Exercise 09

Dennis Perrett

1/11/2022

Exercise 24. Box (1950) describes an experiment studying the effect of two substances on the growth of rats. For this purpose, 27 rats were randomly assigned to three groups, one of which had the hormone `thyroxine (txn)` added to the drinking water, and another group `thiouracil (tcl)`, which inhibits the formation of thyroxine. These two groups were compared to a control group `(ctl)` kept with normal drinking water for four weeks. `Body weight` (bw, in grams) was measured in each case at the beginning (0) and at weekly intervals (1, 2, 3, 4). The data are contained in the file box-weight.txt.

Read the data into R and save it to a data frame dat.w. Each time point is represented by one column `(bw0, . . . , bw4)`. Often, this kind of data (the so-called wide format) is to be converted into a format (the so-called log format) that includes only one dependent variable (bw) and a time variable (week). The function reshape() is available in R for converting data between wide and long format.

```
dat.w <- read.table("../box-weight.txt",header=T)
dat <- reshape(dat.w, idvar="rat", varying=list(names(dat.w[, -(1:2)])),
  direction="long", times=0:4, timevar="week", v.names="bw")
dat$rat <- factor(dat$rat)
dat <- dat[order(dat$grp, dat$rat, dat$week), ]
```

Refer to the data frame dat for further analysis, and base statistical decisions on the probability $\alpha = 0.05$ of a Type I error.

a. For ML parameter estimation with the function `lmer()`, specify a preliminary mixed model with group-specific linear time trends as well as individual intercepts. (Note: To avoid problems with convergence it may be necessary to call the function lmer with the additional option `control=lmerControl(optimizer = c("bobyqa"))` choosing an optimizer different from the default.)

Fixed effects refers to normal coefficients. Random effects in this case refers to coefficients for each individual. In the equation given below as `bw ~ week * grp + (1|rat)`, the body-weight is estimated by week, group, a week*group interaction term (all fixed effects), and an intercept for each rat (random effect). We estimate the variability of the intercepts, not the intercept itself. This is why it is a random effect, as each modelled rat, will be given a random +- value based on the variation presented to the model.

```
# (bw ~ `1 + grp). 1 is implicit. If we want to have 0 then we put 0.
lmm.rat.1 <- lmer(bw ~ week * grp + (1|rat), dat, REML=F,
                  control=lmerControl(optimizer = c("bobyqa")))
```

b. Extend the preliminary model by including individual linear time trends and specify the function call of `lmer()` for a ML estimation of its parameters.

The `1` in the model below, is implicit, and does not need to be included. The `week|rat` part, provides an estimate for each rat, which represents the change in weight gain (slope) for each individual rat. Logically, we can consider that a fat rat will gain weight slower given the same amount of hormone/food than a slim rat.

```
lmm.rat.2 <- lmer(bw ~ week * grp + (1+week|rat), dat, REML=F,
                  control=lmerControl(optimizer = c("bobyqa")))
```

c. Which of these models is to be preferred when compared by a likelihood ratio test?

The anova test, tells us whether the more complex model (npar) describes the data better than a more simple model. In this case the more complex model (lmm.rat.2) is significantly better at the 1% level. We can conclude that a model with individual linear time trends models the data better.

```
anova(lmm.rat.1, lmm.rat.2)
```

```
## Data: dat
## Models:
## lmm.rat.1: bw ~ week * grp + (1 | rat)
## lmm.rat.2: bw ~ week * grp + (1 + week | rat)
##           npar    AIC     BIC  logLik deviance  Chisq Df Pr(>Chisq)
## lmm.rat.1    8 986.19 1009.43 -485.09   970.19
## lmm.rat.2   10 916.78  945.83 -448.39   896.78 73.409  2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

d. Do the mean initial weight and the mean linear rate of weight change differ across groups? Answer this question using incremental F-tests.

Here, we want to see if including a grp parameter/coefficient makes the model better. First we compare the base model (no group coefficient) to the same model with the addition of a group coefficient. The anova results in this case are insignificant, suggesting a group effect is not relevent. We continue by including not just the group parameter, but also a week*group interaction term. We compare this to the model with the main effect (lmm.2b) because we want to ensure that the interaction term is providing additional information not found in the main effect. In this case, the most complex (lmm.2) model is significant.

