# Research Methods and Statistics with R 3

# Week 9 - Linear Mixed Models

## Introduction to the session

In this session, you will be practising using multi-level (Linear Mixed) models in R. The data we will be using is from Belenky et al.’s (2003) study that explored the effects of sleep deprivation on reaction time. Participants were involved in a sleep study for 10 days (incl. ramp up days) and spent either 3, 5, 6, or 9 hours in bed each night (this variable is not included in the dataset – so don’t include in your hypotheses - This is a carry-over from Belenky’s et al study).

Linear mixed methods or *multi-level models* (there are lots of different terms for LMMs)are useful because we can incorporate random effects and largely ignore assumptions of independence and homoscedasticity. However, we should still check the assumptions of general linear models before running a multi-level model. But - in the interest of time – let’s assume that our data are amazing and clean.

## Learning Outcomes

By the end of this session, you will:

1. To apply a multilevel model using R.
2. To decide on and evaluate the best multi-level model to be used.
3. To interpret and summarise the outputs of the multilevel model.

## Pre-workshop Activities

1. Review the lecture material and readings prior to the session, particularly the material regarding assumptions of regression and the lecture on multi-level models.

## General Procedure for Linear Mixed Models

1. Plot the data to help you decide which model will best explain the effect of sleep deprivation on reaction time (*random slope and/or random intercept*)
2. Create a fixed-effect model to explain the sleep data (this should be refresher of what you have already done in other workshops).
3. Create a random-effects model of the sleep data.
4. Create a mixed effects model of the sleep data to explain reaction time due to sleep deprivation.
5. Explain your results.

## Procedure

1. Before you proceed to the exploration and analysis of your data, make sure that you are clear about the hypotheses. Start a new script and include your research question(s) and hypotheses in the comments.
2. Install and load the packages you think will be necessary for this analysis. These will likely include {**lme4}, {ggplot2},** and **{car}**.
3. The *sleepstudy* dataset that we will working with is conveniently contained within the **lme4** package. You can use **str()**, **dplyr::glimpse()**or any other method for inspecting the data you like to get a better feel for the data.
4. Plot the data to visualise the relationship between reaction time and the number of days of sleep deprivation. **Is there a trend here?**
5. Is this trend the same for each subject? Plot the relationship between number of days and sleep deprivation for each subject (**hint:** you can use **color** part of **aes()** inside of **ggplot()** for each *subject*). Try plotting the relationship of reaction time and sleep deprivation on one plot and try plotting each subject on a separate plot in a grid (**hint**: you could use **facet\_wrap(. ~subject)**). Is one easier to interpret than the other? Which would you use in a write-up (Note: there isn’t a right answer – you decide! 😊)?
6. Let’s model the data! Create a fixed-effects model (i.e., a simple linear model treating all the factors as fixed). This is just to help us understand the differences compared to the multilevel model.
   1. ***Optional:*** You can replace the **lm()** function with the **nlme::gls()** function (generalised least squares – this is similar to a linear model). This will allow you to get an AIC and BIC if you wanted to compare the intercept only model to the AIC and BIC of the other models.
7. Do you think it would be better to use a random slope or a random intercept model? You can use the plot that you generated before to help you decide.
8. Let’s evaluate to see if you were right. Create a random intercept model allowing the intercept to vary for each subject. Save the model as something like *intercept.model* if you want to do the follow optional exercise below\*. Remember, to allow the intercept to vary you will need **(1|random effect)** inside your **lmer()** function.
   1. ***Optional***: Are you happy with the model fit? You could plot the residuals to check this. Here is some code that you can use to do so:

> sleepstudy$fitted.mod <- fitted(intercept.model)

> ggplot(data = sleepstudy, mapping = aes(y = resid(lmer.sleepstudy), x = Days)) + geom\_hline(yintercept = 0) + geom\_point() + geom\_line() + facet\_wrap(. ~ Subject)

* 1. As you can see, there is some structure left in the residuals. Let’s keep modelling to see if we can improve this.

1. Create a more complex model. Here, we want to estimate the (fixed) effect of the days of sleep deprivation on response time, while allowing each subject to their have their own effect. That is, we want to estimate a *random slope* for the effect of **day (Days | Subject).** The fixed **Days** effect can be thought of as the average slope over subjects.
   1. ***Optional:*** Check the model fit for this more complex model. Does the model fit the data better than the random intercept model?
2. Now we will want to compare the models the two multilevel models that we have created using the **anova()**function. Remember to set the **test** **=** **“Chisq”**.
3. Take a look at the output of the chi square test. The AIC and BIC are useful statistics to compare between models, where smaller is better. As you night can see, there is a significant difference between the models. How would you interpret this result? (**Hint:** see the explanation in the lecture for how to explain what a non-significant result means).
4. Write up the results. You can follow the suggestion from the lecture.

\*If you get the following error:

> Error in initializePtr() : function 'Rcpp\_precious\_remove' not provided by package 'Rcpp'

then you will need to install and load Rcpp:

install.packages('Rcpp')

library(Rcpp)

## Coding Challenge

Want to explore this deeper? The first two days were adaptation and training days (T1/T2) and the actual sleep deprivation didn’t start until day 2 of the study. Perhaps your models could be further improved by incorporating this into your models? You could use **subset=Days>=2** in your **lmer()** models.