Mixed Effects Models: Exception 3

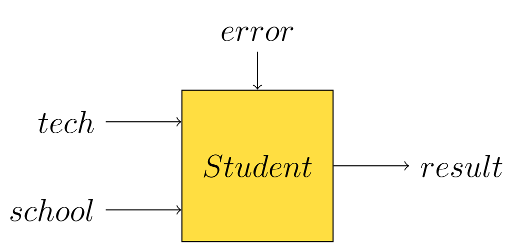
Manas Patnayakuni: BS2017

November 2, 2022

Until now, we have discussed two exceptions of linear mixed effect models that don't belong to the usual set of models that we work with, yet they are meaningful in certain applications. The third such exception that we shall examine now is actually a subset of an important class of models called the Hierarchical linear models(HLM) or Multi-level models.

1 Multi-level Models

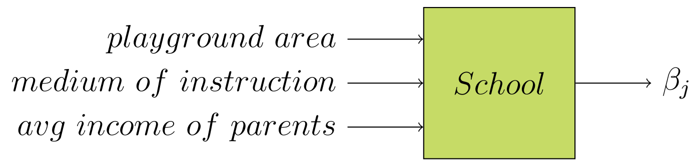
We can understand this exception with the help of an example on teaching statistics. Let's study the effect of technology of teaching on the performance of students in their examinations.



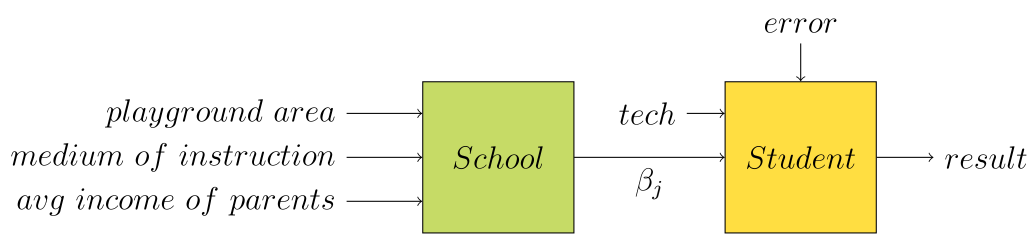
The different types of technology used to teach can be a whiteboard, PowerPoint, smart classroom or an online mode of teaching. In the black box above, we have also taken the school to be a factor since it plays an important role in the performance of the students. In this study, we picked a particular set of schools that we are interested in. We assume that the selection of schools wasn't completely random. The linear model for the above black box can be written as follows:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

In the above model, α_i represents the technology used and β_j represents the effect of school on the performance. Usually, when we conduct such a study, it is natural for us to collect data on other variables pertaining to the school such as the average income of the parents, urban or rural establishment, play area available to the students or the total working hours, etc. This is because these are all factors of the school that can affect the performance of the students. Let us design a black box that shows how these factors can affect β_j.



All of these inputs affect the value of β_j so it makes sense to join the Student black box and the above diagram.



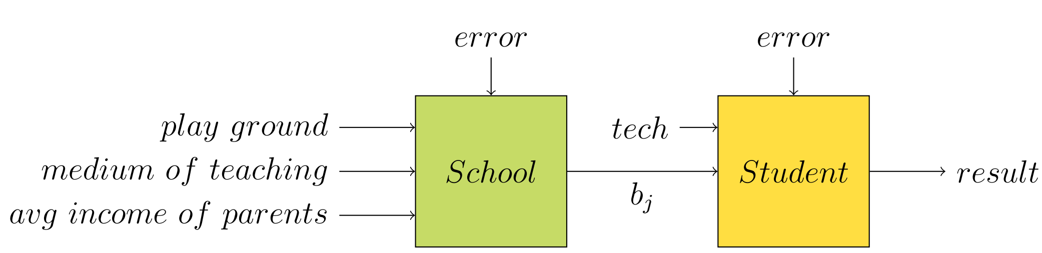
In fact, when we collect the data, we first create a table for the students which contains the technology of teaching provided for them and their result. Next, we create a table for schools separately which contains its corresponding factors. The data collection roughly follows the below template:

Student	Tech	Result
1		
2		

School	avg income of parents	playground area
A		
B		

We can see that the students table is nested within the schools table. This is why we call it a multi-level model where we have student data in the lowest level and the school data on the highest level. If we consider the model to be defined as above, we can observe that the School black box is similar to the Student black box. However, the output for the Student black box is properly defined whereas the output of the School flowchart is β_j which is a bit abstract. Although β_j is a similarly measurable quantity, we cannot say that it can be completely determined by a mathematical expression of playground area, medium of instruction, so on. Hence, we need to add a random error quantity to the School black box similar to how the "result" cannot be completely determined by a mathematical expression of "tech" and "school". As we are adding a random error to the School black box, β_j should be a random variable so we change β_j to b_j. The new linear model for "result" is:

$$y_{ijk} = \mu + \alpha_i + b_j + \epsilon_{ijk}$$



Unlike a typical mixed effects model, instead of assigning a normal distribution to b_j, we can define a proper linear model using the inputs of the School black box.

$$b_j = [...] + \eta_j$$

Furthermore, we can extend the levels of the model and collect data on the cities or states that the schools belong to. This is why it is called a multi-level model. Every input that comes from the higher levels is influenced by a random error making them also a random quantity.

2 Working in R

In R, we can reproduce this multi-level model by using the following piece of code. We can add higher levels by inserting that level of the model in the beginning of the list. The lowest level of the model is in the end whereas the highest level of the model is in the beginning of the list. (Add the inputs of the particular level in the [...] part of the item.)

```
random=list(...,city~[...],school~[...])
```

Such a data set is called a multi-level data set. Normally, all the data is stored in separate tables in a database. However, when we are using the data set in R, we have to combine all the separate tables and present it in a single table as follows:

Student	Tech	Result	School	
			playground area	medium of instruction

Inputting the data in this manner will prompt R to work with the multi-level data set.