```
lmm.2a <- lmer(bw ~ week +       (week|rat), dat, REML=F)
lmm.2b <- lmer(bw ~ week + grp + (week|rat), dat, REML=F)
lmm.2 <-  lmer(bw ~ week * grp + (week|rat), dat, REML=F) # Same as above model. (group spec. lin. time
anova(lmm.2a,lmm.2b)
```

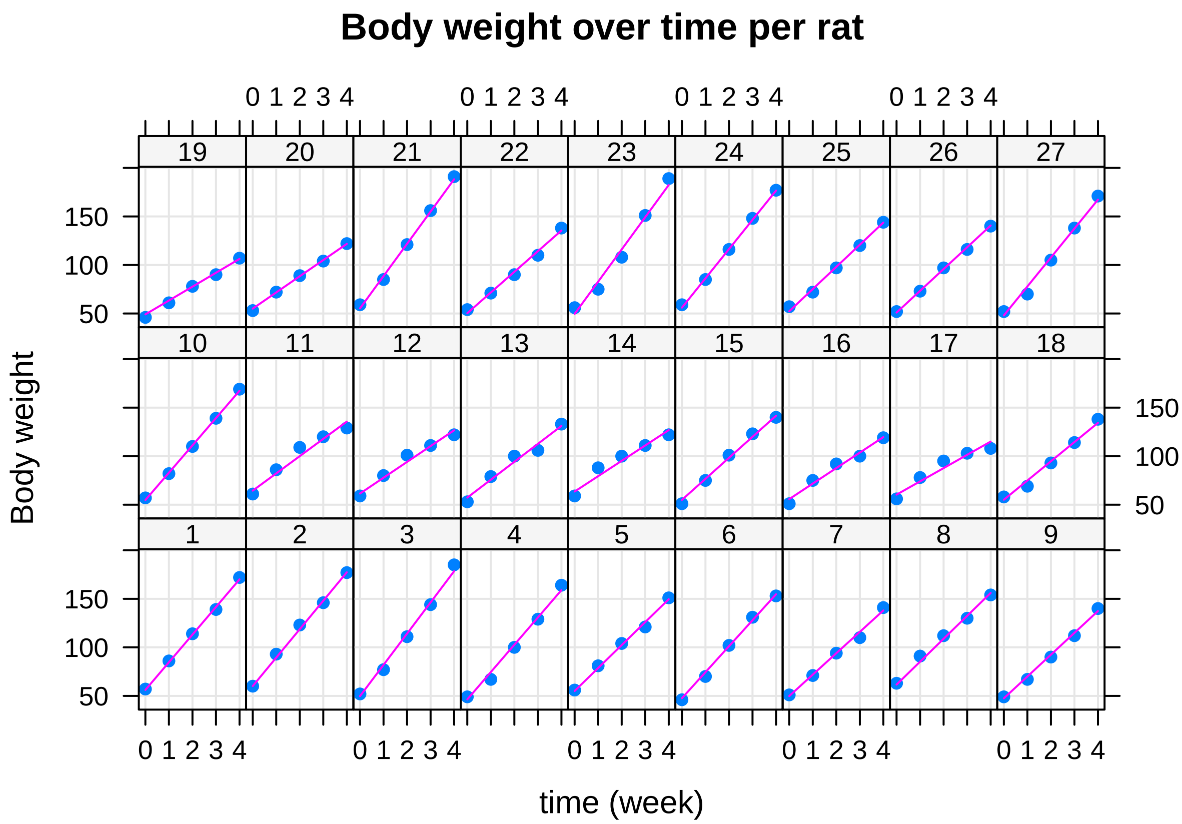
```
## Data: dat
## Models:
## lmm.2a: bw ~ week + (week | rat)
## lmm.2b: bw ~ week + grp + (week | rat)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## lmm.2a    6 933.93 951.36 -460.96   921.93
## lmm.2b    8 937.71 960.95 -460.86   921.71 0.2175  2      0.897
```

