Statistical Modeling Course

Multi-level Modeling Assignment

In this lab we will use the `musicdata.csv` dataset to develop a deeper understanding of multi-level (mixed effect) models.

Objective: To examine models for predicting the happiness of musicians prior to performances, as measured by the positive affect scale from the PANAS (Positive Affect Negative Affect Schedule) instrument, `pa`.

The dataset contains the following variable

Variables in original data set

- id: unique musician identification number
- diary: cumulative total of diaries filled out by musician
- previous: number of previous diary entries filled out
- perform_type: type of performance (solo, large or small ensemble)
- memory: performed from Memory, using Score, or Unspecified
- audience: who attended (Instructor, Public, Students, or Juried)
- pa: positive affect from PANAS
- na: negative affect from PANAS
- age: musician age
- gender: musician gender
- instrument: Voice, Orchestral, or Piano
- years_study: number of years studied the instrument
- mpqsr: stress reaction subscale from MPQ
- mpqab: absorption subscale from MPQ
- mpqpem: positive emotionality composite scale from MPQ
- mpqnem: negative emotionality composite scale from MPQ
- mpqcon: constraint composite scale from MPQ

```
music <- read.csv("musicdata.csv")
music <- music %>% mutate(solo = ifelse(perform_type == "Solo", 1, 0))
```

Problem 1

In this dataset the group is the musician and the unit is the performance. Classify the predictors into unit-level and group-level.

Problem 2

What is the max, min, and median number of diary entries for the musicians?

Problem 3

Write the equations for the model that predicts positive affect, `pa`, with a random intercept term and no predictors. Clearly define all of the terms. Fit this model. What is the estimated mean positive affect across all diary entries and musicians? Use this model to calculate the intraclass correlation coefficient. Interpret this value.

Problem 4

Building on the model from the previous problem, include audience type (`audience`), performing solo (`solo`) and (`years_study`) in your model as fixed effects. Write the equation for this model. Fit the model and interpret the estimates.

Problem 5

Fit the model in the previous problem but now allow the effect of performing solo to vary by musician (random slopes). Write the equation for this model. What are the estimates for the mean effect of solo and the variance of the effect of solo.

Problem 6

Compare the models from the two previous problems using a likelihood ratio test. Which model is better?

Problem 7

Using the model chosen above, predict the happiness score for the first observation using just the fixed effects by (1) creating the model matrix, (2) obtaining the fixed effect coefficients using `fixedf` (3) multiplying them by the first row of the model matrix you created. Compare this result to the output of the `predict` fuction. Now create a new vector that is the same as the first row of music but with an id = 100. Make a prediction for this observation. Use `predictInterval` in the `merTools` package to get intervals for your two predictions.