

N

جلسه بيستم – (بخش اول) بررسي برخي از عناصر جدول مدل رگرسيون

Logistic regression summary

Logit Regression Results

168	No. Observations:	Admitted	Dep. Variable:
166	Df Residuals:	Logit	Model:
1	Df Model:	MLE	Method:
0.7992	Pseudo R-squ.:	Thu, 17 May 2018	Date:
-23.145	Log-Likelihood:	15:51:29	Time:
-115.26	LL-Null:	True	converged:
5.805e-42	LLR p-value:		

 coef
 std err
 z
 P>[z]
 [0.025
 0.975]

 const
 -69.9128
 15.737
 -4.443
 0.000
 -100.756
 -39.070

 SAT
 0.0420
 0.009
 4.454
 0.000
 0.024
 0.060

مدرس: محمد فزوني عضو هيئت علمي دانشگاه گنبدکاووس

Some machine learning terms

$$\Theta = (\beta_0, \dots, \beta_n), x = (1, x_1, \dots, x_n)$$

$$h_{\Theta}(x) = \frac{e^{\Theta x}}{1 + e^{\Theta x}}$$
, Hypothesis function

$$J_{\Theta}(x) = \frac{1}{n} \sum_{i=0}^{n} \left(\frac{e^{\Theta x_i}}{1 + e^{\Theta x_i}} - y_i \right)^2$$
, Cost function

Data Science Course, By Dr. Mohammad Fozouni.



Data Science Course, By Dr. Mohammad Fozouni.





Logistic regression summary



-		
-		



Logit Regression Results

0.0420

0.009

108	-	servations	NO. OI	amitted	-	anable:	Dep. v
166		f Residuals:	D	Logit		Model:	
1	:	Df Model:		MLE		Method:	N
7992	: 0.	udo R-squ.:	Pse	ay 2018	Thu, 17 M	Date:	
3.145	: -23	-Likelihood	Log	5:51:29	1	Time:	
15.26	l: -11	LL-Null:		True		verged:	con
5e-42	5.805	LR p-value:	L				
1	0.975]	[0.025	P> z	z	ef std err	coe	
0	-39.070	-100,756	0.000	-4,443	8 15,737	-69,9128	const

4.454

0.000

0.024

0.060

N

100



Data Science Course, By Dr. Mohammad Fozouni.

SAT

Likelihood function: a function which measures the goodness of fit of a statistical model.

The bigger the likelihood function, the higher the probability that our model is correct

MLE (Maximum Likelihood Estimation) tries to maximize the likelihood function.

The value of the log likelihood is almost but not always negative

The bigger the log likelihood function, the higher the probability that our model is correct

Data Science Course, By Dr. Mohammad Fozouni.

```
reg_log0 = sm.Logit(y, x0)
results log0 = reg log0.fit()
noculte logo cummony()
 Signature:
 reg log0.fit(
     start_params=None,
     method='newton',
     maxiter=35,
     full_output=1,
     disp=1,
```

Using Newton's Method, we try to maximize the Likelihood function

Logit Regr	ession Res	ults					
Dep. Vari	able:	Admitted No. Observations:		16	68		
M	odel:		Logit	Df F	Residuals:	10	65
Me	Method: M				Df Model:		2
	Date: Tue	e, 22 May	2018	Pseud	do R-squ.:	0.824	19
1	Time:	11:	14:24 🔇	Log-L	ikelihood:	-20.18	30
conve	rged:		True		LL-Null:	-115.2	26
				LL	R p-value:	5.118e-4	42
	coef	std err	z	P> z	[0.025	0.975]	
const	-68.3489	16.454	-4.154	0.000	-100.598	-36.100	
SAT	0.0406	0.010	4.129	0.000	0.021	0.060	
Gender	1.9449	0.846	2.299	0.022	0.287	3.603	

Logit Regression Results Admitted No. Observations: Dep. Variable: 168 Model: 166 Logit Df Residuals: MLE Df Model: Method: 1 Date: Tue, 22 May 2018 Pseudo R-squ.: 0.1659 14:03:10 Log-Likelihood: -96.140 Time: converged: True LL-Null: -115.26LLR p-value: 6.283e-10 coef std err z P>|z| [0.025 0.975] const -0.6436 0.222 -2.9010.004 -1.078 -0.209 2.0786 0.363 5.727 0.000 1.367 2.790 Gender

Data Science Course, By Dr. Mohammad Fozouni.

Regression summary

Logit Regression Results

Dep. Va	riable:	A	dmitted	No. Ob	servations	:	168
1	Model:	Logit		D	f Residuals	:	166
M	lethod:		MLE		Df Model	:	1
	Date: T	'hu, 17 Ma	ay 2018	Pse	udo R-squ.	: 0.7	992
	Time:	1	5:51:29	Log	Likelihood	-23	145
conv	erged:		True		LL-Null	-11	5.26
				L	LR p-value	5.805	e-42
	coef	std err	z	P> z	[0.025	0.975]	
const	-69.9128	15.737	-4.443	0.000	-100.756	-39.070	
SAT	0.0420	0.009	4.454	0.000	0.024	0.060	

LL-Null (log likelihood-null):

the log-likelihood of a model which has no independent variables.

 $y = \beta_0 * 1$

Data Science Course, By Dr. Mohammad Fozouni.



You may want to compare the log-likelihood of your model with the LL-null, to see if your model has any explanatory power

summary



The Likelihood-Ratio test (sometimes called the likelihood-ratio chi-squared test) is a hypothesis test that helps you choose the "best" model between two nested models. "Nested models" means that one is a special case of the other. For example, you might want to find out which of the following models is the best fit:

•Model One has four predictor variables (height, weight, age, sex),

•Model Two has two predictor variables (age,sex). It is "nested" within model one because it has just two of the predictor variables (age, sex).

Regression summary

Logit Regression Results



Pseudo R-squared:

McFadden's R-squared

'A good Pseudo R-squared is somewhere between 0.2 and 0.4.

This measure is mostly useful for comparing variations of the same model.

Different models will have completely different and incomparable Pseudo R-squares!

2

How we can interpret the coefficients?

Data Science Course, By Dr. Mohammad Fozouni.

Logit model

SAT - Admittance regression

$$\log\left(\frac{\pi}{1-\pi}\right) = -69.91 + 0.042 * SAT$$

π -> probability of an event occuring 1 - π -> probability of the event NOT occuring

$$log (odds_2) = -69.91 + 0.042 * SAT_2$$

$$log (odds_1) = -69.91 + 0.042 * SAT_1$$

$$log (odds_2) - log (odds_1) = 0.042(SAT_2 - SAT_1)$$

$$\log\left(\frac{odds_2}{odds_1}\right) = 0.042(SAT_2 - SAT_1)$$

Data Science Course, By Dr. Mohammad Fozouni.



Difference of 10 units of SAT:

$$\log\left(\frac{odds_2}{odds_1}\right) = 0.042 * 10$$

 $odds_2 = 152\% * odds_1$

Difference of 100 units of SAT:

 $odds_2 = 6669\% * odds_1$

Data Science Course, By Dr. Mohammad Fozouni.

Time for Jupyter notebook environment coding

