

# 📁 직무성과 OLS 회귀 분석 - Full

## Analysis with 성과지표

### 📄 문제

다음은 직원의 경력, 학력, 근무 시간에 따른 직무 성과 점수 데이터이다. 선형 회귀 분석을 통해 각 변수의 영향력을 평가하고, 회귀모델의 예측 성능을 분석하시오.

#### 📁 변수 설명:

- Experience: 경력 (년)
- Education\_Level: 학력 수준 (1: 고졸, 2: 대졸, 3: 대학원졸)
- Hours\_Per\_Week: 주당 평균 근무 시간
- Performance\_Score: 직무 성과 점수 (종속변수)

#### 🧠 문제:

1. Performance\_Score 를 종속변수로 하여 선형회귀 모델을 구성하시오.
2. 각 변수의 회귀계수 및 p-value 를 해석하시오.
3. 결정계수( $R^2$ ), MSE, RMSE, MAE 등의 회귀 모델 성능 지표를 계산하고 해석하시오.

### 📖 데이터셋 (일부)

Experience	Education_Level	Hours_Per_Week	Performance_Score
3.0	2.0	36.7	69.9
9.0	3.0	36.7	87.8
4.0	1.0	42.4	85.3
4.0	1.0	43.1	81.0
1.0	2.0	39.0	77.5
7.0	2.0	45.3	85.1
9.0	2.0	32.5	102.9
6.0	1.0	52.5	105.6
2.0	1.0	33.2	62.3
9.0	2.0	39.0	102.7

6.0	2.0	33.3	83.5
8.0	3.0	33.5	84.8
6.0	1.0	43.0	73.8
5.0	1.0	43.6	87.0
1.0	3.0	48.7	88.8
1.0	1.0	43.2	68.7
6.0	2.0	34.1	94.7
4.0	1.0	36.6	89.1
5.0	3.0	50.2	108.9
10.0	1.0	37.7	111.9

### ✓ 정답 코드

```
import pandas as pd
import statsmodels.api as sm
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

df = pd.read_csv("job_performance_data.csv")
X = df[['Experience', 'Education_Level', 'Hours_Per_Week']]
y = df['Performance_Score']
X = sm.add_constant(X)

model = sm.OLS(y, X).fit()
print(model.summary())

# 예측 및 성능지표
y_pred = model.predict(X)
print("R2:", r2_score(y, y_pred))
print("MSE:", mean_squared_error(y, y_pred))
print("RMSE:", mean_squared_error(y, y_pred, squared=False))
print("MAE:", mean_absolute_error(y, y_pred))
```

### 🔗 해설

- Experience 변수는 계수가 높고 p-value 가 낮아 성과에 긍정적인 영향을 줄 수 있음.
- Hours\_Per\_Week 또한 긍정적인 상관관계를 가지며, Education\_Level 은 상대적으로 영향력이 작을 수 있음.

- - 결정계수( $R^2$ ): 0.724 → 모델이 전체 성과의 약 72%를 설명함.
- - MSE: 74.829, RMSE: 8.65, MAE: 7.085 → 예측 오차의 평균 크기를 나타냄.

## 📖 코드 실행 결과 (model.summary())

### OLS Regression Results

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

```
Dep. Variable:    Performance_Score  R-squared:            0.724
Model:                OLS  Adj. R-squared:        0.692
Method:             Least Squares  F-statistic:         22.71
Date:              Sun, 15 Jun 2025  Prob (F-statistic):    1.96e-07
Time:              09:36:20  Log-Likelihood:        -107.30
No. Observations:      30  AIC:                222.6
Df Residuals:          26  BIC:                228.2
Df Model:              3
Covariance Type:      nonrobust
```

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

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

```
-----
```

```
const          17.1920    13.379     1.285    0.210   -10.308    44.693
Experience       4.3967     0.577     7.618    0.000     3.210     5.583
Education_Level  0.8543     2.238     0.382    0.706    -3.746     5.455
Hours_Per_Week  1.1343     0.305     3.719    0.001     0.507     1.761
```

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

```
Omnibus:          0.964  Durbin-Watson:          2.057
```

Prob(Omnibus):           0.618   Jarque-Bera (JB):           0.945

Skew:               -0.272   Prob(JB):               0.623

Kurtosis:           2.322   Cond. No.               325.

=====

=====

Notes:

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