**Andre Vargas**

**Homework 5: Mixed-Models course 2016**

**The homework consists of two parts, some R'ing (figure out fixed and random effects, posthocs and follow-up models), and some reading.**

**What you have to hand in:**

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, and upload it to BlackBoard. Deadline is March 14, 2016 at 13:44 sharp!

**(1) ChickWeight data analysis**

R comes with the data set ChickWeight (it is automatically loaded when you start R).

Here's some basic information about this data set. (Do NOT look at the help file, as it is an important learning experience if you can figure out things yourself.)

In an experiment, chicks were fed different diets (variable Diet) and their weight was measured since birth every other day (variable Day: 0, 2, 4, 6, ... up to 20 and also day 21).

**You want to use a "maximal model" to answer the following research questions:**

* Does chick weight change over time?
* Does weight differ as a function of diet in the middle of the observation period?
* Do the diets differ in their change of weight over time?

(Further below, you'll find some additional research questions, but for now, focus on these here.)

**Variables in the data set**

* weight: That's the dependent variable. The weight of the chick in gram
* Time: Age (in days) of the chick (for some unknown reason, they measured the weight every second day starting at day 0 up to day 20, but then again also on day 21 (which is the last day in the observation period))
* Chick: ID variable, tells you which data points come from the same chick (i.e., like 'participant code'); since this variable consists of numbers as entries, it's good practice to turn it into an explicit factor with entries like chick\_1, chick\_2, etc
* Diet: A factor indicating which Diet was fed. Note: This variable also has numbers as entries, so it's safer to turn it into an explicit factor with entries like Diet\_1, Diet\_2, etc

**Do the following steps** (please note that in the interest of time we are skipping some crucial steps, such as looking at and plotting the raw data after writing down your model and hypotheses but before running the model!)**:**

