



ÇOX XƏTTİ REQRESSIYA MODELİ ÜÇÜN NÜMUNƏ PYTHON TƏTBİQİ

Mənzil qiymətlərinin müəyyənədiciləri haqqında tətbiq

Ev qiymətlərinin müəyyənədiciləri üçün bu nümunə proqnozda aşağıdakı Excel cədvəlində təmsil olunan məlumatlardan istifadə edəcəyik.

n	y	x1	x2	x3
1	450000	155	4	3
2	370000	135	7	6
3	275000	105	2	9
4	410000	115	7	4
5	310000	110	4	10
6	355000	130	1	8
7	205000	115	2	16
8	200000	120	1	18
9	490000	160	7	2
10	290000	108	5	16
11	230000	115	1	11
12	370000	141	3	5
13	420000	150	3	7
14	280000	110	3	5
15	250000	110	4	13
16	310000	110	4	9
17	385000	127	7	12
18	380000	140	5	16
19	435000	135	4	2
20	335000	120	1	4

Burada qiymətləndirəcəyimiz çox xətlı regressiya modeli aşağıdakı kimidir:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u_i$$

Bərabərlikdə,

y = Evin satış qiyməti,

x_1 = Evin ölçüsü (m^2),

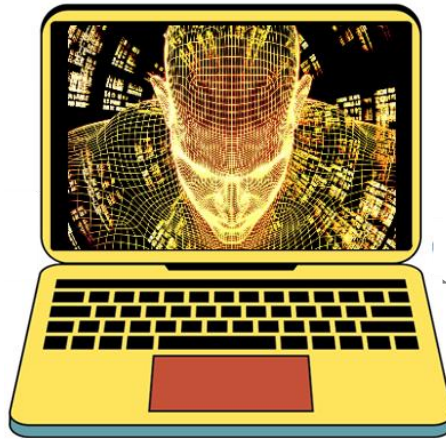
x_2 = Evin yerləşdiyi mərtəbə,

x_3 = Yaşayış yerinin yerləşdiyi binanın yaşını əks etdirir.

İndi əgər istəyirsinizsə, gəlin ev satış qiymətləri ilə bağlı model proqnozumuzu cədvəldəki məlumatlar əsasında Python proqramının köməyi ilə edək.

Regressiya Modelinin Qiymətləndirilməsi

Məlumat ötürüldükdən sonra model proqnozumuzu verə bilərik. Əmsalları öyrəndikdə görürük ki, yaşayış sahəsinin ölçüsü ilə yerləşdiyi mərtəbə ilə yaşayış sahəsinin satış qiyməti, yerləşdiyi binanın yaşı və yaşayış sahəsinin satış qiyməti arasında tərs asılılıqla düzgün əlaqə mövcuddur. Ancaq burada proqnozlaşdırılan modelin statistik cəhətdən əhəmiyyətli olub olmadığını görmək qalır. Eynilə, bu nəticələr hələ modelin təsviri qabiliyyətinə dair bir fikir vermir. Bütün bunları görmək üçün "summary" əmri istifadə olunur.



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

# Verilmiş verilənlərlə dataframe yaradın
data = {
    'n': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20],
    'y': [450000, 370000, 275000, 410000, 310000, 355000, 205000, 200000, 490000, 290000, 230000, 370000, 420000, 280000, 250000,
    'x1': [155, 135, 105, 115, 110, 130, 115, 120, 160, 108, 115, 141, 150, 110, 110, 110, 127, 140, 135, 120],
    'x2': [4, 7, 2, 7, 4, 1, 2, 1, 7, 5, 1, 3, 3, 3, 4, 4, 7, 5, 4, 1],
    'x3': [3, 6, 9, 4, 10, 8, 16, 18, 2, 16, 11, 5, 7, 5, 13, 9, 12, 16, 2, 4]
}
df = pd.DataFrame(data)

# Xətti regressiya modelini uzlaşması
X = df[['x1', 'x2', 'x3']]
X = sm.add_constant(X) # Modelə sabit termin əlavə edin
y = df['y']
model = sm.OLS(y, X).fit()

