

```

# 라이브러리 로드
import numpy as np
import pandas as pd
import statsmodels.api as sm
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
from math import sqrt

#구글드라이브 연결
from google.colab import drive
drive.mount('/content/drive')

#데이터 로드 및 확인
data = pd.read_csv("/content/drive/My Drive/D2.야구선수연봉(투수).csv")
print(data.shape)
print(data.info())
print(data.head())

#train/test 데이터셋 분리
train_data, test_data = train_test_split(data, test_size=0.2, random_state=1)

#회귀모델 생성
#모델1
model1 = sm.OLS.from_formula("SALARY_2018 ~ " + " ".join(train_data.columns.difference(['SALARY_2018'])), data=train_data).fit()
print(model1.summary())

#모델2
model2 = sm.OLS.from_formula("SALARY_2018 ~ VICTORY + DEFEAT + SAVE + WAR", data=train_data).fit()
print(model2.summary())

# 예측
pre_lm = model2.predict(test_data)
result = pd.DataFrame({'actual': test_data['SALARY_2018'], 'predict': pre_lm})
print(result)

# RMSE 계산
rmse_value = sqrt(mean_squared_error(test_data['SALARY_2018'], pre_lm))
print("RMSE:", rmse_value)

```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).



```

(152, 22)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 152 entries, 0 to 151
Data columns (total 22 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   PLAYER                152 non-null   object
 1   TEAM                  152 non-null   object
 2   VICTORY                152 non-null   int64
 3   DEFEAT                152 non-null   int64
 4   SAVE                  152 non-null   int64
 5   HOLD                  152 non-null   int64
 6   BLOWN                 152 non-null   int64
 7   GAME                  152 non-null   int64
 8   STARTER               152 non-null   int64
 9   INNING                152 non-null   float64
10  STRIKEOUT9            152 non-null   float64
11  BALLNET9              152 non-null   float64
12  HOMERUN9              152 non-null   float64
13  BABIP                 152 non-null   float64
14  LOB                   152 non-null   float64
15  ERA                   152 non-null   float64
16  RA9_WAR               152 non-null   float64
17  FIP                   152 non-null   float64
18  KFIIP                 152 non-null   float64
19  WAR                   152 non-null   float64
20  SALARY_2018           152 non-null   int64
21  SALARY_2017           152 non-null   int64
dtypes: float64(11), int64(9), object(2)
memory usage: 26.3+ KB
None

```

	PLAYER	TEAM	VICTORY	DEFEAT	SAVE	HOLD	BLOWN	GAME	STARTER	#
0	Hyun-Jong Yang	KIA	20	6	0	0	0	31	31	
1	Casey Kelly	SK	16	7	0	0	0	30	30	
2	So-Sa	LG	11	11	1	0	0	30	29	
3	Woo-ran Jung	HANWHA	6	4	26	0	5	56	0	
4	Brooks Riley	LOTTE	13	7	0	0	0	30	30	

	INNING	...	HOMRUN9	BABIP	LOB	ERA	RA9_WAR	FIP	KFIP	WAR	#
0	193.1	...	0.79	0.332	72.1	3.44	6.54	3.94	3.82	5.64	
1	190.0	...	0.76	0.342	73.7	3.60	6.91	3.69	3.44	6.62	
2	185.1	...	0.53	0.319	67.1	3.88	6.80	3.52	3.41	6.08	
3	59.0	...	0.76	0.304	79.4	2.75	2.85	3.26	2.69	1.81	
4	187.1	...	0.91	0.323	74.1	3.80	6.13	4.36	4.31	4.38	

	SALARY_2018	SALARY_2017
0	230000	150000
1	140000	85000
2	120000	50000
3	120000	120000
4	111000	85000

[5 rows x 22 columns]

OLS Regression Results

Dep. Variable:	SALARY_2018	R-squared:	1.000
Model:	OLS	Adj. R-squared:	nan
Method:	Least Squares	F-statistic:	nan
Date:	Wed, 26 Mar 2025	Prob (F-statistic):	nan
Time:	12:10:21	Log-Likelihood:	2057.6
No. Observations:	121	AIC:	-3873.
Df Residuals:	0	BIC:	-3535.
Df Model:	120		

