assignment3

Ludan Zhang

2017年10月24日

## Problem 1

Fit a multilevel model in which within patient, pain is a function of time and temperature and between patients these relationships may depend on age, race, income, treatment, sex, occupation, working status and use of NSAIDs.

The model should have the form as:

where,

And, = pain at time t for individual i, = time t or temperature at time t for individual i, = patient level factor (age, sex, race, income, occupation, working status, use of NSAIDs).

### Step1. Check if time and temperature both should be included as random effects

First of all, we need to determine which variables should be included as random effects in the model. Thus, we fit different models including different random effects. And we can see the results of AIC from Table 1.1.

|  |  |  |
| --- | --- | --- |
| Table 1.1 AIC of the four models | | |
| Model | df | AIC | |
| m0 | 3 | 5450.491 | |
| m.date | 6 | 5213.111 | |
| m.temp | 6 | 5423.860 | |
| m.date.temp | 10 | 5219.371 | |
| m0 is the model including only intercept as random effects  m.date is the model including intercept and time as random effects  m.temp is the model including intercept and temperature as random effects  m.date.temp is the model including intercept, time and temperature as random effects | | |

From the AIC results above we can see that the model with only time is the best. Then, we try to compare if the m.date.temp is significant different from m.date.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 1.2 ANOVA test resultof m.date.temp compared to m.date | | | | |
| Model | Df | AIC | BIC | Pr(>Chisq) |
| m.date | 6 | 5213.111 | 5243.20 | NA |
| m.date.temp | 10 | 5219.371 | 5269.52 | 0.7834677 |

### The ANOVA result shows that the model including both time and temperature is not significant different from the model including only time. That is to say, temperature is not a significant variable. Thus, we include only time as the random variable in the model.

### Step2. Check other fixed effects

After determining the random effects, we try to find the variable that should be included as fix effect in the model. We add the variables(age, race2, sex, inccat, retire, nsaid and treat) one by one in the m.date model (pain ~ 1 + WeatherDate + (1 + WeatherDate | ID)). Then we compare the AIC of the new model to the m.date model, as well as doing ANOVA test. We will only include the variable as fix effect if the AIC of the new model is smaller and the p value of ANOVA is below 0.05. Table 1.3 shows the result, from which we can see that no variable should be include as fix effect since none of them meets the criterion.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 1.3 AIC and ANOVA result of the model with certain variable as fixed effect | | | | |
| Model | Variable | AIC | P value of ANOVA | Include or not |
| pain ~ age+ 1 + WeatherDate + (1 + WeatherDate | ID) | age | 5213.18 | 0.16 | Not Include |
| pain ~ race2+ 1 + WeatherDate + (1 + WeatherDate | ID) | race2 | 5213.69 | 0.23 | Not Include |
| pain ~ sex+ 1 + WeatherDate + (1 + WeatherDate | ID) | sex | 5212.85 | 0.13 | Not Include |
| pain ~ inccat+ 1 + WeatherDate + (1 + WeatherDate | ID) | inccat | 2557.08 | 0.58 | Not Include |
| pain ~ retire+ 1 + WeatherDate + (1 + WeatherDate | ID) | retire | 4068.28 | 0.37 | Not Include |
| pain ~ nsaid+ 1 + WeatherDate + (1 + WeatherDate | ID) | nsaid | 5215.08 | 0.85 | Not Include |
| |  | | --- | | pain ~ treat+ 1 + WeatherDate + (1 + WeatherDate | ID) | | treat | 5215.05 | 0.81 | Not Include |

