# Lab 12 - Random Slopes in Mixed Models

May 8, 2017

In this lab we go through a mixed model analysis where the slope is random. This will cover:

1) Interpretation of random interactions
2) Including random interactions in a mixed model in R
3) Visualizing random effects

**Note**: you will not be tested on random interactions, and will not need to fit them on the final.

## Politeness Data

In this analysis we will be adapting the analysis and data as given in *A very basic tutorial for performing linear mixed effects analyses* (Winter, 2013). In this data, a subject was given a question to ask (called a `scenario`) and was asked to ask the question in a formal and informal setting (called `attitude`). The pitch of their voice was measured in order to understand the effect of `attidue` on pitch. The data has the following structure:

**Response**

`frequency` – average pitch of the subject's question that they ask

**Predictors**

`subject` – the unique subject ID

`gender` – the gender of the subject

`scenario` – the question that was asked by the subject (e.g. asking for a favor)

`attitude` – whether the subject was asked to ask their question formally/politely (`pol`) or informally (`inf`)

This data set is a subset of the full dataset that was used in *The Phonetic Profile of Korean Formality* (Winter and Grawunder 2012).

Run the following code to load the data

```
#note that this code reads the data from an online source
politeness= read.csv("http://www.bodowinter.com/tutorial/politeness_data.csv")
head(politeness)
```

```
##   subject gender scenario attitude frequency
## 1      F1      F        1      pol     213.3
## 2      F1      F        1      inf     204.5
## 3      F1      F        2      pol     285.1
## 4      F1      F        2      inf     259.7
## 5      F1      F        3      pol     203.9
## 6      F1      F        3      inf     286.9
```

```
#our data has an incomplete observation
which(!complete.cases(politeness))
```

```
## [1] 39
```

```r
#we'll get rid of this observation
l12dat = politeness[-39,]

#lastly, let's make sure factor variables are treated as such
l12dat$subject = as.factor(l12dat$subject)
l12dat$gender = as.factor(l12dat$gender)
l12dat$scenario = as.factor(l12dat$scenario)
l12dat$attitude = as.factor(l12dat$attitude)
```

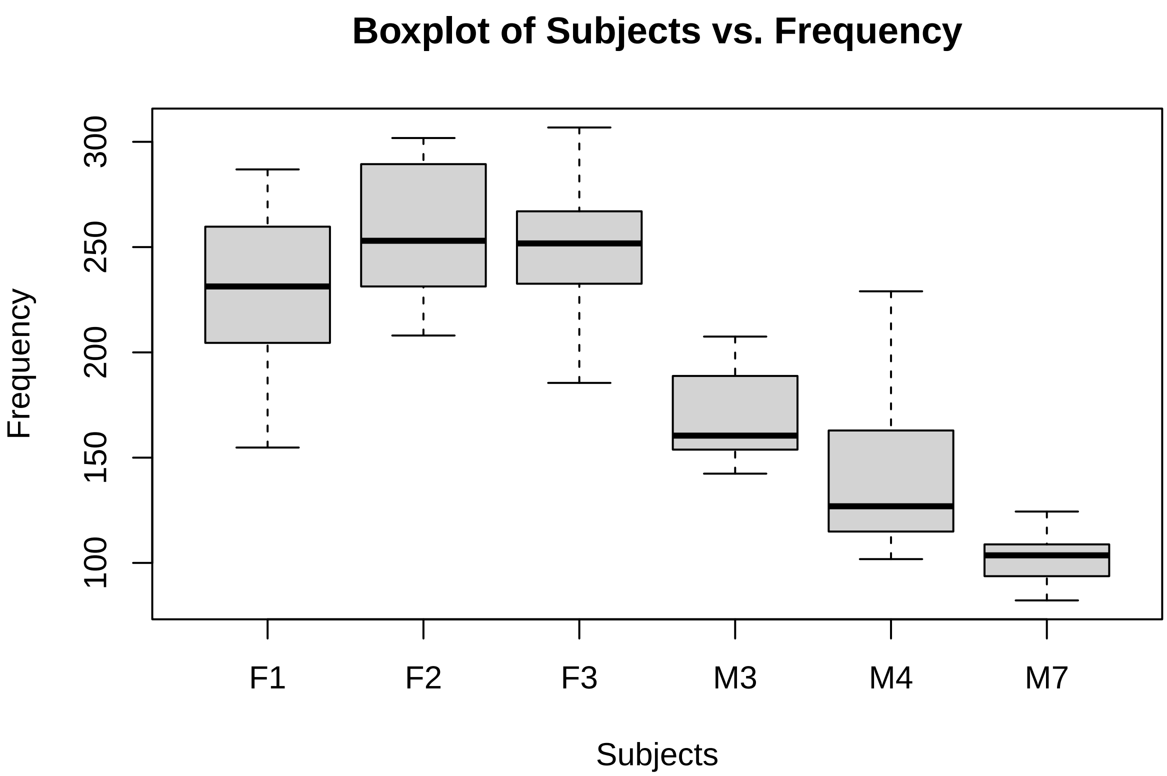
1. If we fit `frequency ~ subject + gender + scenario + attitude`, which predictors should be treated as random effects?

