Chapter 30
Random Intercept Models for Both Outcome and Predictor Categories (55 patients)

General Purpose

Categories are very common in medical research. Examples include age classes, income classes, education levels, drug dosages, diagnosis groups, disease severities, etc. Statistics has generally difficulty to assess categories, and traditional models require either binary or continuous variables. If in the outcome, categories can be assessed with multinomial regression (see the above Chap. 28), if as predictors, they can be assessed with automatic nonparametric tests (see the above Chap. 29). However, with multiple categories or with categories both in the outcome and as predictors, random intercept models may provide better sensitivity of testing. The latter models assume that for each predictor category or combination of categories x_1, x_2,…slightly different a-values can be computed with a better fit for the outcome category y than a single a-value.

$$y = a + b_1 x_1 + b_2 x_2 + ...$$

We should add that, instead of the above linear equation, even better results were obtained with log-linear equations (log = natural logarithm).

$$\log y = a + b_1 x_1 + b_2 x_2 + ...$$

Specific Scientific Question

In a study three hospital departments (no surgery, little surgery, lot of surgery), and three patient age classes (young, middle, old) were the predictors of the risk class of falling out of bed (fall out of bed no, yes but no injury, yes and injury). Are the predictor categories significant determinants of the risk of falling out of bed with or without injury. Does a random intercept provide better statistics.

Example

department	falloutofbed	agecat	patient_id
0	1	1,00	1,00
0	1	1,00	2,00
0	1	2,00	3,00
0	1	1,00	4,00
0	1	1,00	5,00
0	1	,00	6,00
1	1	2,00	7,00
0	1	2,00	8,00
1	1	2,00	9,00
0	1	,00	10,00

Variable 1: department=department class (0=no surgery, 1=little surgery, 2=lot of surgery)
Variable 2: falloutofbed=risk of falling out of bed (0=fall out of bed no, 1=yes but no injury, 2=yes and injury)
Variable 3: agecat=patient age classes (young, middle, old)
Variable 4: patient_id=patient identification

Only the first 10 patients of the 55 patient file is shown above. The entire data file is in extras.springer.com and is entitled "randomintercept.sav". SPSS version 20 and up can be used for analysis. First, we will perform a fixed intercept log-linear analysis.

Command:

click Analyze….Mixed Models....Generalized Linear Mixed Models....click Data Structure….click "patient_id" and drag to Subjects on the Canvas….click Fields and Effects….click Target….Target: select "fall with/out injury"….click Fixed Effects….click "agecat"and "department" and drag to Effect Builder:….mark Include intercept….click Run.

The underneath results show that both the various regression coefficients as well as the overall correlation coefficients between the predictors and the outcome are, generally, statistically significant.

Example 185

Source	F	df1	df2	Sig.
Corrected Model	9,398	4	10	,002
Agecat	6,853	2	10	,013
Department	9,839	2	10	,004

Probability distribution: Multinomial
Link function: Cumulative logit

Model Term		Coefficient	Sig.
	0	2,140	,028
Threshold for falloutofbed=	1	7,229	,000
Agecat=0		5,236	,005
Agecat=1		−0,002	,998
Agecat=2		0,000ª	
Department=0		3,660	,008
Department=1		4,269	,002
Department=2		0,000ª	

Probability distribution: Multinomial
Link function: Cumulative logit
ªThis coefficient is set to zero because it is redundant

Subsequently, a random intercept analysis is performed.

Command:

Analyze….Mixed Models....Generalized Linear Mixed Models....click Data Structure….click "patient_id" and drag to Subjects on the Canvas….click Fields and Effects….click Target….Target: select "fall with/out injury"….click Fixed Effects…. click "agecat"and "department" and drag to Effect Builder:….mark Include intercept….click Random Effects….click Add Block…mark Include intercept….Subject combination: select patient_id….click OK….click Model Options….click Save Fields…mark PredictedValue….mark PredictedProbability….click Save....click Run.

The underneath results show the test statistics of the random intercept model. The random intercept model shows better statistics:

p=0.007 and 0.013 overall for age,
p=0.001 and 0.004 overall for department,
p=0.003 and 0.005 regression coefficients for age class 0 versus 2,
p=0.900 and 0.998 for age class 1 versus 2,
p=0.004 and 0.008 for department 0 versus 2, and
p=0.001 and 0.0002 for department 1 versus 2.

Source	F	df1	df2	Sig.
Corrected Model	7,935	4	49	,000
Agecat	5,513	2	49	,007
Department	7,602	2	49	,001

Probability distribution: Multinomial
Link function: Cumulative logit

Model term		Coefficient	Sig.
Threshold for falloutofbed=	0	2,082	,015
	1	5,464	,000
Agecat=0		3,869	,003
Agecat=1		0,096	,900
Agecat=2		0,000[a]	
Department=0		3,228	,004
Department=1		3,566	,000
Department=2		0,000[a]	

Probability distribution: Multinomial
Link function: Cumulative logit
[a]This coefficient is set to zero because it is redundant

In the random intercept model we have also commanded predicted values (variable 7) and predicted probabilities of having the predicted values as computed by the software (variables 5 and 6).

1	2	3	4	5	6	7 (variables)
0	1	1,00	1,00	,224	,895	1
0	1	1,00	2,00	,224	,895	1
0	1	2,00	3,00	,241	,903	1
0	1	1,00	4,00	,224	,895	1
0	1	1,00	5,00	,224	,895	1
0	1	,00	6,00	,007	,163	2
1	1	2,00	7,00	,185	,870	1
0	1	2,00	8,00	,241	,903	1
1	1	2,00	9,00	,185	,870	1
0	1	,00	10,00	,007	,163	2

Variable 1: department
Variable 2: falloutofbed
Variable 3: agecat
Variable 4: patient_id
Variable 5: predicted probability of predicted value of target accounting the department score only
Variable 6: predicted probability of predicted value of target accounting both department and agecat scores
Variable 7: predicted value of target

Like automatic linear regression (see Chap.) and other generalized mixed linear models (see Chap.) random intercept models include the possibility to make XML files from the analysis, that can subsequently be used for making predictions about the chance of falling out of bed in future patients. However, SPSS uses here slightly different software called winRAR ZIP files that are "shareware". This means that you pay a small fee and be registered if you wish to use it. Note that winRAR ZIP files have an archive file format consistent of compressed data used by

Microsoft since 2006 for the purpose of filing XML (eXtended Markup Language) files. They are only employable for a limited period of time like e.g. 40 days.

Conclusion

Generalized linear mixed models are suitable for analyzing data with multiple categorical variables. Random intercept versions of these models provide better sensitivity of testing than fixed intercept models.

Note

More information on statistical methods for analyzing data with categories is in the Chaps. 28 and 29 of this book.