```
# Mean initial weight
# No sig. difference.
anova(lmm.2b,lmm.2)
```

```
## Data: dat
## Models:
## lmm.2b: bw ~ week + grp + (week | rat)
## lmm.2: bw ~ week * grp + (week | rat)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## lmm.2b    8 937.71 960.95 -460.86   921.71
## lmm.2    10 916.78 945.83 -448.39   896.78 24.932  2  3.855e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# mean linear rate of weight change
```

e. For each rat, plot the observed weights and the weights predicted by the preferred model as a function of



time.

f. Interpret the parameter estimates of the fixed and random effects of the preferred model in the context of the specific application.

```
summary(lmm.rat.2, correlation=F)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: bw ~ week * grp + (1 + week | rat)
##    Data: dat
## Control: lmerControl(optimizer = c("bobyqa"))
##
##      AIC      BIC   logLik deviance df.resid
##    916.8    945.8   -448.4    896.8      125
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.84321 -0.56097  0.05352  0.58506  2.06556
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  rat      (Intercept) 26.86    5.183
##           week        13.23    3.637    -0.10
##  Residual             18.86    4.342
## Number of obs: 135, groups:  rat, 27
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   52.880      1.954  27.065
## week          26.480      1.229  21.539
```

```
## grptcl         4.820      2.763   1.744
## grptxn        -1.080      3.045  -0.355
## week:grptcl   -9.430      1.739  -5.424
## week:grptxn    1.091      1.916   0.570
```

Random effects: These are the differences for each individual, that cannot be explained by the model.

- rat (intercept): individual variability of the initial weight, not explained by the group membership. This difference may be small, and mostly explained by the base intercept, or each rat could start at drastically different weights, yet have very similar weight gains over time.
- rat (week): standard deviation of the differences in slope for each rat. For example, on average, the rats may gain 4 grams per 26 grams per week, but some may be gaining 30 per week, and some may be gaining 10 per week. This standard deviation represents the variance for the weight gain per week between all rats.
- corr: -0.1, Correlation between random effects (start weight and weight gain). Higher initial weight correlated with lower weight gain (compared to group), although this correlation is very small.

Fixed effects:

These effects are nothing new, and are what we get from a normal regression. * intercept: 52.88 = mean initial weight of control group * week: mean weekly weight increase per week (CG) * grptcl: difference of the initial weight of tcl group compared to control group. * grptxn: difference of the initial weight of txn group compared to control group. * week:grptcl: slope adjustment for grptcl (compared to control group) * week:grptxn: slope adjustment for grptxn (compared to control group)

Exercise 25. Consider again the data of the Box (1950) experiment, and analyze the data contained in box-weight.txt in the long format (data frame dat). For hypotheses testing, assume a Type I error rate of $\alpha = 0.05$.

a. Extend the model considered in the preceding exercise, which includes group-specific as well as individual intercepts and linear time trends, by adding group-specific and individual quadratic time trends. Provide ML parameter estimates using the function lmer().

In this case, we include `I(week^2) * grp`, and `(week + I(week^2)| rat)`. The former evaluates coefficents for a quadratic weekly trend, and a quadratic weekly trend per group. The latter evaluates ceofficients for a linear and quadratic trend (and intercept) per rat. The `I()` function simply evaulates the equation prior to modelling.

```
lmm.rat.3 <- lmer(bw ~ week * grp + I(week^2) * grp + (week + I(week^2)| rat), dat, REML =F)
```

```
## boundary (singular) fit: see ?isSingular
```

b. Which of these models is to be preferred when compared by a likelihood ratio test?

```
anova(lmm.rat.2,lmm.rat.3)
```

```
## Data: dat
## Models:
## lmm.rat.2: bw ~ week * grp + (1 + week | rat)
## lmm.rat.3: bw ~ week * grp + I(week^2) * grp + (week + I(week^2) | rat)
##           npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## lmm.rat.2   10 916.78 945.83 -448.39   896.78
## lmm.rat.3   16 866.42 912.91 -417.21   834.42 62.358  6  1.491e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

lmm.rat.3 is significant. Group specific and individual specific quadratic terms are good.

c. Answer the following question using incremental F -tests.