**(a) Check the data frame to see whether there are any missing values.**

How many missing values are there, where do they occur? **5 missing values**

**(b) Figure out and write down the grouping factor(s) of your model**

Chick

**(c) Figure out and write down the fixed-effects structure of your model**

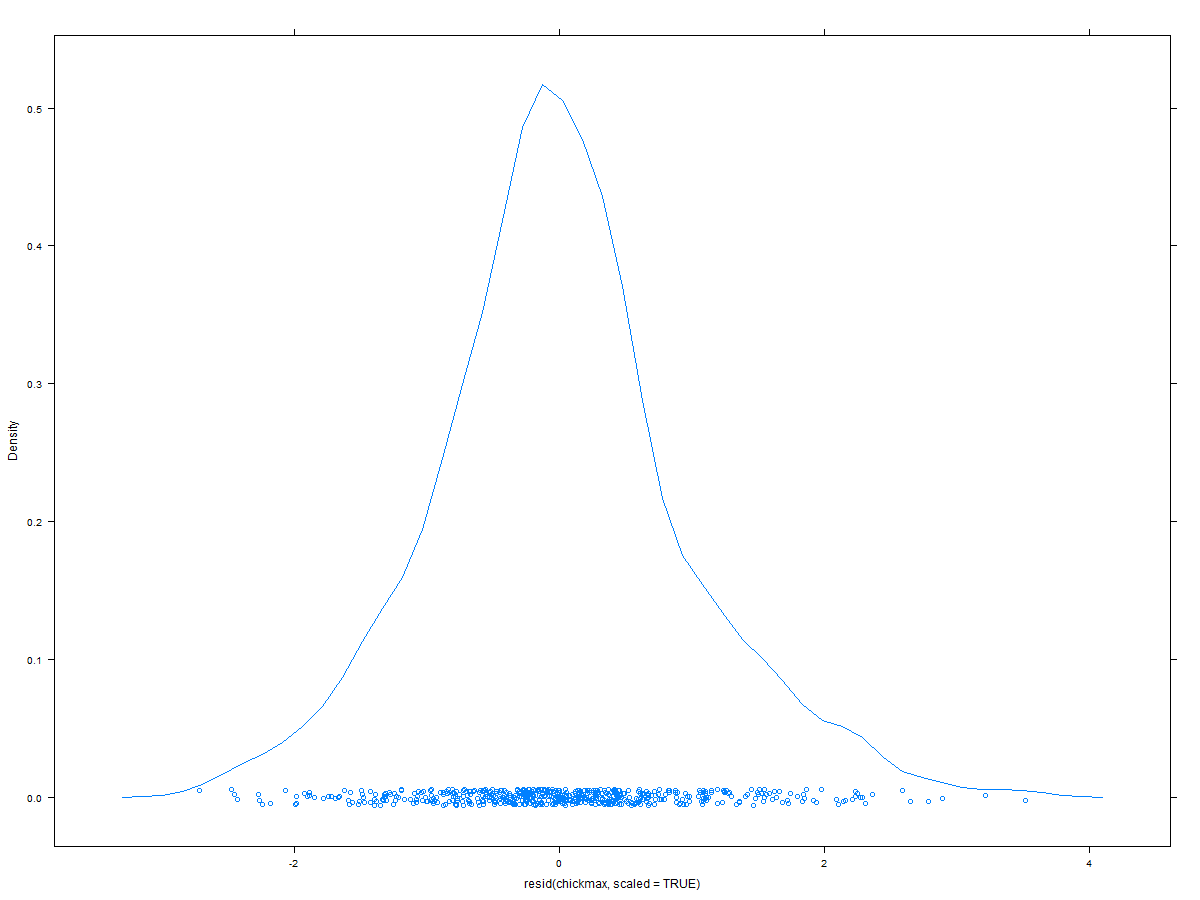
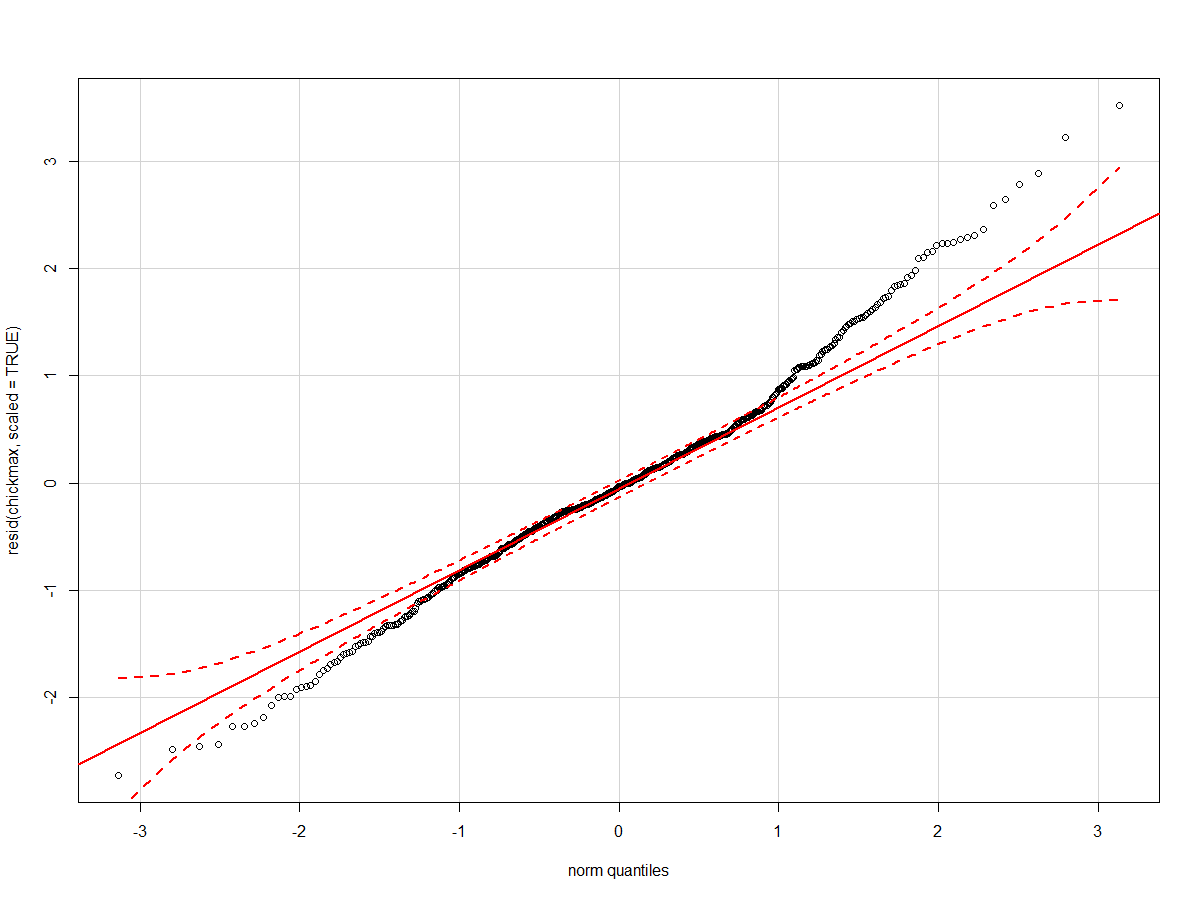
Weight ~ dietf\*timec