### Step3. Check other interaction terms

Furthermore, we want to check if the interaction terms between variables(age, race2, sex, inccat, retire, nsaid and treat) and time can be added as fixed effects in the model. We use the same method as in Step2. Table 1.4 shows the result. We can see that only the variable nsaid\*WeatherDate meets the criterion. Thus, we include this term in the model. Since the interaction term is included, the nsaid itself should be included as well.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 1.4 AIC and ANOVA result of the model with certain variable as fixed effect | | | | |
| Model | Variable | AIC | P value of ANOVA | Include or not |
| Pain ~ age\*WeatherDate+ 1 + WeatherDate + (1 + WeatherDate | ID) | age | 5215.01 | 0.35 | Not Include |
| pain ~ race2\*WeatherDate+ 1 + WeatherDate + (1 + WeatherDate | ID) | race2 | 5215.69 | 0.49 | Not Include |
| pain ~ sex\*WeatherDate+ 1 + WeatherDate + (1 + WeatherDate | ID) | sex | 5214.84 | 0.32 | Not Include |
| pain ~ inccat\*WeatherDate+ 1 + WeatherDate + (1 + WeatherDate | ID) | inccat | 2561.62 | 0.61 | Not Include |
| pain ~ retire\*WeatherDate+ 1 + WeatherDate + (1 + WeatherDate | ID) | retire | 4069.71 | 0.50 | Not Include |
| pain ~ nsaid\*WeatherDate+ 1 + WeatherDate + (1 + WeatherDate | ID) | nsaid | 5210.89 | 0.04 | Include |
| pain ~ treat\*WeatherDate+ 1 + WeatherDate + (1 + WeatherDate | ID) | treat | 5216.27 | 0.66 | Not Include |

### Step4. Final Result

From the screening above, we can know what terms should be included in the final model. Thus, we have our final result as:

## Linear mixed model fit by maximum likelihood ['lmerMod']  
## **Formula: pain ~ 1 + nsaid \* WeatherDate + (1 + WeatherDate | ID)**  
## Data: new\_mcbig  
##   
## AIC BIC logLik deviance df.resid   
## 5210.9 5251.0 -2597.4 5194.9 1105   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -4.3948 -0.4987 0.0007 0.4877 3.5958   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr  
## ID (Intercept) 9.500 3.082   
## WeatherDate 1.083 1.041 0.15  
## Residual 3.111 1.764   
## Number of obs: 1113, groups: ID, 205  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 7.54397 0.50495 14.940  
## nsaid1 -0.09308 0.56346 -0.165  
## WeatherDate -0.28606 0.20903 -1.369  
## nsaid1:WeatherDate -0.58713 0.23419 -2.507  
##   
## Correlation of Fixed Effects:  
## (Intr) nsaid1 WthrDt  
## nsaid1 -0.896   
## WeatherDate 0.129 -0.116   
## nsd1:WthrDt -0.115 0.134 -0.893

The final model should be:

And,

The model has and as random effects and has , and as fixed effects.

After having the equation of the final model, we can have a predictive plot based on the model. Figure 1.1 shows the plot.

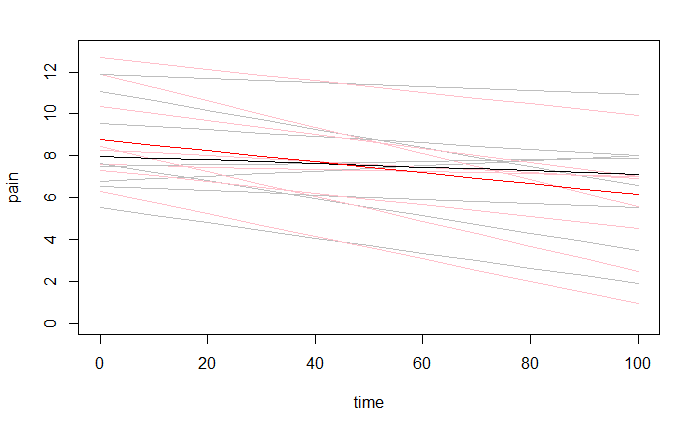


Figure 1.1 Predictive plot of the final model.

In the Figure 1.1, the black line shows how the average pain decreasing with time if NSAID is not taken, the red line shows how the average pain decreasing with time if NSAID is taken. Taking NSAID can result in a quicker speed in pain decreasing. The light red and grey lines show the trend for the first 8 patients in the condition of not having NSAID and having NSAID.

