## PERFORMING Multiple Logistic Regression with SigmaPlot 15

This article is made by Jan Smit, Smit Consult



22 November 2022

Data source: Logistic Regression Data in statistics samples.jnb

In general the data contains one depending variable, which contains the values 0 and 1. Zero, 0, for not having the occurrence, and one, 1, for having the occurrence.

1. Load a datafile

A datafile looks like :

📄 Data 1					
	- DISEAS	2-AGE			
1	0,0000	20,0000			
2	0,0000	23,0000			
3	0,0000	24,0000			
4	0,0000	25,0000			
5	1,0000	25,0000			
6	0,0000	26,0000			
7	0,0000	26,0000			
8	0,0000	28,0000			
9	0,0000	28,0000			
10	0,0000	29,0000			
11	0,0000	30,0000			
12	0,0000	30,0000			
13	0,0000	30,0000			
14	0,0000	30,0000			
15	0,0000	30,0000			
16	1,0000	30,0000			
17	0,0000	32,0000			
18	0,0000	32,0000			
19	0,0000	33,0000			
20	0,0000	33,0000			

1. Perform the analyses : depending is disease (values of 0 and 1), independing is AGE

Multiple Lo	gistic	Regressio	on - Select Data	×
Select dat	a by cli me 55.73 57.88 60.02 62.43 55 19	2-0xyg 78. 75. 76. 73. 82	ksheet columns. Select the independent variable column(s) then dick Finish.	Data for Independent (x): 2-AGE Selected Columns Dependent (y): 1-DISEASE Independent (x): 2-AGE Independent (x):
Help		Cancel	Back	Next Finish

2. Inspect the report

The report shows :

Logit P = -5,309 + (0,111 \* AGE)

N = 100

Estimation Criterion: Maximum likelihood Dependent Variable: DISEASE Positive response (1): 1 Reference response (0): 0 Number of unique independent variable combinations: 43

**Pearson Chi-square Statistic:** 101,943 (P = 0,346) **Likelihood Ratio Test Statistic:** 29,310 (P = <0,001) -**2\*Log(Likelihood) =** 107,353 **Hosmer-Lemeshow Statistic:** 1,160 (P = 0,997)

## **Threshold probability for positive classification:** 0,500 **Classification Table:**

	Predicted Reference	<b>Predicted Positive</b>	Totals
Actual Reference Responses	s 45	12	57
Actual Positive Responses	14	29	43
Totals	59	41	100

## **Details of the Logistic Regression Equation**

Ind. Variable	Coefficient	Standard Error	Wald Statistic	P value	VIF
Constant	-5,309	1,134	21,935	<0,001	
AGE	0,111	0,0241	21,254	<0,001	1,000

Ind. Variable	Odds Ratio	5% Conf. Lower	95% Conf. Upper
Constant	0,00494	0,000536	0,0456
AGE	1,117	1,066	1,171

3. We are only interested in the following part of the report:

In order to generate the graphs, and the CI of the model, we use both the model Coefficient and the corresponding Standard Error

Variable	Coefficient	Standard Error
Constant	-5,309	1,134
AGE	0,111	0,0241

4. We copy this in the data sheet :

8	9-Log of Odds	10	11	12	13	1
56,4706	3,0890		Variable	Coefficient	Standard Error	
46,6386	2,7560		Constant	-5,3090	1,1340	
43,7575	2,6450		AGE	0,1110	0,0241	
41,0545	2,5340					
41,0545	2,5340					

5. We select the cells in Column 11-13, select these, and use from Worksheet -> Titles "Promote" first row to titles.

Column and Row Titles	$\times$
Column Row	
columns 11-13	
Title	
Ind. Variable < Prev < > Next >	
Promote row 1 to titles Promote	
Delete promoted row	
OK Annuleren Toepassen Help	

The result is :

11-Ind. Variable	12-Coefficient	13-Standard Erro
Constant	-5,3090	1,1340
AGE	0,1110	0,0241