**(d) Use crosstables to figure out for each of your fixed effects (don't forget interaction(s) whether you can model it as random slope or not.**

**(e) Based on steps (a) to (d) , write down the 'maximal model' you want to use to answer the research questions above.**

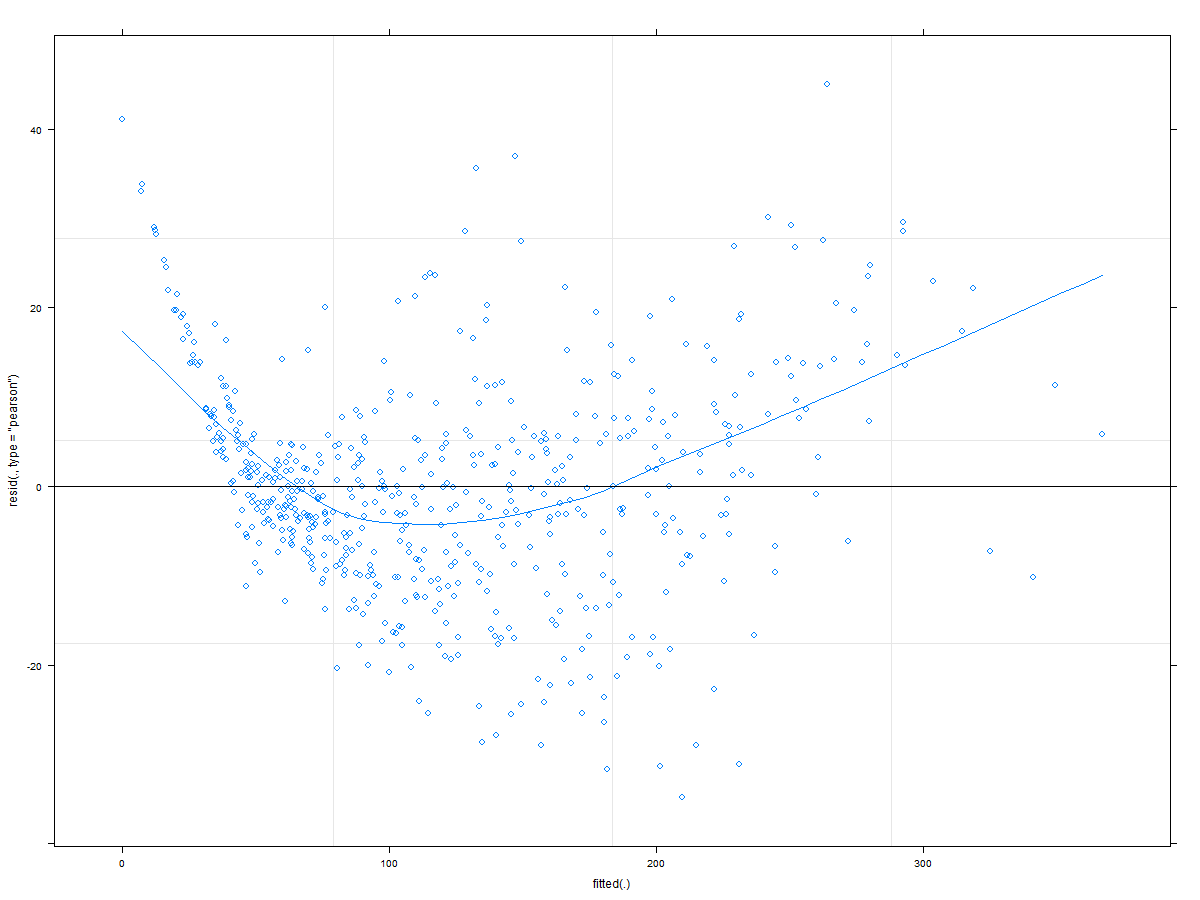
lmer(weight ~ dietf\*timec + (1 + timec| id), data = chick)

