Hierarchical Linear Models - HW5

Siqi Wang

December 18, 2015

First, set working directory, read in the data, and set libraries needed.

```
setwd("~/36-763 Applied Multilevel&Hierarchical Models/hw5/hw5")
library(lme4)
library(R2jags)
library(car)
library(arm)
library(ggplot2)
ratings <- read.csv("ratings.csv", header = T)
#str(ratings)
```

There are three main hypothesis.

- Instrument has the largest influence in rating;
- I-V-vi might be frequently rated as classical, and can also have high ratings on popular;
- Contrary motion would be frequently rated as classical.

Start with the exercises.

1. The three main experimental factors

(a) Conventional linear models/analysis of variance

First, build up conventional linear models of instrument, harmony and voice on classical ratings with all three variables included.

```
ols.IHV <- lm(Classical ~ Instrument + Harmony + Voice, data = ratings)
summary(ols.IHV)
```

```
##
## Call:
## lm(formula = Classical ~ Instrument + Harmony + Voice, data = ratings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8718 -1.7137 -0.0297  1.7576 11.4766
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       4.34016    0.12987  33.420  < 2e-16 ***
## Instrumentpiano   1.37359    0.11298  12.158  < 2e-16 ***
## Instrumentstring  3.13312    0.11230  27.899  < 2e-16 ***
```

```
## HarmonyI-V-IV    -0.03108    0.13008  -0.239 0.811168
## HarmonyI-V-VI     0.76909    0.13008   5.913 3.83e-09 ***
## HarmonyIV-I-V     0.05007    0.12997   0.385 0.700092
## Voicepar3rd      -0.41247    0.11271  -3.660 0.000258 ***
## Voicepar5th      -0.37058    0.11264  -3.290 0.001016 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.297 on 2485 degrees of freedom
##   (27 observations deleted due to missingness)
## Multiple R-squared:  0.255,  Adjusted R-squared:  0.2529
## F-statistic: 121.5 on 7 and 2485 DF,  p-value: < 2.2e-16
```

Most of the coefficients are statistically significant, except for some levels in harmony. Now one by one, examine the influence of the three main experimental factors. Start with instrument, the two levels of instrument are both significant, meaning that different instruments do have impact on the ratings. Piano are more probable to be rated as classical than guitar, and string even more probable than piano. To check with the importance of instrument, fit a model without it.

```r
ols.HV <- lm(Classical ~ Harmony + Voice, data = ratings)
#summary(ols.HV)
#List the coefficients in the summary, and drop the rest.
#Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
#(Intercept)    5.84507    0.12896  45.326  < 2e-16 ***
#HarmonyI-V-IV -0.02824    0.14910  -0.189  0.84979
#HarmonyI-V-VI  0.77194    0.14910   5.177 2.43e-07 ***
#HarmonyIV-I-V  0.05249    0.14898   0.352  0.72461
#Voicepar3rd   -0.41065    0.12919  -3.179  0.00150 **
#Voicepar5th   -0.37075    0.12911  -2.872  0.00412 **
```

The coefficients and the significance of harmony and voice did not change much. Then do a partial F test.

```r
anova(ols.IHV, ols.HV)
```

```
## Analysis of Variance Table
##
## Model 1: Classical ~ Instrument + Harmony + Voice
## Model 2: Classical ~ Harmony + Voice
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1   2485 13108
## 2   2487 17235 -2   -4127.6 391.26 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

P-value is far less than 0.05, indicating that instrument is important in classical ratings.

Then do the similar for harmony and voice. In the full model with all three variables, some levels are not significant. For voice, both the two levels are significant. Now fit two new models, one without harmony and another without voice and check the influence of the two variables. Do partial F tests.