# Model xülasəsini çap edin
print(model.summary())
```

MÜƏLLİF: İBRAHİM İSMAYİL

OLS Regression Results						
=====						
Dep. Variable:	y	R-squared:	0.888			
Model:	OLS	Adj. R-squared:	0.867			
Method:	Least Squares	F-statistic:	42.17			
Date:	Sun, 24 Dec 2023	Prob (F-statistic):	7.99e-08			
Time:	20:06:28	Log-Likelihood:	-232.51			
No. Observations:	20	AIC:	473.0			
Df Residuals:	16	BIC:	477.0			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	3.133e+04	6.54e+04	0.479	0.638	-1.07e+05	1.7e+05
x1	2479.4759	472.272	5.250	0.000	1478.305	3480.647
x2	1.412e+04	3465.535	4.074	0.001	6771.510	2.15e+04
x3	-6598.7583	1540.316	-4.284	0.001	-9864.083	-3333.434
=====						
Omnibus:	3.049	Durbin-Watson:	1.755			
Prob(Omnibus):	0.218	Jarque-Bera (JB):	1.507			
Skew:	-0.336	Prob(JB):	0.471			
Kurtosis:	1.835	Cond. No.	1.23e+03			
=====						

Notes:

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

[2] The condition number is large, 1.23e+03. This might indicate that there are strong multicollinearity or other numerical problems.

Burada proqnozlaşdırılan parametrlərin t-statistik dəyərlərini cədvəlin kritik dəyərləri ilə müqayisə edərək əmsalların fərdi əhəmiyyətini yoxlaya bilərik (cədvəl kritik dəyəri 5% ilə ikitərəfli test üçün 2.120 olaraq müəyyən edilir) əhəmiyyət səviyyəsi və 16 sərbəstlik dərəcəsi) və ya ehtimal dəyərlərinə baxaraq. Ehtimal dəyərləri 0.5-dən az olarsa, sıfır fərziyyə rədd edilir və əmsalların statistik əhəmiyyətli olduğu qənaətinə gəlinir. Buna əsasən,

$$t_{x1} = 5.250 | 5.250 | > | 2.120 | \rightarrow H_0 \text{ rədd edilir, əmsal statistik əhəmiyyətlidir}$$

$$t_{x2} = 4.074 | 4.074 | > | 2.120 | \rightarrow H_0 \text{ rədd edilir, əmsal statistik əhəmiyyətlidir}$$

$$t_{x3} = -4.284 | -4.284 | > | 2.120 | \rightarrow H_0 \text{ rədd edilir, əmsal statistik əhəmiyyətlidir}$$

və ya

$$x1[Pr(> |t|) = 0.00008 < 0.05 \rightarrow H_0 \text{ rədd edilir, əmsal statistik əhəmiyyətlidir}]$$

$$x2[Pr(> |t|) = 0.00088 < 0.05 \rightarrow H_0 \text{ rədd edilir, əmsal statistik əhəmiyyətlidir}]$$

$$x3[Pr(> |t|) = 0.00057 < 0.05 \rightarrow H_0 \text{ rədd edilir, əmsal statistik əhəmiyyətlidir}]$$

MÜƏLLİF: İBRAHİM İSMAYİL

Nəticələr bu kontekstdə araşdırıldığında 5% əhəmiyyətlik səviyyəsində β_1 , β_2 və β_3 parametrlərinin statistik əhəmiyyətli olduğu görülür. Burada təxmin etdiyimiz model xətti model olduğu üçün proqnoz edilən parametrlər xətti formada şərh olunacaq. Buna əsasən,

- Evin ölçüsü 1 vahid (m^2) artdıqda, evin satış qiyməti təxminən 2479 manat artır.
- Evin yerləşdiyi mərtəbə 1 ədəd artdıqda evin satış qiyməti təxminən 14118 manat artır.
- Evin yerləşdiyi binanın yaşı 1 ədəd artdıqda evin satış qiyməti təxminən 6598 manat azalır (10% əhəmiyyətlik səviyyəsində).

Digər tərəfdən, F test statistikasına baxdığımızda ehtimal dəyərinin 0.006312 olduğunu görürük. $0.000079 < 0.05$ olduğundan modelin bütövlükdə 5% əhəmiyyətlik səviyyəsində əhəmiyyətli olduğunu deyə bilərik.

Həmçinin görürük ki, modeldə çoxlu determinasiya əmsalı (R^2) və korreksiya edilmiş determinasiya əmsalı (düzəliş R^2) müvafiq olaraq 0.8877 və 0.8667-dir. Bu əmsallar təxmin edilən modelin izahedici gücünü göstərir. Məsələn, determinasiya əmsalına görə, ev satış qiymətlərindəki dəyişikliyin təxminən 87% – i modeldə bu üç müstəqil dəyişəndəki dəyişikliklərlə izah olunur.