**(f) Run that model, thoroughly inspect the summary output (all ok? something odd?), and do the following 3 diagnostic steps:**

* densityplot of the scaled residuals
* 
* qqplot of the scaled residuals (use qqPlot from car)
* 
* compute the proportions of residuals with values larger than +/- 2, 2.5, 3

**in order: 0.046, 0.012, 0.003**

* plot of the fitted versus residuals (with smoothed line) to check for homo-/heteroskedasticity



**(g) Based on (f), do you think there is a reason for concern, or not? If so, what seems most problematic?** Yes, there are reasons… first there might be some residuals bigger than 3 (outliers), homoscedasticity is an issue, normality for residuals is violated.

**(h) Determine *p* values for the fixed effects using bootstrapped Likelihood Ratio Tests.**

I recommend using mixed() for that; also, use several cores and only 500 simulations, as this will already take a while.

**DietF**

**stat df p.value**

**LRT 18.692 3 0.0003166 \*\*\***

**Timec**

**stat df p.value**

**LRT 104.13 1 < 2.2e-16 \*\*\***

**dietf:timec**

**stat df p.value**

**LRT 15.85 3 0.001217 \*\***

**(i) Do post-hocs (I recommend using lsmeans) to answer the following questions:**

* Which diets differ from each other with respect to the chicks weight in the middle of the experiment?
* **contrast estimate SE df t.ratio p.value**
* **Diet\_1 - Diet\_2 -19.968201 9.850425 45.44 -2.027 0.1932**
* **Diet\_1 - Diet\_3 -39.742520 9.850425 45.44 -4.035 0.0012**
* **Diet\_1 - Diet\_4 -33.136361 9.852451 45.47 -3.363 0.0083**
* **Diet\_2 - Diet\_3 -19.774319 11.298863 44.94 -1.750 0.3106**
* **Diet\_2 - Diet\_4 -13.168160 11.300629 44.97 -1.165 0.6515**
* **Diet\_3 - Diet\_4 6.606159 11.300629 44.97 0.585 0.9362**
* In which diets is there significant evidence of change in weight over time?

**dietf timec.trend SE df lower.CL upper.CL**

**Diet\_1 6.276995 0.7616637 46.80 4.744555 7.809434**

**Diet\_2 8.609136 1.0591352 44.96 6.475871 10.742402**

**Diet\_3 11.422871 1.0591352 44.96 9.289606 13.556136**

**Diet\_4 9.531971 1.0600313 45.11 7.397095 11.666846**

* Which diets differ from each other significantly in the change in weight over time?

**contrast estimate SE df t.ratio p.value**

**Diet\_1 - Diet\_2 -2.3321415 1.304569 45.57 -1.788 0.2924**

**Diet\_1 - Diet\_3 -5.1458762 1.304569 45.57 -3.945 0.0015**

**Diet\_1 - Diet\_4 -3.2549760 1.305296 45.67 -2.494 0.0744**

**Diet\_2 - Diet\_3 -2.8137347 1.497843 44.96 -1.879 0.2516**

**Diet\_2 - Diet\_4 -0.9228345 1.498477 45.03 -0.616 0.9265**

**Diet\_3 - Diet\_4 1.8909002 1.498477 45.03 1.262 0.5915**

**(j) Do follow-up models comparing Diets 1 and 3**

Do not forget to re-center your centered predictor(s) after creating subsets of the full data frame! Use again mixed() with bootstrapped LRTs to obtain *p* values to answer the questions below.