```r
ols.IV <- lm(Classical ~ Instrument + Voice, data = ratings)
ols.IH <- lm(Classical ~ Instrument + Harmony, data = ratings)
anova(ols.IHV, ols.IV)
```

```
## Analysis of Variance Table
##
## Model 1: Classical ~ Instrument + Harmony + Voice
## Model 2: Classical ~ Instrument + Voice
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1   2485 13108
## 2   2488 13381 -3   -273.65 17.293 4.107e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(ols.IHV, ols.IH)
```

```
## Analysis of Variance Table
##
## Model 1: Classical ~ Instrument + Harmony + Voice
## Model 2: Classical ~ Instrument + Harmony
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1   2485 13108
## 2   2487 13193 -2    -85.64 8.1181 0.0003061 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With or without any of the three variables, the coefficients of other two variables would not change much. However, from the analysis of variance, we can see that all three variables are important in classical ratings, as all the p-values are less than 0.05. Looking at the full model, the instrument seems to be the most important as with the change , each change of intrument from guitar to piano to string can cause the classical rating to increae about 1.5 points, while the change of harmony or voice can only change the ratings by less than 1. In harmony, particularly, I-V-vi has the largest influence on classical ratings, increasing the rating by 0.77 compared to I-IV-V, while I-V-IV and IV-I-V are not even significant.

For contrary motion, however, as it is the base in the model, we cannot tell itsimportance by coefficients. to test the hypothesis that it would be frequently rates as classic, change the levels of *Voice*. Then refit the model.

```r
ratings$Voice = factor(ratings$Voice, levels = c("par3rd","par5th","contrary"))
ols.IHV1 <- lm(Classical ~ Instrument + Harmony + Voice, data = ratings)
summary(ols.IHV1)
```

```
##
## Call:
## lm(formula = Classical ~ Instrument + Harmony + Voice, data = ratings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8718 -1.7137 -0.0297  1.7576 11.4766
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)        3.92769    0.12992  30.233  < 2e-16 ***
## Instrumentpiano    1.37359    0.11298  12.158  < 2e-16 ***
## Instrumentstring   3.13312    0.11230  27.899  < 2e-16 ***
## HarmonyI-V-IV     -0.03108    0.13008  -0.239 0.811168
## HarmonyI-V-VI      0.76909    0.13008   5.913 3.83e-09 ***
## HarmonyIV-I-V      0.05007    0.12997   0.385 0.700092
## Voicepar5th        0.04189    0.11267   0.372 0.710092
## Voicecontrary      0.41247    0.11271   3.660 0.000258 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.297 on 2485 degrees of freedom
##   (27 observations deleted due to missingness)
## Multiple R-squared:  0.255,  Adjusted R-squared:  0.2529
## F-statistic: 121.5 on 7 and 2485 DF,  p-value: < 2.2e-16
```

So the voice of par5th does not have a significant change in classical ratings compared with par3rd, but the contrary voice does. It significantly increases the chance that people rate a stimuli more classical by 0.412 compared to par3ed, holding instrument and harmony fixed.

(b) Repeated measures

(i) Write the model

$$classical_i = \alpha_{j[i]} + \beta_{11k} Instrument_i + \beta_{12l} Harmony_i + \beta_{13m} Voice_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$
$$\alpha_j = \beta_{20} + \eta_j, \eta_j \sim N(0, \tau^2)$$

Note that for different levels of *instrument*, *harmony* and *voice* there would be different k, l, m.

(ii) Use two different methods to test whether the random intercept is needed.

Fit a lmer model with random intercept for subject.

```r
lmer.intercept <- lmer(Classical ~ Instrument + Harmony + Voice +
                         (1|Subject), data = ratings, REML = F)
display(lmer.intercept)
```

```
## lmer(formula = Classical ~ Instrument + Harmony + Voice + (1 |
##     Subject), data = ratings, REML = F)
##                  coef.est coef.se
## (Intercept)       3.93     0.19
## Instrumentpiano   1.38     0.09
## Instrumentstring  3.13     0.09
## HarmonyI-V-IV    -0.03     0.11
## HarmonyI-V-VI     0.77     0.11
## HarmonyIV-I-V     0.05     0.11
## Voicepar5th       0.04     0.09
## Voicecontrary     0.42     0.09
##
## Error terms:
##  Groups   Name        Std.Dev.
```

```
##  Subject  (Intercept) 1.29
##  Residual             1.89
## ---
## number of obs: 2493, groups: Subject, 70
## AIC = 10468.9, DIC = 10448.9
## deviance = 10448.9
```

```r
fixef(lmer.intercept)
```

```
##      (Intercept)  Instrumentpiano Instrumentstring    HarmonyI-V-IV
##       3.92867713       1.37704509       3.13160679      -0.03250823
##     HarmonyI-V-VI    HarmonyIV-I-V      Voicepar5th    Voicecontrary
##       0.77095828       0.04989468       0.04068211       0.41506472
```

```r
#ranef(lmer.intercept)
as.data.frame.list(ranef(lmer.intercept))[1:15,]
```

```
##  [1] -0.26489601  0.07605034 -1.34018836 -1.65490807 -0.84188215
##  [6] -0.81565551 -1.68113471 -0.89433543 -0.50093580 -0.11303152
## [11]  1.91191532 -0.55338908 -0.52716244  0.04982370  1.64964889
```

