Mixed Models

What is a Mixed Model?

- A model with both fixed and random effects (more on what these are in a minute!)
- These are a very broad collection of models. Can be used to analyze a plethora of experiments such as Randomized Complete Block Designs (RCBD), Split Plots Designs, RCBD with sub-sampling, Crossover Designs, and Repeated Measure Designs.

Random Effect Versus Fixed Effect Terminology

- In most experiments, we hypothesize that a (response) variable will vary as we change the value of two or more independent variables. We call these independent variables **factors**. The values of the factor are called the **factor levels** (the values the independent variable can take).
- In experiments, we are either interested in the specific levels of the factor that actually appear in the experiment, or the levels may be a subset of a broader set of factor levels.
- **Fixed Effect** - The factor levels in the experiment are the only levels of interest.
- **Random Effect** - The factor levels are a subset of a broader set of factor levels of potential interest. Here the levels that appear in the experiment were randomly chosen from those broad set.
- **Example** : Suppose we are studying the reaction rate of five specific chemical agents of interest. Due to equipment restraints, we can only analyze five chemicals a day, and we'd like to have five total replicates of each chemical agent.
  - **Factors** - Chemical Agent and Day
  - Here since we are interested in those five specific agents, chemical agents are a fixed effect. Moreover, we are not specifically interested in those five days, so we want to treat day as a random blocking factor.

Implementation in R

- In general, mixed models are best done in SAS's PROC GLIMMIX (in my humble opinion). In R, there is a function for this and function for that, but GLIMMIX allows analysis to be done in a more uniform environemnt with less code. But, we can use the lme function for our analysis today which is in the nlme library.
- **Caution**

1. There are many intricacies and technicalities to consider when fitting a mixed model such as normality of response, is the model fitting appropriately, is a factor fixed or random, degrees of freedom adjustments, post-hoc comparisons, how to model the correlation with measurements on the same subject, etc.
2. Second, and most importantly, a poorly designed experiment can almost never be salvaged by any statistical analysis.

So, before you perform an experiment, I would highly recommend you consult a statistician to make sure there are no issues of confounding, and then obtain advice on how to analyze the experiment. Set up an appointment with me and Sarah; it is free! The nice folks at the Department of Agriculture pay for it!

The only package we need to today is the nlme package, which is included in base R. So, lets go ahead and load it with

```
library(nlme)
```

Data Analysis

- **Example 1** - Suppose we want to study the effect of six fertilization programs on the nitrate content of wheat. Now we would like to make recommendations that apply across the state of Kentucky. So, we should run our experiment at multiple locations since there is natural variability in climate, soil type, and fertility of soil across the state. Assume we select 4 locations at random to run our experiment. Here
  - Response - nitrate content
  - Factors - fertilization program (fixed) and sites (random).

  Now let's fit a mixed model :

```
##Load the nlme package ##
library(nlme)
install.packages('lsmeans')


## I manually inputted the data ##
content <- c(40.89,37.99,37.18,34.98,34.89,42.07,41.22,49.42,45.85,50.15,
41.99,46.69,44.57,52.68,37.61,36.94,46.65,40.23,41.90,39.20,43.29,
40.45,42.91,39.97)

block <- as.factor(sort(rep(1:4,6)))
nitrogen <- as.factor(c(2,5,4,1,6,3,1,3,4,6,5,2,6,3,5,1,2,4,2,4,6,5,3,1))

## Made a data frame called wheat
wheat <- data.frame(content,block,nitrogen)

## Always a good idea to make a boxplot or visualize the data before
## performing a formal analysis

boxplot(content ~ nitrogen,data = wheat)

## Our model ##
## basic syntax resp ~ (FIXED Effect) , Data set , random ~ 1|(RANDOM EFFECT) ##

model <- lme(content ~ nitrogen , data = wheat, random = ~ 1|block, method = 'REML')
anova.lme(model)

## So we see that not all treatments are the same.
We now ask : What treatments do differ? ##
## This is where pairwise contrasts become useful ##

## Compare each treatment to one another to see what treatment differs ##

library(lsmeans)
lsmeans(model,pairwise ~ nitrogen , adjust = 'Tukey')
```

- **Example 2** - Let's consider a model with sub-sampling. **Experiment** We want to see if pesticide residue on cotton plants differ by two standard chemical treatments (call them A and B). We have six batches of plants, each batch from a single field. Three batches were randomly assigned to each treatment. To actually measure the amount of pesticide residue, we select two plants from each batch and measure the residue.
  - Factors - chemical treatments (fixed) and 2 plants from the batch.
  - We call the 2 plants from the batch a **sub-sample**.
  - **Important** The 6 batches of plants from the field are the experimental units (they receive the treatments). It is important to not confuse replication of the experimental unit with sub-sampling.

```
## Cotton Data ##

residue <- c(120,110,120,100,140,130,71,71,70,76,63,68)
trt <- as.factor(sort( rep(c("A","B"),6) ))
batch <- as.factor(sort(rep(1:6,2)))
sample <- as.factor(1:12)

cotton <- data.frame(residue,trt,batch,sample)
View(cotton)

boxplot(residue ~  trt)

out <- lme(residue ~ trt , data = cotton , random = ~ 1|sample(batch) , method = "REML" )
anova.lme(out)
lsmeans(out,pairwise~trt)
```

Example 3 We will look at repeated measures data. Repeated measures allows us to see how factors compare over the length of time. **Warning :** There are many challenges to modeling repeated measures data due to the strong assumptions present in the models. Since we have multiple measurements on the same subject, we now have to model correlations within subject. I would seek help before modeling on your own.
Experiment Suppose we want see how the temperatures of rabbits compare by three different drugs. (This study done was in context of studying phlebetis) Suppose we have 15 total rabbits, 5 get each treatment, and we take observations at 0 ,30 , 60 , and 90 minutes.

```
## Rabbit Data ##

temp <- c(-.3,-.2,1.2,3.1,-.5,2.2,3.3,3.7,-1.1,2.4,2.2,2.7,1.0,1.7,
2.1,2.5,-.3,.8,.6,.9,-1.1,-2.2,.2,.3,
-1.4,-.2,-.5,-.1,-.1,-.1,-.5,-.3,-.2,
.1,-.2,.4,-.1,-.2,.7,-.3,-1.8,.2,.1,.6,-.5,0.0,
1.0,.5,-1.0,-.3,-2.1,.6,.4,.4,-.7,-.3,-.5,.9,-.4,-.3)

treat <- as.factor(sort(rep(c('A','B','C'),20)))
time <- as.factor(rep(c(0,30,60,90),15))
rabbit <- as.factor(sort(rep(1:15,4)))

data <- data.frame(rabbit,temp,time,treat)
```

```
## These are called Profile Plots -- Very useful for longitudinal data analysis##
interaction.plot(time,treat,temp,mean)

##model##

o <- lme(temp ~ treat + time + treat*time, random = ~time|rabbit ,
correlation = corCompSymm(form = ~time|rabbit))
anova.lme(o)
```