

연습문제 폴이

패키지 가져오기

문제 1

데이터 가져오기

회귀분석 수행

결과보고

문제 2

결과보고

모델 구축

문제3

모델구축

계수를 리턴받기

연습문제 폴이

패키지 가져오기

```
from pandas import read_excel
import os, sys

sys.path.append(os.path.dirname(os.path.dirname(os.getcwd()))))
from helper import ext_ols
```

문제 1

데이터 가져오기

```
df = read_excel("https://data.hossam.kr/E04/beans.xlsx")
df.head()
```

	수분함유량	강도
0	36	29
1	40	32
2	34	29

연습문제 풀이

패키지 가져오기

문제 1

데이터 가져오기

회귀분석 수행

결과보고

문제 2

결과보고

모델 구축

문제3

모델구축

계수를 리턴받기

	수분함유량	강도
3	44	40
4	33	31

회귀분석 수행

```
model, fit, summary, table, result, goodness, varstr = ext_ols(df, x="수", y="강도")
summary
```

```
c:\Users\leekh\AppData\Local\Programs\Python\Python311\Lib\site-packages
warnings.warn("kurtosistest only valid for n ≥ 20 ... continuing ")
```

OLS Regression Results

Dep. Variable:	강도	R-squared:	0.789
Model:	OLS	Adj. R-squared:	0.762
Method:	Least Squares	F-statistic:	29.85
Date:	Tue, 25 Jul 2023	Prob (F-statistic):	0.000599
Time:	14:35:36	Log-Likelihood:	-19.614
No. Observations:	10	AIC:	43.23
Df Residuals:	8	BIC:	43.83

연습문제 폴이

패키지 가져오기

문제 1

데이터 가져오기

회귀분석 수행

결과보고

문제 2

결과보고

모델 구축

문제3

모델구축

계수를 리턴받기

Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	5.4991	4.707	1.168	0.276	-5.355	16.353
수분함유량	0.7123	0.130	5.463	0.001	0.412	1.013

Omnibus:	0.502	Durbin-Watson:	1.420
Prob(Omnibus):	0.778	Jarque-Bera (JB):	0.530
Skew:	0.264	Prob(JB):	0.767
Kurtosis:	2.004	Cond. No.	280.

Notes:

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

결과보고

table

연습문제 풀이

패키지 가져오기

문제 1

데이터 가져오기

회귀분석 수행

결과보고

문제 2

결과보고

모델 구축

문제3

모델구축

계수를 리턴받기

		B	표준오차	β	t	유의확률	VIF
종속변수	독립변수						
강도	수분함유량	0.7123	0.130	0	5.463*	0.001	1.0

result

'R(0.789), R^2(0.762), F(29.85), 유의확률(0.000599), Durbin-Watson(1.420)

goodness

'강도에 대하여 수분함유량로 예측하는 회귀분석을 실시한 결과, 이 회귀모형은 통계적으로

varstr

['수분함유량의 회귀계수는 0.7123(p<0.05)로, 강도에 대하여 유의미한 예측변인인 것으

문제 2

연습문제 풀이

패키지 가져오기

문제 1

데이터 가져오기

회귀분석 수행

결과보고

문제 2

결과보고

모델 구축

문제3

모델구축

계수를 리턴받기

```
df = read_excel("https://data.hossam.kr/E04/aircon.xlsx")
df.head()
```

	예약대수	판매대수
0	19	33
1	23	51
2	26	40
3	29	49
4	30	50

```
model, fit, summary, table, result, goodness, varstr = ext_ols(df, x="예약대수", y="판매대수")
summary
```

```
c:\Users\leekh\AppData\Local\Programs\Python\Python311\Lib\site-packages
warnings.warn("kurtosistest only valid for n ≥ 20 ... continuing ")
```

OLS Regression Results

Dep. Variable:	판매대수	R-squared:	0.897
Model:	OLS	Adj. R-squared:	0.884
Method:	Least Squares	F-statistic:	69.40

연습문제 풀이

패키지 가져오기

문제 1

데이터 가져오기

회귀분석 수행

결과보고

문제 2

결과보고

모델 구축

문제3

모델구축

계수를 리턴받기

Date:	Tue, 25 Jul 2023	Prob (F-statistic):	3.26e-05
Time:	14:35:36	Log-Likelihood:	-32.667
No. Observations:	10	AIC:	69.33
Df Residuals:	8	BIC:	69.94
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	5.0689	7.113	0.713	0.496	-11.333	21.471
예약대수	1.5751	0.189	8.331	0.000	1.139	2.011

Omnibus:	1.408	Durbin-Watson:	2.910
Prob(Omnibus):	0.495	Jarque-Bera (JB):	0.497
Skew:	-0.541	Prob(JB):	0.780
Kurtosis:	2.849	Cond. No.	119.