Covariance Type: nonrobust



	coef	std err	t	P> t	[0.025	0.975]
Intercept	1500.8461	inf	0	nan	nan	nan
PLAYER[T.Brooks Raley]	131.1369	inf	0	nan	nan	nan
PLAYER[T.Byung-Yong Chae]	2333.0598	inf	0	nan	nan	nan
PLAYER[T.Casey Kelly]	6424.3669	inf	0	nan	nan	nan
PLAYER[T.Chan-Heon Jung]	-5030.9110	inf	-0	nan	nan	nan
PLAYER[T.Chang-Min Im]	3582.7009	inf	0	nan	nan	nan
PLAYER[T.Chang-Sik Song]	3015.6537	inf	0	nan	nan	nan
PLAYER[T.Chang-Yong Lim]	-4635.4869	inf	-0	nan	nan	nan
PLAYER[T.Dae-Hyun Kim]	3436.7182	inf	0	nan	nan	nan
PLAYER[T.Dae-Yoo Kim]	-2898.1629	inf	-0	nan	nan	nan
PLAYER[T.Deok-Ju Ham]	-1.159e+04	inf	-0	nan	nan	nan
PLAYER[T.Do-Young Kim]	-3618.3094	inf	-0	nan	nan	nan
PLAYER[T.Dong-Hwan Choi]	3032.4826	inf	0	nan	nan	nan
PLAYER[T.Dong-Hyun Lee]	-2638.3880	inf	-0	nan	nan	nan
PLAYER[T.Dong-Seob Shim]	644.8804	inf	0	nan	nan	nan
PLAYER[T.Dong-Yoon Jeong]	-1876.3065	inf	-0	nan	nan	nan
PLAYER[T.Eun-Bum Song]	-615.8839	inf	-0	nan	nan	nan
PLAYER[T.Geon-Hee Hong]	-1597.6898	inf	-0	nan	nan	nan
PLAYER[T.Geon-Uk Yeo]	-3440.9749	inf	-0	nan	nan	nan
PLAYER[T.Geon-Yeop Heo]	-1235.2219	inf	-0	nan	nan	nan
PLAYER[T.Geum-Kang Choi]	5318.4235	inf	0	nan	nan	nan
PLAYER[T.Geun-Hong Park]	-554.2066	inf	-0	nan	nan	nan
PLAYER[T.Geun-Jong Cho]	8191.3176	inf	0	nan	nan	nan
PLAYER[T.Geun-Young Yoon]	-1208.3481	inf	-0	nan	nan	nan
PLAYER[T.Gyu-Hyun Ahn]	-2947.1434	inf	-0	nan	nan	nan
PLAYER[T.Gyu-Jin Yoon]	-3304.2871	inf	-0	nan	nan	nan
PLAYER[T.Gyun Seo]	799.3423	inf	0	nan	nan	nan
PLAYER[T.Hee-Kwan Yoo]	-1.167e+04	inf	-0	nan	nan	nan
PLAYER[T.Hee-Sang Yoon]	-9758.6840	inf	-0	nan	nan	nan
PLAYER[T.Hee-Su Park]	342.6035	inf	0	nan	nan	nan
PLAYER[T.Hee-Woon Ryu]	9594.2322	inf	0	nan	nan	nan
PLAYER[T.Hyeong-Beom Lee]	1707.1225	inf	0	nan	nan	nan
PLAYER[T.Hyuk Kwon]	1965.7669	inf	0	nan	nan	nan
PLAYER[T.Hyun-Ho Lee]	3421.9022	inf	0	nan	nan	nan
PLAYER[T.Hyun-Jong Yang]	3.793e+04	inf	0	nan	nan	nan



