

Chapter 28

Multinomial Regression for Outcome Categories (55 Patients)

General Purpose

To assess whether multinomial regression can be trained to make predictions about (1) patients being in a category and (2) the probability of it.

Specific Scientific Question

Patients from different hospital departments and ages are assessed for falling out of bed (0=no, 1=yes without injury, 2=yes with injury). The falloutofbed categories are the outcome, the department and ages are the predictors. Can a data file of such patients be trained to make predictions in future patients about their best fit category and probability of being in it.

department	falloutofbed	age(years)
,00	1	56,00
,00	1	58,00
,00	1	87,00
,00	1	64,00
,00	1	65,00
,00	1	53,00
,00	1	87,00
,00	1	77,00
,00	1	78,00
,00	1	89,00

This chapter was previously published in “Machine learning in medicine-cookbook 2” as Chap. 4, 2014.

Only the first 10 patients are given, the entire data file is entitled “categoriasa-soutcome” and is in extras.springer.com.

The Computer Teaches Itself to Make Predictions

SPSS versions 18 and later can be used. SPSS will produce an XML (eXtended Markup Language) file of the prediction model from the above data. We will start by opening the above data file.

Command:

click Transform....click Random Number Generators....click Set Starting Point.... click Fixed Value (2000000)....click OK....click Analyze.... Regression Multinomial Logistic Regression....Dependent: falloutofbed.... Factor: department....Covariate: age....click Save....mark: Estimated response probability, Predicted category, Predicted category probability, Actual category probability.... click Browse....various folders in your personal computer come up....in "File name" of the appropriate folder enter "exportcategoriasasoutcome"....click Save....click Continue....click OK.

Parameter estimates							95 % Confidence interval for Exp (B)	
Fall with/out injury ^a	B	Std. error	Wald	df	Sig.	Exp(B)	Lower bound	Upper bound
0 Intercept	5,337	2,298	5,393	1	,020			
Age	-,059	,029	4,013	1	,045	,943	,890	,999
[department = ,00]	-1,139	,949	1,440	1	,230	,320	,050	2,057
[department = 1,00]	0 ^b			0				
1 Intercept	3,493	2,333	2,241	1	,134			
Age	-,022	,029	,560	1	,454	,978	,924	1,036
[department = ,00]	-1,945	,894	4,735	1	,030	,143	,025	,824
[department = 1,00]	0 ^b			0				

^aThe reference category is: 2

^bThis parameter is set to zsf0 because it is redundant

The above table is in the output. The independent predictors of falloutofbed are given. Per year of age there are 0,943 less “no falloutofbeds” versus “falloutofbeds with injury”. The department 0,00 has 0,143 less falloutofbeds with versus without injury. The respective p-values are 0,045 and 0,030. When returning to the main data view, we will observe that SPSS has provided 6 novel variables for each patient.

1. EST1_1 estimated response probability (probability of the category 0 for each patient)
2. EST2_1 idem for category 1

3. EST3_1 idem for category 2
4. PRE_1 predicted category (category with highest probability score)
5. PCP_1 predicted category probability (the highest probability score predicted by model)
6. ACP_1 actual category probability (the highest probability computed from data)

With the Scoring Wizard and the exported XML file entitled "exportcategoriesa-soutcome" we can now try and predict from the department and age of future patients (1) the most probable category they are in, and (2) the very probability of it. The department and age of 12 novel patients are as follow.

department	age
,00	73,00
,00	38,00
1,00	89,00
,00	75,00
,00	84,00
,00	74,00
1,00	90,00
1,00	72,00
1,00	62,00
1,00	34,00
1,00	85,00
1,00	43,00

Enter the above data in a novel data file and command:

Utilities....click Scoring Wizard....click Browse....Open the appropriate folder with the XML file entitled "exportcategoriesasoutcome"....click on the latter and click Select....in Scoring Wizard double-click Next....mark Predicted category and Probability of it....click Finish.

department	age	probability of being in predicted category	predicted category
,00	73,00	,48	1,00
,00	38,00	,48	1,00
1,00	89,00	,36	2,00
,00	75,00	,47	1,00
,00	84,00	,48	2,00
,00	74,00	,48	1,00
1,00	90,00	,37	2,00
1,00	72,00	,55	,00
1,00	62,00	,65	,00
1,00	34,00	,84	,00

(continued)

department	age	probability of being in predicted category	predicted category
1,00	85,00	,39	,00
1,00	43,00	,79	,00

0=no falloutofbed

1=falloutofbed without injury

2=falloutofbed with injury

In the data file SPSS has provided two novel variables as requested. The first patient from department 0,00 and 73 years of age has a 48 % chance of being in the “falloutofbed without injury”. His/her chance of being in the other two categories is smaller than 48 %.

Conclusion

Multinomial, otherwise called polytomous, logistic regression can be readily trained to make predictions in future patients about their best fit category and the probability of being in it.

Note

More background theoretical and mathematical information of analyses using categories as outcome is available in Machine learning in medicine part two, Chap.10, Anomaly detection, pp 93–103, Springer Heidelberg Germany, 2013, from the same authors.