3.8	7.3	1.0	44.9
8.3	7.0	0.0	66.5
0.1	9.4	0.0	7.6
4.1	7.4	3.0	60.9
7.5	8.0	0.0	67.2
3.3	9.2	0.0	29.6

3.6	5.7	3.0	42.2
4.8	6.0	0.0	36.1
8.0	8.7	0.0	70.8
3.7	6.2	1.0	31.4
4.1	7.0	0.0	40.7
4.1	8.1	0.0	46.1
9.4	7.9	0.0	78.9
9.4	8.8	2.0	86.2
7.0	8.5	3.0	72.6
5.8	8.1	2.0	58.4
6.5	6.2	1.0	54.6
8.0	7.8	1.0	71.8
3.1	7.3	0.0	43.9
7.4	5.7	2.0	58.7
2.5	8.4	2.0	28.3
3.7	7.8	1.0	36.6
6.8	7.0	2.0	69.0
2.1	6.8	2.0	34.2
4.7	5.8	0.0	52.2
3.3	7.2	3.0	40.3
4.5	7.5	2.0	52.6
-0.6	6.2	0.0	5.2
1.5	8.2	0.0	20.5
3.6	5.9	1.0	34.7
6.9	4.9	2.0	57.5
4.7	8.0	3.0	44.6
5.0	6.6	1.0	47.1
6.4	6.9	3.0	68.2
3.2	6.2	1.0	39.2
5.6	5.4	0.0	51.3

✓ 정답 코드

```
import pandas as pd
import statsmodels.api as sm

df = pd.read_csv("exam_score_data.csv")
X = df[['Study_Hours', 'Sleep_Hours', 'Coffee_Intake']]
y = df['Exam_Score']
X = sm.add_constant(X)
model = sm.OLS(y, X).fit()
print(model.summary())
```

해설

- - Study_Hours 는 양의 상관관계를 보이며, 시험 점수에 유의미한 영향을 준다.
- - Sleep_Hours 의 계수는 작거나 음수일 수 있으나, 해석은 p-value 를 기준으로 판단한다.
- - Coffee_Intake 도 성적에 긍정적인 영향이 있지만 유의성 여부는 p-value 에 따라 달라질 수 있다.
- - 모델의 결정계수(R^2)는 약 0.917 로, 이 회귀모형이 시험 점수의 약 92%를 설명함을 의미한다.

코드 실행 결과 (model.summary())

OLS Regression Results

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

```
Dep. Variable:      Exam_Score  R-squared:      0.917
Model:              OLS  Adj. R-squared:    0.910
Method:             Least Squares  F-statistic:    132.2
Date:              Sun, 15 Jun 2025  Prob (F-statistic):  1.71e-19
Time:              09:27:38  Log-Likelihood:   -123.14
No. Observations:      40  AIC:              254.3
Df Residuals:          36  BIC:              261.0
Df Model:              3
Covariance Type:      nonrobust
```

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

```
      coef  std err      t  P>|t|  [0.025  0.975]
```

```
-----
const      0.9070   6.151   0.147   0.884  -11.568   13.382
Study_Hours  7.3220   0.375  19.541   0.000   6.562   8.082
```

Sleep_Hours	1.2685	0.790	1.607	0.117	-0.333	2.870
Coffee_Intake	2.2062	0.781	2.826	0.008	0.623	3.790

=====

=====

Omnibus:	1.984	Durbin-Watson:	1.985
Prob(Omnibus):	0.371	Jarque-Bera (JB):	1.669
Skew:	0.493	Prob(JB):	0.434
Kurtosis:	2.824	Cond. No.	62.9

=====

=====

Notes:

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