PLAYER[T.Hyun-Jun Im]	2130.9544	inf	0	nan	nan	nan
PLAYER[T.Hyun-Seung Lee]	806.3347	inf	0	nan	nan	nan
PLAYER[T.Jae-Guk Ryu]	-1.682e+04	inf	-0	nan	nan	nan
PLAYER[T.Jae-Hak Lee]	-1171.1092	inf	-0	nan	nan	nan
PLAYER[T.Jae-Hwan Bae]	-4629.2418	inf	-0	nan	nan	nan
PLAYER[T.Jae-Won Jung]	-1064.1437	inf	-0	nan	nan	nan
PLAYER[T.Jae-Woong Shin]	-490.3910	inf	-0	nan	nan	nan
PLAYER[T.Jae-Yong Cha]	6180.2336	inf	0	nan	nan	nan
PLAYER[T.Jae-Young Kim]	-3692.4405	inf	-0	nan	nan	nan
PLAYER[T.Jang-Ho Bae]	3208.8205	inf	0	nan	nan	nan
PLAYER[T.Jang-San Kang]	-1547.7578	inf	-0	nan	nan	nan
PLAYER[T.Jeong-Woo Im]	-6900.6604	inf	-0	nan	nan	nan
PLAYER[T.Ji-Gwang Choi]	2241.6353	inf	0	nan	nan	nan
PLAYER[T.Ji-Hoon Park]	-1259.7791	inf	-0	nan	nan	nan
PLAYER[T.Ji-Yong Kim]	1226.8077	inf	0	nan	nan	nan
PLAYER[T.Jin-Hyung Park]	3808.8333	inf	0	nan	nan	nan
PLAYER[T.Jin-Sung Kim]	-5386.2146	inf	-0	nan	nan	nan
PLAYER[T.Jin-Woo Im]	2123.1709	inf	0	nan	nan	nan
PLAYER[T.Jin-Woo Kim]	-3300.5496	inf	-0	nan	nan	nan
PLAYER[T.Jin-Yong Seo]	-731.0895	inf	-0	nan	nan	nan
PLAYER[T.Jin-Young Kim]	2382.3574	inf	0	nan	nan	nan
PLAYER[T.Jong-Hoon Kim]	-7437.7919	inf	-0	nan	nan	nan
PLAYER[T.Jong-Hyeok Lee]	-1655.7335	inf	-0	nan	nan	nan
PLAYER[T.Jong-Hyun Won]	-739.1341	inf	-0	nan	nan	nan
PLAYER[T.Joo-Han Kim]	2191.1866	inf	0	nan	nan	nan
PLAYER[T.Ju-Young Son]	3202.7827	inf	0	nan	nan	nan
PLAYER[T.Jun-Hyuk Lim]	-1858.6022	inf	-0	nan	nan	nan
PLAYER[T.Jung-Bae Park]	7104.1275	inf	0	nan	nan	nan
PLAYER[T.Jung-Jin Park]	-1.048e+04	inf	-0	nan	nan	nan
PLAYER[T.Jung-Min Lee]	-3992.1122	inf	-0	nan	nan	nan
PLAYER[T.Jung-Rak Shin]	1165.0040	inf	0	nan	nan	nan
PLAYER[T.Kang-Yool Kim]	5438.5197	inf	0	nan	nan	nan
PLAYER[T.Ki-Jun Lim]	2416.8990	inf	0	nan	nan	nan
PLAYER[T.Ki-Young Lim]	-7569.8358	inf	-0	nan	nan	nan
PLAYER[T.Kwang-Eun Moon]	200.9865	inf	0	nan	nan	nan
PLAYER[T.Kwon Ju]	4170.9894	inf	0	nan	nan	nan
PLAYER[T.Kyu-Min Woo]	-5527.8687	inf	-0	nan	nan	nan
PLAYER[T.Kyung-Eun No]	-428.8437	inf	-0	nan	nan	nan
PLAYER[T.Kyung-Tae Park]	-785.8110	inf	-0	nan	nan	nan