The new model does not change the fixed effect of the three variables much, However, the random effect of different subjects differ a lot. For example, the first subject is more inclined to have lower score in classical than average, while the 11th subject rate stimuli 1.9 higher in classical than average. So the model with random intercept for eah subject is better than the OLS model.

Test whether the random intercept is needed in the model, using AIC, BIC and simulation.

- **AIC, BIC**

```r
AIC(ols.IHV, lmer.intercept)
```

```
##                df      AIC
## ols.IHV         9 11230.45
## lmer.intercept 10 10468.86
```

```r
BIC(ols.IHV, lmer.intercept)
```

```
##                df      BIC
## ols.IHV         9 11282.84
## lmer.intercept 10 10527.07
```

Comparing AIC and BIC with the OLS model, the difference in AIC and BIC can be up to about 750, which is significantly larger than 3 (rule of thumb), indicating that the random intercept is needed in the model. Each subject has different classical scores on average.

- **Simulation**

Test whether the random intercept is needed using simulation.

```r
library(RLRsim)
#formula(ols.IHV)
#formula(lmer.intercept)
exactLRT(m0 = ols.IHV1, m = lmer.intercept)
```

```
##
##  simulated finite sample distribution of LRT. (p-value based on
##  10000 simulated values)
##
## data:
## LRT = 763.59, p-value < 2.2e-16
```

The p-value of the test is far smaller than 0.05, so we can reject the null and say that the model with random intercept is better than the OLS model. The random effect is needed and significant.

(iii) Influence of the three experimental factors

First, look at the model with random intercept. The coefficients and their significance are similar to the OLS model, and nothing interesting in the three variables happens in this model. We can obtain similar conclusions from the model except that different subjects have personal biases toward classical ratings.

Start with the model with random intercept, try other models with random intercept but without one of the three variables to examine the influence of them.

```r
lmer.intercept1 <- lmer(Classical ~ Harmony + Voice + (1|Subject),
                        data = ratings, REML = F)
lmer.intercept2 <- lmer(Classical ~ Instrument + Voice + (1|Subject),
                        data = ratings, REML = F)
lmer.intercept3 <- lmer(Classical ~ Instrument + Harmony + (1|Subject),
                        data = ratings, REML = F)
AIC(lmer.intercept, lmer.intercept1, lmer.intercept2, lmer.intercept3)
```

```
##                 df      AIC
## lmer.intercept  10 10468.86
## lmer.intercept1  8 11408.45
## lmer.intercept2  7 10538.79
## lmer.intercept3  8 10489.10
```

```r
BIC(lmer.intercept, lmer.intercept1, lmer.intercept2, lmer.intercept3)
```

```
##                 df      BIC
## lmer.intercept  10 10527.07
## lmer.intercept1  8 11455.02
## lmer.intercept2  7 10579.54
## lmer.intercept3  8 10535.67
```

The AIC and BIC both suggeats that the full model is better. I also did analysis of variance to do partial F test to test whether the three variables are influential, and all the tests show that the full model is significantly better. Therefore, the three main experimental factors are all influential in classical ratings. From the coefficients and standard errors, instrument is the most influential, and in harmony I-V-vi is significantly frequently rated as classical, which is also the case contrary motion voice.