* Do Diets 1 and 3 differ from each other significantly in the weight of the chicks in the middle of the experiment? **Yes,**

**stat df p.value**

**LRT 12.882 1 0.0003318 \*\*\***

* Do Diets 1 and 3 differ from each other significantly in how much the chicks' weight changes over time? **Yes,**

**stat df p.value**

**LRT 12.902 1 0.0003282 \*\*\***

**(k) Do follow-up models to investigate whether there is a significant change in weight over time:**

* **in Diet 1**
* **stat df p.value**
* **LRT 30.31 1 3.682e-08 \*\*\***
* **in Diet 3**
* **stat df p.value**
* **LRT 25.201 1 5.166e-07 \*\*\***

Notes

* As before, do not forget to center again the centered predictor(s) in each new data frame.
* Use again mixed() with bootstrapped LRTs to obtain the *p* values to answer the research questions.

So, what do you conclude from the post-hocs and the follow-up models? Do the conclusions differ or not? **No, they do not differ; they all show a significant effect for diet 1 and 3**

**(l) Write-up**

Write up the following things:

* Describe your main model in enough detail with respect to your fixed and random effects so that a person with lme4 knowledge could write down the model syntax (but do not use expressions that **only** lme4 users would understand, i.e., do not simply copy/paste the lme4 syntax, as somebody who for example uses SAS to do mixed-models would not understand this).
* Describe how you obtained *p* values for your main model.
* Describe the statistic results you obtained from your main model (including, where appropriate, the coefficients (+standard error), test statistics, and *p* values)
* Write down in a few sentences (really only 1 or 3 sentences) what you conclude with respect to the research questions.
* Describe how you did the post-hocs and follow-up models and what you conclude from them. (You can start this section by saying something like "We were particularly interested in comparing Diets 1 and 3 and thus conducted additonal analyses investigating them: First, we did a ....")

**The ChickWeight was analyzed with a linear mixedeffects model approach, using the lmer function of the lme4 package (version 1.1.-10; Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2015). Following Barr, Levy, Scheepers, and Tily's (2013) advice to use a maximal random-effects structure: The repeated measures nature of the data was accordingly modeled by including a per-participant random adjustment to the fixed intercept ("random intercept"), as well as per-participant random adjustments to the time slope ("random slopes"); in addition, we included all possible random correlation terms among the random effects.**

**Specifically, the model included a fixed intercept, a fixed effect for the factor diet (coded using sum-to-zero contrasts, with diet 4 as -1 and with diet 1, diet 2 and diet 3 as +1, with their respective levels), a fixed slope for the continuous predictor time (which was centered before we entered it into the model), and a fixed effect for the interaction between diet and time. Moreover, the random slope was the same centered version of diet, and the random intercept was the id code of the chicks. The dependent variable was the weight of the chicks.**

**P values were determined using the function mixed from the package afex (Singmann, Bolker, & Westfall, 2015), using type 3 tests and the parametric bootstrap method (with 500 simulations), which in turn calls the function PBmodcomp from the package pbkrtest (Halekoh & Højsgaard, 2014, version 0.4.6). Furthermore the P values of interested where calculated with the likelihood ratio test.**

**The maximal model reported that the effect of time on weight was significant (coef = 8.96, se = .49, LRT(1) = 104.13, p < .001). There was a significant main effect of diet difference (LRT(3) = 18.69, p < .001). However, there was a significant interaction between time and diet (LRT(3) = 15.85, p = .001). Therefore we can conclude that chicks weight change over time. That weight differs as a function of the diet in the middle of the observation perios, and that dissimilar diets change weight in a different way over time.**

