Chapter 44
Multinomial Regression for Outcome Categories (55 Patients)

1 General Purpose

In clinical research it is not uncommon that outcome variables are categorical, e.g., the choice of food, treatment modality, type of doctor etc. If such outcome variables are binary, then binary logistic regression is appropriate (Chaps. 36, 37, 38, 39). If, however, we have three or more alternatives, then multinomial logistic regression must be used. It works, essentially, similarly to the recoding procedure reviewed in Chap. 8 on categorical predictors variables. Multinomial logistic regression should not be confounded with ordered logistic regression, which is used in case the outcome variable consists of categories, that can be ordered in a meaningful way, e.g., anginal class or quality of life class (Chap. 48).

2 Schematic Overview of Type of Data File

Outcome categorical	predictor
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.

T.J. Cleophas, A.H. Zwinderman, *SPSS for Starters and 2nd Levelers*,
DOI 10.1007/978-3-319-20600-4_44

3 Primary Scientific Question

Do the predictor values significantly predict the outcome categories.

4 Data Example

In a study of 55 hospitalized patients the primary question was the following. The numbers of patients falling out of bed with and without injury were assessed in two hospital departments. It was expected that the department of internal medicine would have higher scores. Instead of binary outcomes, "yes or no falling out of bed", we have three possible outcomes

no falling,
falling without injury,
falling with injury.

Because the outcome scores may indicate increasing severities of falling from the scores 0 to 2, a linear or ordinal regression may be adequate (Chap.48). However, the three possible outcomes may also relate to different types of patients and different types of morbidities, and may, therefore, be presented with nominal rather than increasing values like increasing severities. A multinomial logistic regression may, therefore, be an adequate choice.

Fall out of bed cats 0, 1, 2	Department
1,00	,00
1,00	,00
1,00	,00
1,00	,00
1,00	,00
1,00	,00
1,00	,00
1,00	,00
1,00	,00
1,00	,00
2,00	,00

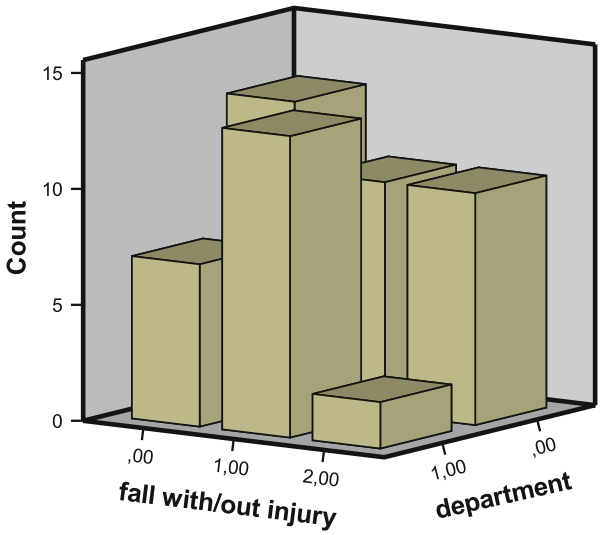
cats 0 = no fall out of bed, 1 = fall out of bed without injury, 2 = fall out of bed with injury; department 0 = internal medicine, 1 = surgery

The entire data file is entitled "chapter44multinomialregression", and is in extras.springer.com. Start by opening the data file in SPSS.

5 3-D Bar Chart

We will first draw a graph of the data.

Command:
Graphs. . .. Legacy Dialogs....3-D Charts. . ...X-Axis: Groups of cases. . ...Z-Axis: Groups of cases. . ..Define. . ..X Category Axis: falloutofbed. . ..Z Category Axis: department. . ..click OK.



The above graph shows that at the department of surgery fewer no-falls and fewer fall with injury are observed. In order to test these data we will first perform a linear regression with fall as outcome and department as predictor variable.

6 Linear Regression

For analysis the statistical model Linear in the module Regression is required.

Command:
Analyze. . ..Regression. . ..Linear. . ..Dependent: falloutofbed. . ..Independent (s): department. . ..click OK.

Coefficients[a]

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	,909	,132		6,874	,000
	Department	−,136	,209	−,089	−,652	,517

[a]Dependent Variable: fall with/out injury

The above graph shows that difference between the departments is not statistically significant. However, the linear model applied assumes increasing severities of the outcome variable, while categories without increasing severities may be a better approach to this variable. For that purpose a multinomial logistic regression is performed.

7 Multinomial Regression

For analysis the statistical model Multinomial Logistic Regression in the module Regression is required.

Command:
Analyze. . ..Regression. . ..Multinomial Logistic Regression. . .. Dependent: falloutofbed. . ..Factor: department. . ..click OK.

Parameter estimates

Fall with/out injury[a]		B	Std. error	Wald	df	Sig.	Exp(B)	95 % confidence interval for Exp (B)	
								Lower bound	Upper bound
,00	Intercept	1,253	,802	2,441	1	,118			
	[VAR00001=,00]	−,990	,905	1,197	1	,274	,371	,063	2,191
	[VAR00001 = 1,00]	0[b]	.	.	0	.	.	.	.
1,00	Intercept	1,872	,760	6,073	1	,014	.		
	[VAR00001=,00]	−1,872	,881	4,510	1	,034	,154	,027	,866
	[VAR00001 = 1,00]	0[b]	.	.	0	.	.	.	.

[a]The reference category is: 2,00.
[b]This parameter is set to zero because it is redundant.

The above graph shows that the odds of falling with injury versus no falling is smaller at surgery than at internal medicine with an odds ratio of 0.371 ($p = 0.274$), and that the odds of falling with injury versus falling without injury is also smaller at surgery than at internal medicine with and odds ratio of 0.154 ($p = 0.034$).

And, so, surgery seems to perform better, when injuries are compared with no injuries. This effect was not observed with linear regression.

8 Conclusion

In research it is not uncommon that outcome variables are categorical, e.g., the choice of food, treatment modality, type of doctor etc. If such outcome variables are binary, then binary logistic regression is appropriate. If, however, we have three or more alternatives, then multinomial logistic regression must be used. It works, essentially, similarly to the recoding procedure reviewed in Chap. 8 on categorical predictors variables. It can be considered a multivariate technique, because the dependent variable is recoded from a single categorical variable into multiple dummy variables (see Chap. 8 for explanation). More on multivariate techniques are reviewed in the Chaps. 17 and 18. Multinomial logistic regression should not be confounded with ordered logistic regression which is used in case the outcome variable consists of categories, that can be ordered in a meaningful way, e.g., anginal class or quality of life class. Also ordered logistic regression is readily available in the regression module of SPSS (Chap. 48).

9 Note

More background, theoretical and mathematical information of categorical variables is given Statistics applied to clinical studies 5th edition, Chap. 21, Springer Heidelberg Germany, 2012, and in Machine learning in medicine a complete overview, chaps 9–11 and 28–30, Springer Heidelberg Germany, 2015, from the same authors.