(c) Random effect on instrument, harmony and voice

(i) Fit models and compare

Fit a model with personal bias with the type of instrument, type of harmony and type of voice leading as three random effects. I guess we should consider the possible correlation between these random effects, but using lmer we are not able to do that with a model like this. So I ignore this issue.

```
lmer.type <- lmer(Classical ~ Instrument + Harmony + Voice +
                    (1|Subject:Instrument) + (1|Subject:Harmony) +
                    (1|Subject:Voice), data = ratings, REML = F)
display(lmer.type)
```

```
## lmer(formula = Classical ~ Instrument + Harmony + Voice + (1 |
##     Subject:Instrument) + (1 | Subject:Harmony) + (1 | Subject:Voice),
##     data = ratings, REML = F)
##                  coef.est coef.se
## (Intercept)       3.93     0.21
## Instrumentpiano   1.36     0.26
## Instrumentstring  3.13     0.26
## HarmonyI-V-IV    -0.03     0.14
## HarmonyI-V-VI     0.77     0.14
## HarmonyIV-I-V     0.06     0.14
## Voicepar5th       0.04     0.08
## Voicecontrary     0.41     0.08
##
## Error terms:
##  Groups             Name        Std.Dev.
##  Subject:Harmony    (Intercept) 0.66
##  Subject:Voice      (Intercept) 0.16
##  Subject:Instrument (Intercept) 1.47
##  Residual                       1.56
## ---
## number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210
## AIC = 10057.5, DIC = 10033.5
## deviance = 10033.5
```

```
#ranef(lmer.type)
as.data.frame(ranef(lmer.type)$`Subject:Instrument`)[1:10,]
```

```
##  [1] -0.8338064 -0.4806770  0.6490096 -0.6475938 -0.4468627  1.2924171
##  [7] -2.3003133 -1.4137898  0.3254900 -1.5753964
```

```
AIC(ols.IHV1, lmer.intercept, lmer.type)
```

```
##                df      AIC
## ols.IHV1        9 11230.45
## lmer.intercept 10 10468.86
## lmer.type      12 10057.53
```

```r
BIC(ols.IHV1, lmer.intercept, lmer.type)
```

```
##                 df      BIC
## ols.IHV1         9 11282.84
## lmer.intercept  10 10527.07
## lmer.type       12 10127.38
```

AIC and BIC both suggests that the model with random effect of personal bias with different types of instrument, harmony and voice is better than models in 1(a) and 1(b). Also, comparing the DIC, the DIC of the model in 1(b) is 10448.9, while the DIC of the new model is 10033.5, which is much smaller. So all the evidence supports the new model with random effect of personal bias with different types of instrument, harmony and voice.

(ii) Re-examine the influence of the three variables

Fit 6 models, each with one or two fixed and random effects in the model, with the same fixed effects. Then compare the AIC, BIC of the models, together with the full model.

```r
lmer.ran1 <- lmer(Classical ~  Harmony + Voice+
                    (1|Subject:Harmony) +
                    (1|Subject:Voice), data = ratings, REML = F)
lmer.ran2 <- lmer(Classical ~ Instrument + Voice+
                    (1|Subject:Instrument) +
                    (1|Subject:Voice), data = ratings, REML = F)
lmer.ran3 <- lmer(Classical ~ Instrument + Harmony +
                    (1|Subject:Instrument) + (1|Subject:Harmony),
                    data = ratings, REML = F)
lmer.ran4 <- lmer(Classical ~ Voice+
                    (1|Subject:Voice), data = ratings, REML = F)
lmer.ran5 <- lmer(Classical ~ Instrument +
                    (1|Subject:Instrument), data = ratings, REML = F)
lmer.ran6 <- lmer(Classical ~  Harmony +
                    (1|Subject:Harmony), data = ratings, REML = F)

AIC(lmer.type, lmer.ran1, lmer.ran2, lmer.ran3, lmer.ran4, lmer.ran5, lmer.ran6)
```

```
##           df      AIC
## lmer.type 12 10057.53
## lmer.ran1  9 11606.02
## lmer.ran2  8 10245.03
## lmer.ran3  9 10086.20
## lmer.ran4  5 11654.93
## lmer.ran5  5 10267.98
## lmer.ran6  6 11599.60
```

```r
BIC(lmer.type, lmer.ran1, lmer.ran2, lmer.ran3, lmer.ran4, lmer.ran5, lmer.ran6)
```

```
##           df      BIC
## lmer.type 12 10127.38
## lmer.ran1  9 11658.41
## lmer.ran2  8 10291.60
```

```
## lmer.ran3  9 10138.60
## lmer.ran4  5 11684.03
## lmer.ran5  5 10297.09
## lmer.ran6  6 11634.53
```

The random effect of personal bias of classical ratings on voice is not as important as the other two random effects, suggested by AIC and BIC, as the difference is not as big in AIC and BIC with and without voice as the harmony and instrument. Instrument has the largest effect. Still, the full model with all the three random effects and all three variables as fixed effects are the best.