PLAYER[T.Young-Hoon Seong]	-4657.6774	inf	-0	nan	nan	nan
PLAYER[T.Young-Il Jeong]	-5451.8805	inf	-0	nan	nan	nan
PLAYER[T.Young-Min Son]	-1757.0711	inf	-0	nan	nan	nan
PLAYER[T.Young-Myung Ahn]	1.303e+04	inf	0	nan	nan	nan
PLAYER[T.Young-Pyo Ko]	-1.774e+04	inf	-0	nan	nan	nan
PLAYER[T.Yu-Su Jeon]	-2099.6516	inf	-0	nan	nan	nan
TEAM[T.HANWHA]	5976.7839	inf	0	nan	nan	nan
TEAM[T.KIA]	3138.7719	inf	0	nan	nan	nan
TEAM[T.KT]	6090.4202	inf	0	nan	nan	nan
TEAM[T.LG]	5090.4072	inf	0	nan	nan	nan
TEAM[T.LOTTE]	2008.9109	inf	0	nan	nan	nan
TEAM[T.NC]	3256.3777	inf	0	nan	nan	nan
TEAM[T.SAMSUNG]	1761.8448	inf	0	nan	nan	nan
TEAM[T.SK]	4233.4990	inf	0	nan	nan	nan
BABIP	-665.5175	inf	-0	nan	nan	nan
BALLNET9	1178.0156	inf	0	nan	nan	nan
BLOWN	346.5720	inf	0	nan	nan	nan
DEFEAT	-60.1621	inf	-0	nan	nan	nan
ERA	78.9977	inf	0	nan	nan	nan
FIP	837.1066	inf	0	nan	nan	nan
GAME	61.3792	inf	0	nan	nan	nan
HOLD	-324.8788	inf	-0	nan	nan	nan
HOMRUN9	2373.7649	inf	0	nan	nan	nan
INNING	-52.9580	inf	-0	nan	nan	nan
LOB	19.2529	inf	0	nan	nan	nan
RA9_WAR	130.8345	inf	0	nan	nan	nan
SALARY_2017	0.9904	inf	0	nan	nan	nan
SAVE	-382.7034	inf	-0	nan	nan	nan
STARTER	-260.3662	inf	-0	nan	nan	nan
STRIKEOUT9	-1312.7768	inf	-0	nan	nan	nan
VICTORY	659.6962	inf	0	nan	nan	nan
WAR	8796.2431	inf	0	nan	nan	nan
kFIP	-1950.8065	inf	-0	nan	nan	nan

Omnibus:	104.631	Durbin-Watson:	1.573
Prob(Omnibus):	0.000	Jarque-Bera (JB):	812.408
Skew:	-3.129	Prob(JB):	3.87e-177
Kurtosis:	14.045	Cond. No.	1.06e+06

Notes:



- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The input rank is higher than the number of observations.
- [3] The condition number is large, 1.06e+06. This might indicate that there are strong multicollinearity or other numerical problems.

OLS Regression Results

Dep. Variable:	SALARY_2018	R-squared:	0.693
Model:	OLS	Adj. R-squared:	0.682
Method:	Least Squares	F-statistic:	65.47
Date:	Wed, 26 Mar 2025	Prob (F-statistic):	7.37e-29
Time:	12:10:21	Log-Likelihood:	-1359.3
No. Observations:	121	AIC:	2729.
Df Residuals:	116	BIC:	2743.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	4820.0776	2543.807	1.895	0.061	-218.252	9858.407
VICTORY	2817.1971	845.880	3.330	0.001	1141.826	4492.568
DEFEAT	-1796.7433	822.673	-2.184	0.031	-3426.150	-167.337
SAVE	1158.2981	336.387	3.443	0.001	492.042	1824.554
WAR	1.45e+04	2239.886	6.472	0.000	1.01e+04	1.89e+04

Omnibus:	46.631	Durbin-Watson:	2.109
Prob(Omnibus):	0.000	Jarque-Bera (JB):	283.046
Skew:	1.104	Prob(JB):	3.45e-62
Kurtosis:	10.160	Cond. No.	11.4

Notes:

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



```

actual      predict
59  10000  39974.619819
33  20000  14498.889582
69   8500  11656.423474
127  3200   3660.275219
31  20000  49535.237615
107   4000  11551.607630
97   5200  20753.344459
14   50000  34522.360453
53  11500  33330.787315
44  15000  27195.257084
146  2800   5979.879921
138  3000   2355.497574
78   7100 -5343.083468
102   4500  2232.154340
5   100000  87538.406815
85   6300  4049.414877
16  50000  7699.066206
100  4700 -4954.818843
89   6000  6993.202690
95  5400  6848.227396
91   6000  123.628397
124  9300  4675.102276
92   5800  8883.508283
99   4700  16051.468513
93   5700 -2449.204364
40  15500  52790.122512
29  23000  23944.153710
42  15000  8103.096332
54  11000  39639.359853
122   3400  5840.531400
35  19000  20492.785338
RMSE: 15760.477097309951
/usr/local/lib/python3.11/dist-packages/statsmodels/regression/linear_model.py:1795: RuntimeWarning: divide by zero encountered in divide
    return 1 - (np.divide(self.nobs - self.k_constant, self.df_resid)
/usr/local/lib/python3.11/dist-packages/statsmodels/regression/linear_model.py:1795: RuntimeWarning: invalid value encountered in scalar multiply
    return 1 - (np.divide(self.nobs - self.k_constant, self.df_resid)

```