`scenario` could be specific scenarios of interest (which would correspond to a fixed effect), but in the case of this experiment scenarios were selected to be a random selection of questions that a person may ask a professor, or a person may ask their friend. In this sense, we will treat scenario as a random effect.

`subject` will also be treated as a random effect. Even though subjects are independent from eachother, we observe the same subject multiple times, so there is a consistent subject effect. (On the other hand, if a subject was only observed once, then their random effect is captured in the error term.)

2. Take a look at the data (through a plot) to make sure we need to account for subject in order to account for extra variability in our dataset.

```r
boxplot(frequency~subject, data = l12dat, xlab = "Subjects", ylab = "Frequency", main = "Boxplot of Sub
```




The leading "M" or "F" corresponds to subject gender. Note that the M3, M4, M7 observations appear quite different.

3. Write out the mixed model where `subject`, `scenario` are random effects, and `gender`, `attitude` are fixed effects. Fit the model in R.

The model for subject $i$, during scenario $j$ with attitude $k$ is

$frequency_{ijk} = \mu + subject_i + scenario_j + \beta_M * (gender_i == M) + \beta_{pol} * (attitude_k == pol) + \epsilon_{ijk}$

- $\mu$ is the overall mean
- $subject_i$ is the random effect from the $i^{th}$ subject with distribution $N(0, \sigma_s^2)$
- $scenario_j$ is the random effect from the $j^{th}$ scenario with distribution $N(0, \sigma_i^2)$
- $beta_M, gender_i$ is the fixed effect from `gender`, with `F` as the baseline
- $beta_{pol}, attitude_k$ is the fixed effect from `attitude`, with `inf` as the baseline
- $\epsilon_{ijk}$ is the error with distribution $N(0, \sigma_e^2)$

```
library(lme4)

## Loading required package: Matrix

modI = lmer(frequency ~ attitude + gender + (1|subject) + (1|scenario), data=l12dat, REML=FALSE)
```

4. Under this model, does the estimated difference in `frequency` between attitudes change with subject?

No. If we keep all other variables fixed and change `attitude` from `inf` to `pos`, then the estimaed frequency changes by -19.722. This difference does not depend on what the subject was fixed at.