- Does the mean linear trend differ across groups?

  - Here we drop the baseline linear trend per group and compare to the model with.
  - The results are insignificant, implying there linear time trend per group is not a powerful predictor of rat weight (ie. no difference across groups).

```
lmm.3a <- lmer(bw ~ week +I(week^2) * grp + (week + I(week^2)|rat),dat,REML=F)
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(lmm.rat.3,lmm.3a)
```

```
## Data: dat
## Models:
## lmm.3a: bw ~ week + I(week^2) * grp + (week + I(week^2) | rat)
## lmm.rat.3: bw ~ week * grp + I(week^2) * grp + (week + I(week^2) | rat)
##           npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## lmm.3a      14 863.34 904.01 -417.67   835.34
## lmm.rat.3   16 866.42 912.91 -417.21   834.42 0.9176  2      0.632
```

- Does the mean quadratic trend differ across groups?
  - Here we include the linear time trend per group, and remove the quadratic time trend per group, and compare with a model with quadratic time trends and a linear time trend, and a quadratic time trend per group
  - This time the results are significant, implying that the more complex model (with quadratic time trend per group), describes the data better.

```
lmm.3b <- lmer(bw ~ week*grp +I(week^2)  + (week + I(week^2)|rat),dat,REML=F)
```

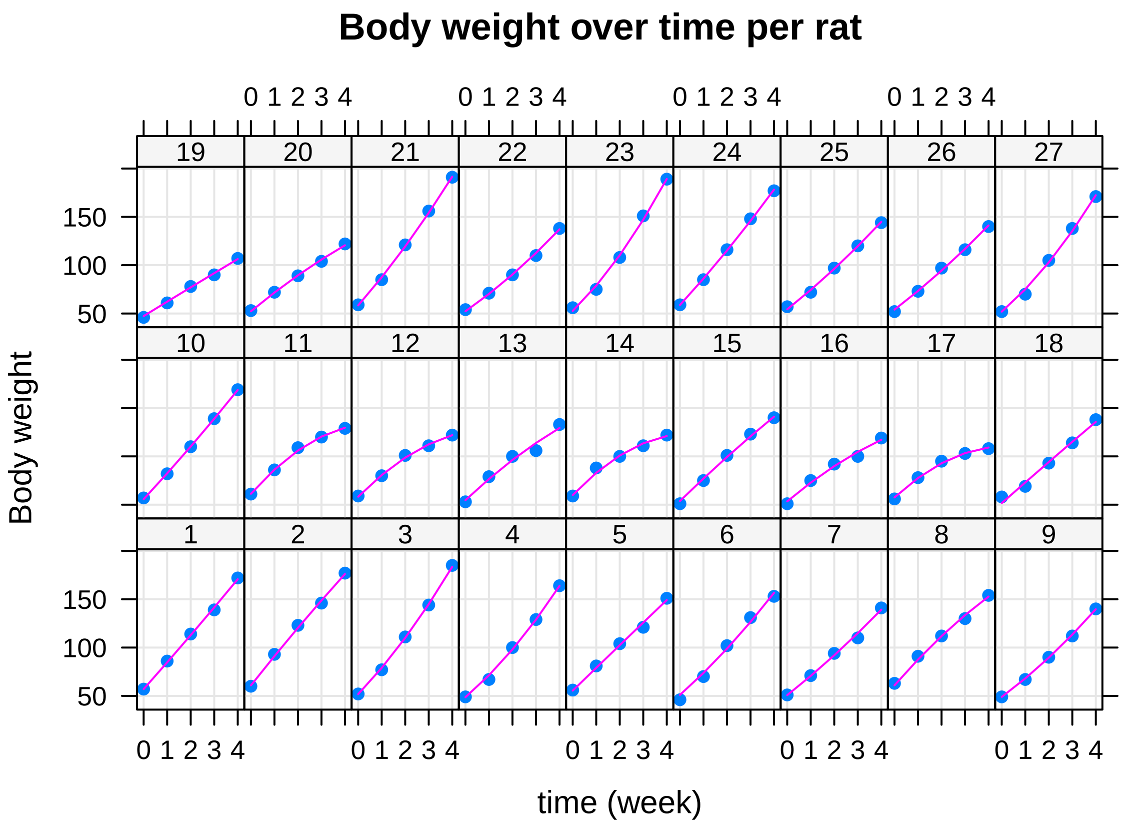
```
## boundary (singular) fit: see ?isSingular
```

```
anova(lmm.rat.3,lmm.3b)
```

```
## Data: dat
## Models:
## lmm.3b: bw ~ week * grp + I(week^2) + (week + I(week^2) | rat)
## lmm.rat.3: bw ~ week * grp + I(week^2) * grp + (week + I(week^2) | rat)
##           npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## lmm.3b      14 880.46 921.13 -426.23   852.46
## lmm.rat.3   16 866.42 912.91 -417.21   834.42 18.034  2  0.0001213 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

d. For each rat, plot the observed weights and the weights predicted by the preferred model as a function of time.

- What we see is that each line of weight gain is now nicely described by the model. Some rats (or groups) such as rats 19 and 20 have very linear weight gains. Rats 23 and 27 both appear to have accelerating weight gains, and 11 and 17 have decelerating weight gains.



e. Interpret the parameter estimates of the fixed and random effects of the preferred model in the context of the specific application.