Look at the coefficients and standard errors in the model with all the three new random effects in it.

The coefficients of the two levels of instrument are significant, and are the biggest among all the coefficients, indicating that instruments have the largest influence on classical ratings among the three. For harmony, only harmony I-V-vi is significant with a positive sign, indicating that it is more likely to be rated classical than I-IV-V. Also, for voice, only the voice contrary is significant with a positive sign, suggesting that contrary motion is more likely to be rated classical than par3rd.

Also, take a look at the estimated variances. For harmony, the standard deviation is 0.66, and that for voice is 0.16 while for instrument is 1.47. Instrument has the biggest variance, meaning that different subjects have bigger variance on classical ratings with different instruments. The next is harmony, but the standard deviation of voice is small, indicating that there might not be a lot of difference in classical ratings with the type of voice. The random effect of person/voice might not be very important in the model, although included. Compared with the residual standard deviation, which is 1.56, we can see that the variance in subject with instrument is close to it. This is further evidence that the random effect of person/instrument is needed and instrument is the most influential in classical ratings among the three variables.

(iii) Write the model

$$classical_i = \alpha_{I[ij]} + \alpha_{H[ik]} + \alpha_{V[il]} + \beta_{11j} Instrument_i + \beta_{12k} Harmony_i + \beta_{13l} Voice_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$

$$\alpha_{I[j]} = \beta_{20j} + \eta_{jh}, \eta_{jh} \sim N(0, \tau_j^2)$$

$$\alpha_{H[k]} = \beta_{20k} + \eta_{km}, \eta_{km} \sim N(0, \tau_m^2)$$

$$\alpha_{V[l]} = \beta_{20l} + \eta_{ln}, \eta_{ln} \sim N(0, \tau_n^2)$$

Note that for different levels of *instrument*, *harmony* and *voice* there would be different j, k, l.

2. Individual Covariates

Begin with the model with 3 random effects in 1(c), which is the best model in questin 1. Look at the individual covariates and do some EDA, and there are a lot of NA's. Note that CollegeMusic and APTheory are binary variables, so they should be turned into factors. For the rest, as they are basically scores and are ordered and do not have gaps or other patterns, I consider them as numbers and can be fit in the model without any transformation.

```
#str(ratings)
ratings$CollegeMusic =as.factor(ratings$CollegeMusic)
ratings$APTheory=as.factor(ratings$APTheory)
table(ratings$KnowRob)
```

```
##
##    0    1    5
## 1836  180  324
```

```r
table(ratings$KnowAxis)
```

```
##
##    0    1    5
## 1800   36  396
```

```r
ratings$KnowRob1 = as.numeric(ratings$KnowRob==5)
ratings$KnowAxis1 = as.numeric(ratings$KnowAxis==5)
```

It is worth mentioning that for the variables KnowRob and KnowAxis, the scores should be between 0 to 5, but there is no 2, 3 and 4 in the dataset, meaning that the students are either very sure they've heard of them or they didn't. These two variables can also be turned into binary variables. Number 5 means heard of and 0 or 1 means didn't or at least not familiar with.

For the two variables X1stInstr and X2ndInstr, most of the subjects (for X1stInstr only 28 out of 70 have data and for X2ndInstr only 9 out of 70 subjects have data) haven't played instruments so there are missing data. As far as I believe, the piano and the guitar are the most common instruments for beginners, and other variables cover the music knowledge of the subject, so I tend to drop these two variables in the data.

One more thing to mention is that the information about concentration on instrument and concentration on notes are already in the variable Instr.minus.Notes, so I will only include the aggregate variable in the model to avoid correlation between covariates. Later I may check whether the three are all needed. The similar situation is with CollegeMusic and Noclass. If number of music class is 0, that means CollegeMusic is also 0. But there are some missing values in Noclass. So I would try them both in the model but may consider it later.

