

로지스틱 회귀 + 표준화

#01. 작업준비

패키지 참조

그래프 초기화

데이터 가져오기

#02. 데이터 표준화

#03. 로지스틱 회귀 분석 (모듈기능 활용)

로지스틱 회귀 + 표준화

#01. 작업준비

패키지 참조

```
from pandas import read_excel
from matplotlib import pyplot as plt
import seaborn as sb
import sys
import os
```

```
sys.path.append(os.path.dirname(os.path.dirname(os.getcwd())))
from helper import my_logit, scaling
```

그래프 초기화

```
plt.rcParams["font.family"] = 'AppleGothic' if sys.platform == 'darwin'
plt.rcParams["font.size"] = 12
plt.rcParams["figure.figsize"] = (10, 5)
plt.rcParams["axes.unicode_minus"] = False
```

데이터 가져오기

로지스틱 회귀 + 표준화

#01. 작업준비

패키지 참조

그래프 초기화

데이터 가져오기

#02. 데이터 표준화

#03. 로지스틱 회귀 분석 (모듈기능 활용)

```
df = read_excel("https://data.hossam.kr/E05/gradeuate.xlsx")
df.head()
```

	합격여부	필기점수	학부성적	병원경력
0	0	380	3.61	3
1	1	660	3.67	3
2	1	800	4.00	1
3	1	640	3.19	4
4	0	520	2.93	4

#02. 데이터 표준화

```
df_tmp = df.drop('합격여부', axis=1)
std_df = scalling(df_tmp)
std_df['합격여부'] = df['합격여부']
std_df.head()
```

	필기점수	학부성적	병원경력	합격여부
0	-1.800263	0.579072	0.545968	0
1	0.626668	0.736929	0.545968	1
2	1.840134	1.605143	-1.574296	1

로지스틱 회귀 + 표준화

#01. 작업준비

패키지 참조

그래프 초기화

데이터 가져오기

#02. 데이터 표준화

#03. 로지스틱 회귀 분석 (모듈기능 활용)

	필기점수	학부성적	병원경력	합격여부
3	0.453316	-0.525927	1.606100	1
4	-0.586797	-1.209974	1.606100	0

#03. 로지스틱 회귀 분석 (모듈기능 활용)

```
logit_result = my_logit(std_df, y='합격여부', x=['필기점수', '학부성적', '병원경력'])
print(logit_result.summary)
```

Optimization terminated successfully.

Current function value: 0.574302

Iterations 5

Logit Regression Results

Dep. Variable:	합격여부	No. Observations:	0
Model:	Logit	Df Residuals:	-
Method:	MLE	Df Model:	-
Date:	Mon, 31 Jul 2023	Pseudo R-squ.:	0.000
Time:	12:01:23	Log-Likelihood:	-1.088
converged:	True	LL-Null:	-1.088
Covariance Type:	nonrobust	LLR p-value:	8.2

	coef	std err	z	P> z	[0.025
Intercept	-0.8591	0.117	-7.345	0.000	-1.088

로지스틱 회귀 + 표준화

#01. 작업준비

패키지 참조

그래프 초기화

데이터 가져오기

#02. 데이터 표준화

#03. 로지스틱 회귀 분석 (모듈기능 활용)

03-로지스틱회귀+표준화.ipynb

필기점수	0.2647	0.126	2.101	0.036	0.018
학부성적	0.2953	0.124	2.373	0.018	0.051
병원경력	-0.5283	0.120	-4.405	0.000	-0.763

=====

logit_result.cmdf

	Positive	Negative
True	253	20
False	98	29

logit_result.odds_rate_df

	odds_rate
Intercept	0.423557
필기점수	1.302986
학부성적	1.343577
병원경력	0.589627

logit_result.prs

로지스틱 회귀 + 표준화

#01. 작업준비

패키지 참조

그래프 초기화

데이터 가져오기

#02. 데이터 표준화

#03. 로지스틱 회귀 분석 (모듈기능 활용)

0.08107331586891475

logit_result.result_df

	설명력 (Pseudo-Rsqe)	정확도 (Accuracy)	정밀도 (Precision)	재현율 (Recall, TPR)	위양성 율 (Fallout, FPR)	특이성 (Specificity, TNR)	RAS
0	0.081073	0.705	0.591837	0.228346	0.07326	0.92674	0.577543