## Problem2

### Step1. Add random effects

First we need to check if the three house level variables(floor, single and basement) should all be included as random effects. We compare the model including each of them and intercepts as random effects with the model having inly intercepts as random effects. Table 2.1, Table 2.2 and Table 2.3 shows the ANOVA result. From the three tables, we can see that the three variables are all significant.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 2.1 ANOVA result of model include intercepts and floor as random effects compared to the model include only intercepts as random effects | | | | |
| Model | Df | AIC | BIC | Pr(>Chisq) |
| m.0 | 3 | 2898.022 | 2913.267 | NA |
| m.floor | 6 | 2757.063 | 2787.553 | 0 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 2.2 ANOVA result of model include intercepts and single as random effects compared to the model include only intercepts as random effects | | | | |
| Model | Df | AIC | BIC | Pr(>Chisq) |
| m.0 | 3 | 2898.022 | 2913.267 | NA |
| m.single | 6 | 2884.505 | 2914.995 | 0.0002137 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 2.3 ANOVA result of model include intercepts and basement as random effects compared to the model include only intercepts as random effects | | | | |
| Model | Df | AIC | BIC | Pr(>Chisq) |
| m.0 | 3 | 2898.022 | 2913.267 | NA |
| m.basement | 10 | 2841.164 | 2891.981 | 0 |

Then we fit a model including all three variables as random effects and call it m.floor.basement.single. Since the m.floor model has the least AIC, we compare m.floor.basement.single with m.floor. The ANOVA result in Table 2.4 indicates that these two models are not significant different, which means the more variables added to the m.floor are not significant actually. Thus, we do not need to include the variables basement and single as random effects in the model. We only include floor and intercepts as random effects in the model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 2.4 ANOVA result of model include intercepts, single, basement and floor as random effects compared to the model include only intercepts and floor | | | | |
| Model | Df | AIC | BIC | Pr(>Chisq) |
| m.floor | 6 | 2757.063 | 2787.553 | NA |
| m.floor.basement.single | 21 | 2763.572 | 2870.288 | 0.0742488 |

## Step2. Add fixed effects

Then, we can add fixed effects to the model. In this case, the fixed effects can only be the variable uranium. Table 2.5 shows that uranium is a significant variable in the model. Thus we need to include it as fixed effects in the model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 2.5 ANOVA result of model include intercepts, uranium and floor as random effects compared to the model include only intercepts and floor as random effects | | | | |
| Model | Df | AIC | BIC | Pr(>Chisq) |
| m.floor | 6 | 2736.447 | 2766.907 | NA |
| m.floor.uranium | 7 | 2683.548 | 2719.085 | 0 |

**Appendix**

**Problem 1**

library(lme4)

library(nlme)

setwd("D:/BROWN/dropbox/Dropbox/PHP2550/assignment3")  
mcbig <- read.csv("McAlindon\_Big.csv")  
vars <- c("age","race2","sex","inccat","retire","nsaid")  
painvar <- paste0("pain.",c(1,2,3,4,5,6,7))  
timevar <- paste0("lastdt",c(1,2,3,4,5,6,7))

#rearrange the data  
id <- unique(mcbig$ID)  
library(dplyr)

new\_mcbig <- c()  
for (k in id) {  
 date <- as.numeric(mcbig[which(mcbig$ID == k),timevar][1,])  
 pain <- as.numeric(mcbig[which(mcbig$ID == k),painvar][1,])  
 newbig <- data.frame(ID = rep(k,7))  
 newbig$WeatherDate <- date  
 newbig$pain <-pain  
 temp <- mcbig[which(mcbig$ID == k),c("ID","WeatherDate","avgtemp",vars)]  
 newbig <- newbig %>% left\_join(temp, by = c("ID", "WeatherDate"))  
 newbig$WeatherDate <- newbig$WeatherDate - newbig$WeatherDate[1]  
 new\_mcbig <- rbind(new\_mcbig,newbig)  
}

#pre-processing  
#remove NAs rows  
new\_mcbig <- new\_mcbig[!is.na(new\_mcbig$WeatherDate) & !is.na(new\_mcbig$avgtemp) &!is.na(new\_mcbig$pain),]  
#since date, temprature and age are not in the same measurement scale, we need to scale the variables  
new\_mcbig$avgtemp <- scale(new\_mcbig$avgtemp)  
new\_mcbig$age <- scale(new\_mcbig$age)  
new\_mcbig$WeatherDate <- scale(new\_mcbig$WeatherDate)