6. Now make use of the logistic regression function :

Using the logistic model, the probability of a binary event is

$$\Pr(Y = 1|X) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)} = \frac{1}{1 + \exp(-\beta_0 - \beta_1 X)}$$

×I

Here we use the following commands in

Col(1)= Disease Col(2) = AGE

Col(3) = 1/(1+exp(-(cell(12;1)+cell(12;2)\*col(2)))) Quick Transform

quation	Transforme	•
✓ Updating	Transforms	
🗹 Auto On	New Equation	
🗹 Auto On	col(3)=1/(1+exp(-(cell(10;1)+cell(10;2)*col(2	X
Auto On	col(4)=1/(1+exp(-(cell(10;1)+(cell(10;2)+nor	
Auto On	col(5)=1/(1+exp(-(cell(10;1)+(cell(10;2)-nor	
🖌 Auto On	col(9)=col(6)/col(7)	¥
ol(3) = 1/(1+e	exp(-(cell(12; 1)+cell(12; 2)*col(2))))	

7. This will result in one line, with an S-curve. This uses only the Coefficients b0=-5,30 and b1=0,11







A (1 -  $\alpha$ )% two-sided confidence interval for  $\beta_1$  is

$$\hat{\beta}_1 \pm z_{1-\frac{\alpha}{2}} s_{\hat{\beta}_1}$$

We use the following formulas :

For "higher 95% CI"

Col(4) = 1/(1+exp(-(cell(12;1)+(cell(12;2)+norminv(0,975;0;1)\*cell(12;2))\*col(2))))

And for "Lower 95% CI" Col(5) = 1/(1+exp(-(cell(12;1)+(cell(12;2) - norminv(0,975;0;1)\*cell(12;2))\*col(2))))

Quick Transform		×
	Enable auto update timer for	Quick Transforms
Equation		
Updating	Transforms	^
Auto On	col(10)=col(7)/100	
Auto On	col(10)=col(8)/100	X
Auto On	col(3)=1/(1+exp(-(cell(12;1)+cell(12;2)*col	(2
Auto On	col(4)=1/(1+exp(-(cell(12;1)+(cell(12;2)+r	nor
🗹 Auto On	col(5)=1/(1+exp(-(cell(12;1)+(cell(12;2)-n	or 🗸
col(5) =	1/(1+exp(-(cell(12;1)+(cell(12;2)- norminv(0,975;0;1	l)*cell(13;2
<< Less	Help Options Run	Cancel

9. We place this in the data file, after we have given the correct column labels :

Data 1*						
	- DISEASI	2-AGE	3-logistic regression model	4-higher 95% CI	5-Lower 95% CI	
1	0,0000	20,0000	0,0436	0,1049	0,0174	
2	0,0000	23,0000	0,0597	0,1585	0,0210	
3	0,0000	24,0000	0,0663	0,1807	0,0223	
4	0,0000	25,0000	0,0735	0,2054	0,0238	
5	1,0000	25,0000	0,0735	0,2054	0,0238	
6	0,0000	26,0000	0,0814	0,2324	0,0253	
7	0,0000	26,0000	0,0814	0,2324	0,0253	
8	0,0000	28,0000	0,0997	0,2935	0,0286	
9	0,0000	28,0000	0,0997	0,2935	0,0286	
10	0,0000	29,0000	0,1101	0,3274	0,0305	

Let's plot the data:

We use the option "Multiple Line Graphs", and "One X many Y".

Date  I*    1  1.Time    1  55.7300    2  57.8800    3  60.0200    4  62.4300	2-0 xy⊆ 78.00 75.00 76.00 73.00	Select the column to plot by clicking the column in the worksheet.	Data for Y 4    5-Lower 95% Cl    Selected columns    X:  2-AGE    Y 1:  3-logistic regression    Y 2:  4-higher 95% Cl    Y 3:  5-Lower 95% Cl    Y 4:
Help	Annuler	en < Vorige	Volgende > Voltooien

Graph of Logistic regression Disease by Age

