**Homework 1: Mixed-Models course 2016**

**The homework consists of 4 parts that include installing a few packages relevant for mixed models, some reading, and doing some analyses and figures in R.**

**What you have to hand in:**

For (1) and (3), you don't have to hand in anything. For (2) and (4), you hand in your R code and the output and figures you created with it: Copy/paste the code and the output/figures into a word document, print it out, and bring it to the class/lab session on Feb 15, 2016 (=deadline!).

**(1) Install packages, check versions, do a quick check whether lme4 (the mixed-model) package runs without problems.**

**If you work on your own computer:** Download and install the packages described below; then run your very first lmer model to see whether everything works fine.

**If you work on a lab computer:** Run only the very last commands (from library(lme4) on) and see whether it works or whether you get an error message. If you get an error message, try to install the packages as described earlier in the code below, and then try to run the lmer model again.

In any case, if you download and install packages, after downloading and installing make sure to check the versions of these packages to see whether you have a version that is as new (or newer) than the version I recommend.

As you see towards the end of the code, run then a simple lmer model; you don't yet have to understand yet what it means, this is just to see whether it runs or gives an error. If you get an error warning, let me know.

**Here's the code to install the relevant packages and check the versions (plus some explanations):**

# first you need to install the needed packages; some of them are 'development' versions not yet available on the default repository, thus for some packages, slightly different than the usual commands to install them have to be used

# I recommend to install them in the following order

# 1. package Matrix (lme4 needs this for its computation; when you load lme4, it will automatically load the package 'Matrix;' thus, once 'Matrix' is installed, you can basically forget about it)

install.packages('Matrix', repos='http://cran.us.r-project.org')

# check which version was installed with the following command

packageVersion('Matrix') # it's probably 1.2.3 (the most recent); but even 1.1.1.1 (or higher) is most likely fine

# 2. package pbkrtest (this will be used later in the course, as it is one way to get p values for predictors); this command will download several required packages

install.packages('pbkrtest', repos='http://cran.us.r-project.org')

# check the version

packageVersion('pbkrtest') # it's likely to be 0.4.6 (if it's something close to that version number, it should be fine)

# 3. package lme4; this is our main mixed-models package

install.packages("lme4", repos = c("http://lme4.r-forge.r-project.org/repos", getOption("repos")))

# check the version

packageVersion('lme4') # it's likely to be 1.1.10 (again, a version number close to that should be fine)

**Now let's check whether you can load lme4 and run a simple mixed-effects model:**

library(lme4)

fm1 <- lmer(Reaction ~ Days + (Days | Subject), sleepstudy)

summary(fm1)

**Did the model run without any error? Good! If you got an error, let me or Jesse know, so we can check what the problem might be.**

**(2) Read and work through the lm tutorial**

It is available here: http://www.bodowinter.com/tutorial/bw\_LME\_tutorial1.pdf

Full reference: Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]

It's recap from the last statistics course, as it's about linear regression. But mixed-effects models are closely related to linear regression and therefore it's a good idea to repeat and refresh linear regression at the start of this course.

In the tutorial, there are several things that the author tells you to compute yourself (create a data frame, run models, etc): Do all these things yourself in R; and as proof that you did it, you hand in this R script.

Do not yet continue with the lmer/lme4 tutorial; we'll do that later (for didactic reasons, of course).

**(3) Read the Aarts et al. (2014) Nature Neuroscience paper**

Available for example here: http://www.nature.com/neuro/journal/v17/n4/pdf/nn.3648.pdf

**Comments**

* I like this paper because it does a nice job of introducing the idea of non-independent data. However, some of the things they talk about is less relevant for us and their examples might sound a bit unfamiliar.
* For example, they talk about clusters of data (e.g., from the same bunch of neurons, or from the same mouse). To make things perhaps a bit more similar to the kind of data sets that we often have, you can think of a "cluster" as a participant who participates in a experiment with repeated measures. I.e., each participant provides several data points and therefore, these data points are not independent of each other.
* The paper focuses mainly on one relatively simple data structure that results from an experiment with 1 between-subjects factor and involves repeated measures. In our course, we will cover that type of data structure as well, but we will discuss other (and more complex) cases as well.
* They describe approaches and formulas to computing things like *p* values, CIs, and SEs that are, well, not always optimal. They are easy to compute and thus very convenient, but there is evidence that these approaches can lead to biased and/or unreliable results. Thus, read these parts with a grain of salt: It's good to be familiar with these approaches and formulas, but in our course, we will focus more strongly on methods that have been shown to be more reliable, robust, and precise (and complex... there's always a price to pay).

**(4) Recreate the 3 plots (dotplot, densityplot, xyplot) I showed pretty much at the end of my slides.**

These types of plots (we'll encounter more of those in the future) are handy ways to understand the data set that you want to analyze using mixed models (this should only be done **after** you have written down your hypotheses you want to investigate, but we'll discuss that later, too).

Load the "CherryPit\_small\_N5\_30Jan2015\_correct.csv" data file (available on BlackBoard: Week 1) and create the **point plot**, the **densityplot**, and the **xyplot** I showed in my slides. Note, I used ggplot2 for the first one and lattice for the latter two. Feel free to use ggplot2 for all of them.

Please note: The ggplot2 command I used was geom\_point; you could also use other commands, e.g., geom\_dotplot, it will look slightly different than what I did. Either is fine, so use what you like better.

The variables in the data frame are:

* pid = participant code
* trial: cp1 = first cherry pit spit trial, cp2 = second trial, etc
* distance: how far the pit flew (in meters)