**We were interested in knowing which diets differ from each other with respect to the chicks weight in the middle of the experiment, in which diets is there significant evidence of change in weight over time, and to know which diets differ from each other significantly in the change in weight over time. Therefore, we conducted additional analyses investigate this questions.**

**To answer the first question we used the package “lsmeans” (version 2.23; Lenth, 2016) to do pairwise comparisons (Tukey method) that resulted in only two significant results. Diet 1 (ls*M* = 100.93, *SE* = 5.76, df = 47.69, 95% CI [89.35, 112.52]) and diet 3 (ls*M* = 140.68, *SE* = 7.98, df = 45.98, 95% CI [124.59, 156.76]) differ from each other with respect to the chicks’ weight in the middle of the experiment (p = .001). Furthermore, Diet 1 (ls*M* = 100.93, *SE* = 5.76, df = 47.69, 95% CI [89.35, 112.52]) and diet 4 (ls*M* = 134.07, *SE* = 7.99, df = 46.04, 95% CI [117.98, 150.16]) differ from each other with respect to the chicks’ weight in the middle of the experiment (p = .008). To answer the second question we used the package “lsmeans” (version 2.23; Lenth, 2016) to do pairwise comparisons. We found that all diets changed weight over time: Diet 1 (Trend = 6.27, *SE* =.76, df = 48.24, 95% CI [4.74, 7.8]), diet 2 (Trend = 8.6, *SE* =1.05, df = 46.54, 95% CI [6.47, 10.74]), diet 3 (Trend = 11.42, *SE* =1.05, df = 45.54, 95% CI [9.29, 13.55]), and diet 4 (Trend = 9.53, *SE* = 1.06, df = 46.69, 95% CI [7.39, 11.66]). Finally for the last question we used the package “lsmeans” (version 2.23; Lenth, 2016) to do pairwise comparisons. We found that diets 1 and 3 differ from each other significantly in the change in weight over time (Estimate = -5.14, *SE* = 1.3, df = 47.11; p = .001)**

**(2) Some reading**

The 2 papers you are asked to read have become very very influential in the mixed-models world. They give very clear and hands-on advice how to proceed to do confirmatory hypothesis testing with mixed models (i.e., to use "maximal" models) and how to proceed when you have problems that your model doesn't converge.

The second paper (Barr et al., 2013) is kind of a little add-on to the main (and much longer) Barr et al. (2013) paper.

**(a) Barr, Levy, Scheepers, & Tily (2013)**

This paper is available, e.g., here: http://www.sciencedirect.com/science/article/pii/S0749596X12001180/pdfft?md5=7f4d4c5f8d38e046320d5f05642b473e&pid=1-s2.0-S0749596X12001180-main.pdf

**(b) Barr (2013)**

available, e.g., here: http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00328/full

**Important**

These two papers (in particular the first) are not that easy to understand in all respects. For example, there are lots of formulas and different types of simulations. Further, they focus on a type of analysis that is not that common our usual research, but they are pretty common in linguistic/language research (these F1 and F2 tests) because these experiments typically involve random samples of participants **and** random samples of stimuli.

Thus, try not to be overwhelmed by those details but focus on the more important and more general messages which I list below (I know that this is easier to write than to do).

So: Please read everything, but be aware that you don't have to understand all the details and formulas. The most important things are the following main messages:

* Why is it important to specify "maximal" random effects, and how do you do it?

**Why:**

1. **To avoid anti-conservativity and maximal LMEMs showed better retention of their power relative to ANOVA-based approaches.**
2. **It avoids some problems with F1 and F2 analysis by-item random intercept variation is mixed with trial-level noise in the F1 analysis, and by-subject random intercept variation is mixed with trial-level noise in the F2 analysis**

**How:**

1. **For confirmatory analyses, a design-driven approach is preferable to a data-driven approach for specifying random effects.**
2. **There are few circumstances in which we can know a priori that a random-intercepts-only model is truly justified, and by knowing which factors are within-unit, and which are between.**
3. **General rule of thumb: for whatever fixed effects are of critical interest, the corresponding random effects should be present in that analysis.**