#fit model with only intercpts  
m0 <- lmer(pain ~ 1|ID, REML = FALSE, data = new\_mcbig)  
#time as random effect  
m.date <-lmer(pain ~ 1 + WeatherDate + (1+WeatherDate | ID), REML = FALSE, data = new\_mcbig)  
#temprature as random effect  
m.temp <- lmer(pain ~ 1 + avgtemp + (1+avgtemp | ID), REML = FALSE, data = new\_mcbig)  
#include time and temprature as random effect  
m.date.temp <- lmer(pain ~ 1 + avgtemp + WeatherDate + (1 + avgtemp + WeatherDate | ID), REML = FALSE, data = new\_mcbig)  
knitr::kable(AIC(m0,m.date,m.temp,m.date.temp))

knitr::kable(anova(m.date,m.date.temp))

#check other fixed effects

new\_fix <- c()  
old\_formula <- "+ 1 + WeatherDate + (1 + WeatherDate | ID)"  
m.old <- m.date  
for (var in c("race2","sex","inccat","retire","nsaid")) {  
 new\_mcbig[,var] <- as.factor(new\_mcbig[,var])  
}  
for (var in vars) {  
 temp\_mcbig <- new\_mcbig[!is.na(new\_mcbig[,var]),]  
 formula <- paste0("pain ~", var, old\_formula)  
 formula <- as.formula(formula)  
 m.old <- lmer(pain ~ 1 + WeatherDate + (1+WeatherDate | ID), REML = FALSE, data = temp\_mcbig)  
 m.new <- lmer(formula, REML = FALSE, data = temp\_mcbig)  
 aic <- AIC(m.new)  
 p\_anova <- anova(m.old,m.new)$`Pr(>Chisq)`[2]  
 if (aic < AIC(m.old) & p\_anova < 0.05){  
 result <- "Include"  
 #old\_formula <- paste0("+", var, old\_formula)  
 #m.old <- m.new  
 }else{  
 result <- "Not Include"  
 }  
 model <- paste0("pain ~", var, old\_formula)  
 new\_fix <- rbind(new\_fix,c(model,var,aic,p\_anova,result))  
}  
new\_fix <- data.frame(new\_fix)  
names(new\_fix) <- c("model","Variable","AIC","P value of anova","Include or not")  
knitr::kable(new\_fix)

#check other fixed effects with interactions

new\_inter <- c()  
old\_formula <- "+ 1 + WeatherDate + (1 + WeatherDate | ID)"  
for (var in c("race2","sex","inccat","retire","nsaid")) {  
 new\_mcbig[,var] <- as.factor(new\_mcbig[,var])  
}  
for (var in vars) {  
 temp\_mcbig <- new\_mcbig[!is.na(new\_mcbig[,var]),]  
 formula <- paste0("pain ~", paste0(var,"\*WeatherDate"), old\_formula)  
 formula <- as.formula(formula)  
 m.old <- lmer(pain ~ 1 + WeatherDate + (1+WeatherDate | ID), REML = FALSE, data = temp\_mcbig)  
 m.new <- lmer(formula, REML = FALSE, data = temp\_mcbig)  
 aic <- AIC(m.new)  
 p\_anova <- anova(m.old,m.new)$`Pr(>Chisq)`[2]  
 if (aic < AIC(m.old) & p\_anova < 0.05){  
 result <- "Include"  
 #old\_formula <- paste0("+", var, old\_formula)  
 #m.old <- m.new  
 }else{  
 result <- "Not Include"  
 }  
 model <- paste0("pain ~", paste0(var,"\*WeatherDate"), old\_formula)  
 new\_inter <- rbind(new\_inter,c(model,var,aic,p\_anova,result))  
}  
new\_inter <- data.frame(new\_inter)  
names(new\_inter) <- c("model","Variable","AIC","P value of anova","Include or not")  
knitr::kable(new\_inter)

#final model

m.final <- lmer(pain ~ 1 + nsaid\*WeatherDate + (1+WeatherDate | ID), REML = FALSE, data = new\_mcbig)  
summary(m.final)

