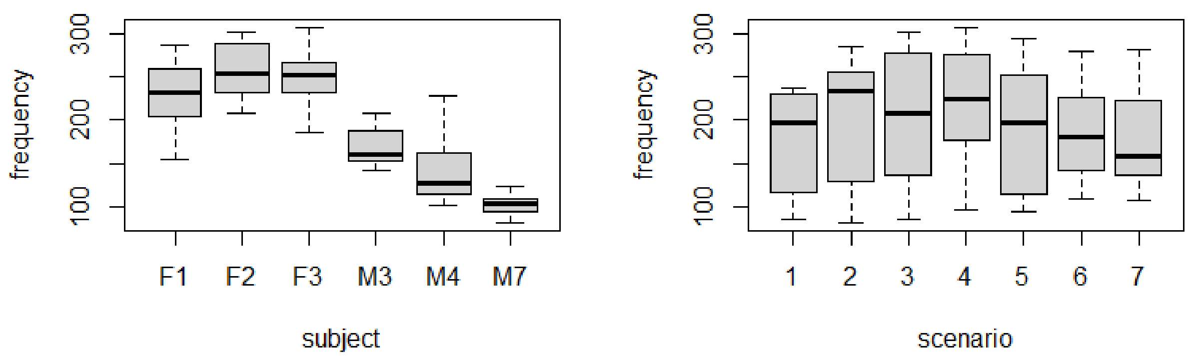
CS614 – Prof. Berardi
Module 3
December 9, 2020

This module will have you run some basic hierarchical linear models that will allow you to demonstrate your knowledge in this area. We will use data from a cognitive science study that aimed to examine the relationship between politeness and the pitch of someone's voice. More specifically, six participants (**subject**) were enrolled, three male and three female (**gender**). Each one was given a phrase to say in seven different situations (**scenario**), e.g. asking for change, ordering a drink. They repeated each scenario twice, once when told to be informal and again when told to be polite (**attitude**). The average pitch of their voice (**frequency**) was measured each time. So we have 6 subjects with 14 observations per subject (7 scenarios*2 attitudes), for a total of 84 observations.

The nature of this data is appropriate for hierarchical linear models. As a demonstration, consider the following boxplots of frequency-by-subject and frequency-by-scenario, which demonstrate that different groupings of data have characteristic frequencies. Therefore, observations cannot be considered independent.



Your goal is to build the following three models and analyze them:

<u>Model 1:</u> Random intercept model with subject as the random effect and gender + attitude as predictor variables.

Model 1 is a basic model that allows each participant to have their own frequency intercept.

<u>Model 2:</u> Random slope model with subject as the random effect, but with attitude as the slope effect. Still have gender + attitude as predictor variables.

Model 2 advances things a bit and allows each participant to have their own frequency intercept as well as their own unique effect for the attitude variable.

<u>Model 3:</u> Random intercept model with subject and scenario as the random effects and gender + attitude as predictor variables. Still have gender + attitude as predictor variables.

Model 3 is similar to Model 1, but it takes advantage of the fact that this is a fully-crossed design (every participant sees every scenario), which allows each participant to have a unique intercept, but also each scenario to also have a unique intercept. We touched on this only briefly in class, so I'll tell you how to specify the fully-crossed random effect for this case: (1|subject) + (1|scenario). Ignore the warning this fit produces.

Your tasks are as follows:

1. Summarize the fixed effects of each model. Don't just recite what the output says. Use plain language to tell me what you've learned by the modeling efforts.

2. For Models 1 and 3, calculate the proportion of total variance that is associated with each random effect. Can you relate your results to the figures provided above? (For extra credit, you can do this for Model 2 by calculating the attitude-dependent variance partitioning coefficient [VPC].)

3. Using AIC and/or ANOVA, determine which model best fits the data. Remember to set REML to false to perform inferential statistics tests.

4. Confirm that random effects are normally distributed by extracting them from the model and plotting a kernel-density estimate (kde) for them. Note that for Model 3 you'll need a 2D kde (`MASS::kde2d`) – leave this one for last.