* How to use model comparisons to get *p* values (i.e., which fixed/random effects to remove, which to leave in the models that you want to compare)

1. **Compare a model containing the fixed effect of interest to a model that is identical in all respects except the fixed effect in question. (restricted vs. not restricted)**
   1. **Don’t remove random effect associated with fixed effect**
   2. **Probably not applicable to datasets of the typical size of (?)**

* What to do if models don't converge; i.e., steps how to simplify your model. An important section for this is Barr et al. (2013), pages 275-277

1. **Check for possible misspecifications or data problems that might account for the error.**
2. **Use standard outlier removal methods and to center or sumcode the predictors.**
3. **Increase the maximum number of iterations in the estimation procedure.**
4. **Identify simplification of random effects (first check whether the nonconvergence might be attributable to the presence of a few subjects with small numbers of observations in particular cells.)**
   1. **Removing random correlations and within-unit random intercepts, might be a good strategy**
   2. **Preferable to remove (or replace) these few subjects (or items)**
   3. **For severe convergence problems use a data-driven approach (building from simple model)**
5. **Alternative to perform separate by-subject and by-item LMEMs, similar in logic to F1 x F2, each with appropriate maximal random effect structures.**

###############Rcode##############

install.packages('lme4')

library (lme4)

install.packages('car')

library (car)

install.packages('psych')

library (psych)

install.packages('afex')

library (afex)

install.packages('pbkrtest')

library (pbkrtest)

install.packages('lattice')

library (lattice)

install.packages('parallel')

library (parallel)

install.packages('lsmeans')

library (lsmeans)

chick <- ChickWeight

options(contrasts=c("contr.sum", "contr.poly"))

chick$dietf <-as.factor(paste("Diet", chick$Diet, sep = '\_'))

chick$id <-as.factor(paste("ID", chick$Chick, sep = '\_'))

chick$timec <- scale(chick$Time, center = T, scale = F)

#(a) Check the data frame to see whether there are any missing values.

which(is.na(chick))

summary(chick)

describe(chick)

is.na(chick)

as.data.frame(table(chick$Time)) # five missing values.. chicks died

contrasts(chick$dietf)

#(b) Figure out and write down the grouping factor(s) of your model #chick

#the above is just confirming what theoretically makes sense

#(c) Figure out and write down the fixed-effects structure of your model

#dietf and timec... we want to know how things change in time and how diet affects the weight

#(d) Use crosstables to figure out for each of your fixed effects (don't forget interaction(s) whether you can model it as random slope or not.

with(chick, table(dietf, id)) #between subject factor because diest stays the same only one for each pp

with(chick, table(timec, dietf, id))

with(chick, table(timec, id))#time as a slope becasue its continues you need 3 or more points.. and it-s repeated

#(e) Based on steps (a) to (d) , write down the 'maximal model' you want to use to answer the research questions above

chickmax <- lmer(weight ~ dietf\*timec + (1 + timec| id), data = chick)

summary(chickmax)

#desnity plot reiduals

dplot <- lattice::densityplot(resid(chickmax, scaled = TRUE))

#qqplot

qqPlot(resid(chickmax, scaled = TRUE))

#proportion of residuals

sum(abs(resid(chickmax, scaled = TRUE)) > 3)/ length(resid(chickmax))

sum(abs(resid(chickmax, scaled = TRUE)) > 2.5)/ length(resid(chickmax))

sum(abs(resid(chickmax, scaled = TRUE)) > 2)/ length(resid(chickmax))

#fitted vs residuals hetero

plot(chickmax, type = c('p', 'smooth'))

#Determine p values for the fixed effects using bootstrapped Likelihood Ratio Tests

# Create the cluster

n\_cores <- detectCores()

MyCluster <- makeCluster(rep("localhost", n\_cores - 1))

#Run your mixed() command (LRT, KR, PB)

chicklrt <- mixed(weight ~ dietf\*timec + (1 + timec| id), type = 3, method = "PB", data = chick, cl = MyCluster, args.test = list(nsim = 500, cl = MyCluster))

#Each CPU runs a (sub)model! Can save a lot of time.