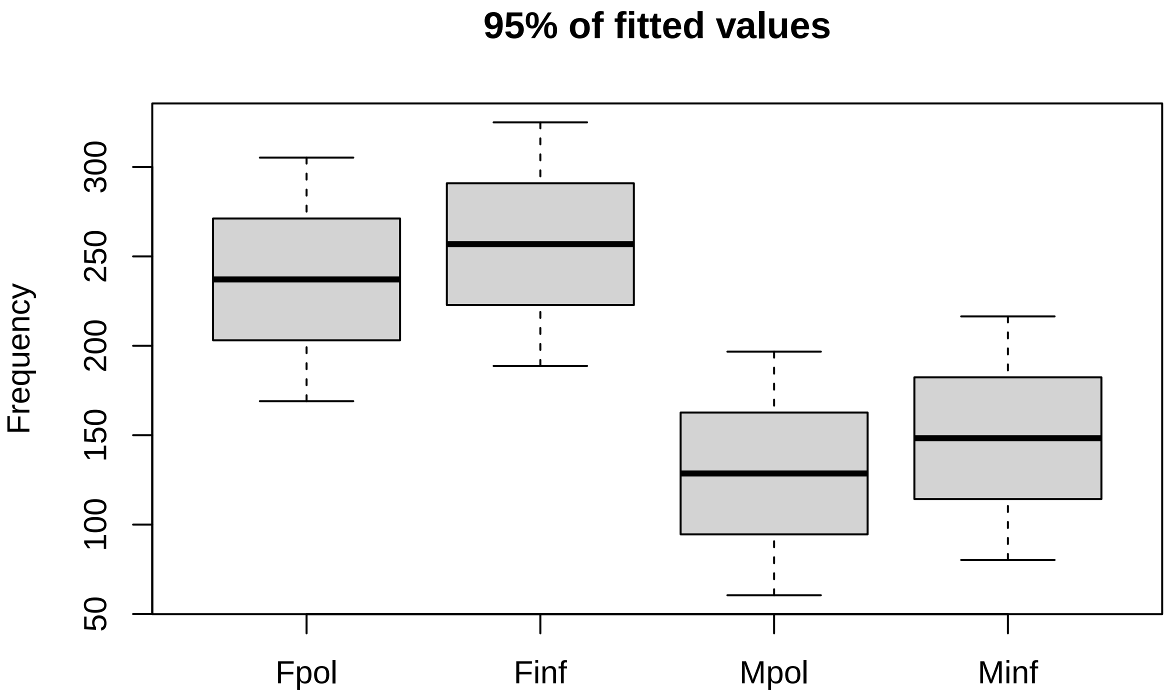
5. Visually show the model fit of this dataset. This is for pedagogical purposes, in general you don't need to plot your model fit.

There are four fixed groups, `F-pol`, `F-inf`, `M-pol`, `M-inf`. The means for this group can be calculated from the fixed effects summary table as 237.125, 256.847, 128.608, 148.33 respectively.

Within each group the fitted values will vary by the estimated variance from the scenario and subject random effects. This corresponds to a standard deviation of $14.33 + 20.42 = 34.75$.

Then, a 95% prediction interval for fitted values corresponds to the following plot

```
Fpol = c(237.125 - 34.75*1.96, 237.125, 237.125 + 34.75*1.96)
Finf = c(256.847 - 34.75*1.96, 256.847, 256.847 + 34.75*1.96)
Mpol = c(128.608 - 34.75*1.96, 128.608, 128.608 + 34.75*1.96)
Minf = c(148.33 - 34.75*1.96, 148.33, 148.33 + 34.75*1.96)
fitDat = data.frame(cbind(Fpol, Finf, Mpol, Minf))
boxplot(fitDat, ylab = "Frequency", main = "95% of fitted values")
```




What if we thought that not only does the frequency change with subject, but the change in frequency between attitudes also changes between subject? We can include that by changing `(1|subject)` to `(1 + attitude|subject)` in the model specification. This is analagous to an interaction term. It is saying, "subject changes the overall mean (the intercept), **and** the attitude effect".

6. Fit the model described above.

```
modR = lmer(frequency ~ attitude + gender + (1+attitude|subject) + (1|scenario), data=l12dat, REML=FALSE
```

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

7. Under this model, does the estimated difference in `frequency` between attitudes change with subject? In what way does it change?

It changes with respect to the `attitudepol` random effect. The estimated difference in `frequency` between attitudes has a mean of 19.715 and a standard deviation of 1.121 for a randomly selected subject.