```
summary(lmm.rat.3)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: bw ~ week * grp + I(week^2) * grp + (week + I(week^2) | rat)
##    Data: dat
##
##      AIC      BIC   logLik deviance df.resid
##    866.4    912.9   -417.2    834.4      119
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.76659 -0.42026  0.06121  0.56599  2.23321
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  rat      (Intercept) 16.1845  4.0230
##           week        14.2976  3.7812    0.87
##           I(week^2)    0.9432  0.9712   -0.87 -0.51
##  Residual              8.1366  2.8525
## Number of obs: 135, groups:  rat, 27
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)        54.1086     1.5294  35.379
## week               24.0229     1.5624  15.376
```

```
## grptcl             0.9200     2.1629   0.425
## grptxn             0.5486     2.3834   0.230
## I(week^2)          0.6143     0.3904   1.573
## week:grptcl       -1.6300     2.2095  -0.738
## week:grptxn       -2.1657     2.4348  -0.889
## grptcl:I(week^2)  -1.9500     0.5522  -3.532
## grptxn:I(week^2)   0.8143     0.6084   1.338
##
## Correlation of Fixed Effects:
##             (Intr) week   grptcl grptxn I(w^2) wk:grptc wk:grptx grptc:I(^2)
## week         0.293
## grptcl      -0.707 -0.207
## grptxn      -0.642 -0.188  0.454
## I(week^2)   -0.372 -0.690  0.263  0.239
## week:grptcl -0.207 -0.707  0.293  0.133  0.488
## week:grptxn -0.188 -0.642  0.133  0.293  0.443  0.454
## grptc:I(^2)  0.263  0.488 -0.372 -0.169 -0.707 -0.690   -0.313
## grptx:I(^2)  0.239  0.443 -0.169 -0.372 -0.642 -0.313   -0.690    0.454
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

Random Effects:

- Rat - Intercept: Same as above. The variance/standard deviation of the unique starting weights for each rat.
- Rat - Week: The variance of the different/unique slopes for each rat.
- Rat - Week^2: As above, except for the quadratic term.

Fixed Effects:

- Intercept: Mean starting weight of the control group.
- Week: Lienar increase in weight of the control group per week.
- grptcl: Difference in starting weight of the tcl group compared to the control group.
- grptxn: Difference in starting weight of the txn group compared to the control group.
- I(week^2): Control group quadratic increase in weight gain per week.
- week:grptcl: Difference in slope of linear weight gain per week for group tcl.
- week:grptxn: Difference in slope of linear weight gain per week for group txn.
- grptc:I(^2): Difference in slope of the quadratic weight gain per week for group tcl.
- grptx:I(^2): Difference in slope of the quadratic weight gain per week for group txn.