#Once you're done, stop the cluster

stopCluster(MyCluster)

summary(chicklrt)

#check p-values

chicklrt$tests

########post-hocs########

#Which diets differ from each other with respect to the chicks weight in the middle of the experiment?

chickpost <- lsmeans(chickmax, pairwise ~ dietf)

chickpost

#In which diets is there significant evidence of change in weight over time?

lstrends(chickmax, "dietf", var = "timec")

#Which diets differ from each other significantly in the change in weight over time?

pairs(lstrends(chickmax, "dietf", var ="timec"))

#to report?

describeBy(chick$weight, list(chick$dietf))

#(j) Do follow-up models comparing Diets 1 and 3

# selecting data

#exclude variables

chick2 <- subset(chick, Diet == "1" | Diet == "3" , select = c(weight, Chick, Time, Diet))

describe(chick2)

chick2$timec <- scale(chick2$Time, center = T, scale = F)

chick2$id <-as.factor(paste("ID", chick2$Chick, sep = '\_'))

chick2$dietf <-as.factor(paste("Diet", chick2$Diet, sep = '\_'))

chickmax2 <- mixed(weight ~ dietf\*timec + (1 + timec| id), data = chick2)

summary(chickmax2)

# Create the cluster

n\_cores <- detectCores()

MyCluster <- makeCluster(rep("localhost", n\_cores - 2))

#Run your mixed() command (LRT, KR, PB)

chicklrt2 <- mixed(weight ~ dietf\*timec + (1 + timec| id), type = 3, method = "PB", data = chick2, cl = MyCluster, args.test = list(nsim = 500, cl = MyCluster))

#Each CPU runs a (sub)model! Can save a lot of time.

#Once you're done, stop the cluster

stopCluster(MyCluster)

summary(chicklrt2)

#check p-values

chicklrt2$tests

######################

## these models dont have diet as an IV, because it's just one level

#(k) Do follow-up models to investigate whether there is a significant change in weight over time:

#in Diet 1

#exclude variables

contrast()

chick3 <- subset(chick, Diet == "1" , select = c(weight, Chick, Time, Diet))

describe(chick3)

chick3$timec <- scale(chick3$Time, center = T, scale = F)

chick3$id <-as.factor(paste("ID", chick3$Chick, sep = '\_'))

chickmax3 <- mixed(weight ~ timec + (1 + timec| id), data = chick3)

summary(chickmax3)

# Create the cluster

n\_cores <- detectCores()

MyCluster <- makeCluster(rep("localhost", n\_cores - 2))

#Run your mixed() command (LRT, KR, PB)

chicklrt3 <- mixed(weight ~ timec + (1 + timec| id), type = 3, method = "PB", data = chick3, cl = MyCluster, args.test = list(nsim = 500, cl = MyCluster))

#Each CPU runs a (sub)model! Can save a lot of time.

#Once you're done, stop the cluster

stopCluster(MyCluster)

summary(chicklrt3)

#check p-values

chicklrt3$tests

#####################

#in Diet 3

#exclude variables

chick4 <- subset(chick, Diet == "3" , select = c(weight, Chick, Time, Diet))

describe(chick4)

chick4$timec <- scale(chick4$Time, center = T, scale = F)

chick4$id <-as.factor(paste("ID", chick4$Chick, sep = '\_'))

chickmax4 <- mixed(weight ~ timec + (1 + timec| id), data = chick4)

summary(chickmax4)

# Create the cluster

n\_cores <- detectCores()

MyCluster <- makeCluster(rep("localhost", n\_cores - 2))

#Run your mixed() command (LRT, KR, PB)

chicklrt4 <- mixed(weight ~ timec + (1 + timec| id), type = 3, method = "PB", data = chick4, cl = MyCluster, args.test = list(nsim = 500, cl = MyCluster))

#Each CPU runs a (sub)model! Can save a lot of time.

#Once you're done, stop the cluster

stopCluster(MyCluster)

summary(chicklrt4)

#check p-values

chicklrt4$tests

citation("lsmeans")

sessionInfo()