Notes:

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

결과보고

연습문제 풀이

패키지 가져오기

문제 1

데이터 가져오기

회귀분석 수행

결과보고

문제 2

결과보고

모델 구축

문제3

모델구축

계수를 리턴받기

table

		B	표준오차	β	t	유의확률	VIF
종속변수	독립변수						
판매대수	예약대수	1.5751	0.189	0	8.331*	0.000	1.0

result

'R(0.897), R^2(0.884), F(69.40), 유의확률(3.26e-05), Durbin-Watson(2.910)

goodness

'판매대수에 대하여 예약대수로 예측하는 회귀분석을 실시한 결과, 이 회귀모형은 통계적으로

varstr

['예약대수의 회귀계수는 1.5751(p<0.05)로, 판매대수에 대하여 유의미한 예측변인인 것

연습문제 풀이

패키지 가져오기

문제 1

데이터 가져오기

회귀분석 수행

결과보고

문제 2

결과보고

모델 구축

문제3

모델구축

계수를 리턴받기

모델 구축

예약대수 = 150
 $5.0689 + 1.5751 * \text{예약대수}$

241.3339

```
fit.predict({"예약대수": [150]})
```

```
0    241.334067
dtype: float64
```

문제3

```
df = read_excel("https://data.hossam.kr/E04/galton.xlsx")
df.head()
```

	child	parent
0	61.7	70.5
1	61.7	68.5
2	61.7	65.5

연습문제 풀이

패키지 가져오기

문제 1

데이터 가져오기

회귀분석 수행

결과보고

문제 2

결과보고

모델 구축

문제3

모델구축

계수를 리턴받기

	child	parent
3	61.7	64.5
4	61.7	64.0

```
model, fit, summary, table, result, goodness, varstr = ext_ols(df, x="pa
summary
```

OLS Regression Results

Dep. Variable:	child	R-squared:	0.210
Model:	OLS	Adj. R-squared:	0.210
Method:	Least Squares	F-statistic:	246.8
Date:	Tue, 25 Jul 2023	Prob (F-statistic):	1.73e-49
Time:	14:35:36	Log-Likelihood:	-2063.6
No. Observations:	928	AIC:	4131.
Df Residuals:	926	BIC:	4141.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	23.9415	2.811	8.517	0.000	18.425	29.458

연습문제 폴이

패키지 가져오기

문제 1

데이터 가져오기

회귀분석 수행

결과보고

문제 2

결과보고

모델 구축

문제3

모델구축

계수를 리턴받기

parent	0.6463	0.041	15.711	0.000	0.566	0.727
---------------	--------	-------	--------	-------	-------	-------

Omnibus:	11.057	Durbin-Watson:	0.046
Prob(Omnibus):	0.004	Jarque-Bera (JB):	10.944
Skew:	-0.241	Prob(JB):	0.00420
Kurtosis:	2.775	Cond. No.	2.61e+03

Notes:

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

[2] The condition number is large, 2.61e+03. This might indicate that there are strong multicollinearity or other numerical problems.

table

		B	표준오차	β	t	유의확률	VIF
종속변수	독립변수						
child	parent	0.6463	0.041	0	15.711*	0.000	1.0

result

연습문제 풀이

패키지 가져오기

문제 1

데이터 가져오기

회귀분석 수행

결과보고

문제 2

결과보고

모델 구축

문제3

모델구축

계수를 리턴받기

```
'R(0.210), R^2(0.210), F(246.8), 유의확률(1.73e-49), Durbin-Watson(0.046)
```

```
goodness
```

```
'child에 대하여 parent로 예측하는 회귀분석을 실시한 결과, 이 회귀모형은 통계적으로
```

```
varstr
```

```
['parent의 회귀계수는 0.6463(p<0.05)로, child에 대하여 유의미한 예측변인인 것으
```

모델구축

```
parent = 68.8976
23.9415 + 0.6463 * parent
```

```
68.47001888
```

```
fit.predict({"parent": [68.8976]})
```

연습문제 풀이

패키지 가져오기

문제 1

데이터 가져오기

회귀분석 수행

결과보고

문제 2

결과보고

모델 구축

문제3

모델구축

계수를 리턴받기

```
0      68.4694  
dtype: float64
```

계수를 리턴받기

```
b, a = fit.params  
b + a * 68.8976
```

```
68.469400182378
```