#plot

coef\_fix <- fixed.effects(m.final)  
coef\_random <- random.effects(m.final)  
time <- seq(0,100,by= 10)  
time\_s <- scale(seq(0,100,by= 10))  
y0 <- coef\_fix[1] + coef\_fix[3]\*time\_s  
y1 <- coef\_fix[1] + coef\_fix[3]\*time\_s + coef\_fix[2] + coef\_fix[4]\*time\_s  
plot(time, y0, type = "l", col = "black",ylim = c(0,13))  
  
for (i in 1:8) {  
 yi0 <- coef\_fix[1] + coef\_fix[3]\*time\_s + coef\_random$ID[i,1] + coef\_random$ID[i,2]\*time\_s  
 yi1 <- coef\_fix[1] + coef\_fix[3]\*time\_s + coef\_fix[2] + coef\_fix[4]\*time\_s + coef\_random$ID[i,1] + coef\_random$ID[i,2]\*time\_s  
 points(time, yi0, type = "l", col = "grey")  
 points(time, yi1, type = "l", col = "pink")  
}  
points(time, y0, type = "l", col = "black")  
points(time, y1, type = "l", col = "red")

**Problem 2**

#rearrange the data  
srrs <- read.table("srrs2.txt",sep = ",",header = T)  
cty <- read.table("cty.txt",sep = ",",header = T)  
srrs\_col <- c("idnum","state2","stfips","typebldg","floor","basement","activity","cntyfips")  
cty\_col <- c("stfips","ctfips","Uppm")  
srrs\_new <- srrs[srrs\_col]  
cty\_new <- cty[cty\_col]  
names(cty\_new) <- c("stfips","cntyfips","uranium")  
newdata <- srrs\_new %>% left\_join(cty\_new, by = c("stfips","cntyfips"))  
newdata$floor[which(newdata$floor == 2)] <- 1  
newdata$floor[which(newdata$floor == 3)] <- 1

#pre-processing   
mndata <- newdata[which(newdata$state2 == 'MN'),]  
mndata$floor[which(mndata$floor == 9)] <- NA  
mndata$floor <- as.factor(mndata$floor)  
mndata$single <- 0  
mndata$single[which(mndata$typebldg == 1)] <- 1  
mndata$single <- as.factor(mndata$single)  
mndata$basement[which(mndata$basement == 0)] <- NA  
mndata$basement <- as.factor(mndata$basement)  
mndata$activity <- log(mndata$activity)  
mndata$uranium <- log(mndata$uranium)  
mndata <- mndata[which(mndata$activity != -Inf),]

#fit the model  
#check the random effects  
mndata <- mndata[!is.na(mndata$floor) & !is.na(mndata$basement),]  
m.0 <- lmer(activity ~ 1 | cntyfips, REML = FALSE, data = mndata)  
m.floor <- lmer(activity ~ 1 + floor + (1 + floor | cntyfips), REML = FALSE, data = mndata)  
m.single <- lmer(activity ~ 1 + single + (1 + single | cntyfips), REML = FALSE, data = mndata)  
m.basement <- lmer(activity ~ 1 + basement + (1 + basement | cntyfips), REML = FALSE, data = mndata)  
knitr::kable(anova(m.0, m.floor))

knitr::kable(anova(m.0, m.single))

knitr::kable(anova(m.0, m.basement))

m.floor.basement.single <- lmer(activity ~ 1 + basement + single + floor + (1 + basement + single + floor | cntyfips), REML = FALSE, data = mndata)  
#anova(m.0, m.floor.basement)  
knitr::kable(anova(m.floor, m.floor.basement.single))

#add fixed effects  
fix\_mndata <- mndata[!is.na(mndata$uranium),]  
m.floor <- lmer(activity ~ 1 + floor + (1 + floor | cntyfips), REML = FALSE, data = fix\_mndata)  
m.floor.uranium <- lmer(activity ~ 1 + floor + uranium + (1 + floor | cntyfips), REML = FALSE, data = fix\_mndata)  
knitr::kable(anova(m.floor, m.floor.uranium))