Then begin to fit models. First try the full model with all the variables in consideration. Compare it with the model using ConsInstr and ConsNotes instead of Instr.minus.Notes. AIC and BIC favor the second model.

```r
lmer.full <- lmer(Classical ~ Instrument + Harmony + Voice + Selfdeclare+
                    OMSI+X16.minus.17+Instr.minus.Notes+PachListen+ClsListen+
                    KnowRob1+KnowAxis1+X1990s2000s+X1990s2000s.minus.1960s1970s+
                    CollegeMusic+NoClass+APTheory+Composing+PianoPlay+GuitarPlay+
                    (1|Subject:Instrument) + (1|Subject:Harmony) +
                    (1|Subject:Voice), data = ratings, REML = F)
#display(lmer.full)
lmer.full2 <- lmer(Classical ~ Instrument + Harmony + Voice + Selfdeclare+
                    OMSI+X16.minus.17+ConsInstr+ConsNotes+PachListen+ClsListen+
                    KnowRob1+KnowAxis1+X1990s2000s+X1990s2000s.minus.1960s1970s+
                    CollegeMusic+NoClass+APTheory+Composing+PianoPlay+GuitarPlay+
                    (1|Subject:Instrument) + (1|Subject:Harmony) +
                    (1|Subject:Voice), data = ratings, REML = F)
#display(lmer.full2)
AIC(lmer.full, lmer.full2)
```

```
## Warning in AIC.default(lmer.full, lmer.full2): models are not all fitted to
## the same number of observations
```

```
##            df      AIC
## lmer.full  28 7225.477
## lmer.full2 29 6232.921
```

After looking at the coefficients and their standard deviations, I found that only x16.minus.17, ConsNotes, ClsListen, and PianoDisplay are significant. Only use the variables significant in the two full models to fit a reduced model.

```
lmer.reduced1 <- lmer(Classical ~ Instrument + Harmony + Voice+
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Instrument) + (1|Subject:Harmony) +
                    (1|Subject:Voice), data = ratings, REML = F)
#display(lmer.reduced1)
```

After looking at the coefficients and their standard deviations, I found that ConsNotes is not significant. Compare the fit with the second full model. The signs of the coefficients didn't change. So I'm not really worried about confounders. Drop the ConsNotes and fit again.

```
lmer.reduced2 <- lmer(Classical ~ Instrument + Harmony + Voice+
                    X16.minus.17+ClsListen+PianoPlay+
                    (1|Subject:Instrument) + (1|Subject:Harmony) +
                    (1|Subject:Voice), data = ratings, REML = F)
#display(lmer.reduced2)
```

After looking at the coefficients and their standard deviations, I found that the fit changed, especially for PianoPlay, and the significance of the coefficients also changed. I would include ConsNotes in the model although it is not significant itself. It might be a confounder that may both influence the covariates and the classical ratings. Balancing between lower BIC and the number of covariates in the model, this is my final model.

So the individual covariates I include in the model as fixed effects are:

- X16.minus.17: measure of listener's ability to distinguish classical vs popular music,
- ConsNotes: how concentrate on the notes while listening,
- ClsListen: how much listen to classical music,
- PlayPiano: whether play piano.

(b) Re-examine the random effects

Fit 6 models with one or two random effects in the model, with the same fixed effects. Then compare the AIC, BIC of the models.

```
lmer.raneff1 <- lmer(Classical ~ Instrument + Harmony + Voice +
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Harmony) +
                    (1|Subject:Voice), data = ratings, REML = F)
lmer.raneff2 <- lmer(Classical ~ Instrument + Harmony + Voice +
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Instrument) +
                    (1|Subject:Voice), data = ratings, REML = F)
lmer.raneff3 <- lmer(Classical ~ Instrument + Harmony + Voice +
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Instrument) + (1|Subject:Harmony),
                    data = ratings, REML = F)
```

```r
lmer.raneff4 <- lmer(Classical ~ Instrument + Harmony + Voice +
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Voice), data = ratings, REML = F)
lmer.raneff5 <- lmer(Classical ~ Instrument + Harmony + Voice +
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Instrument), data = ratings, REML = F)
lmer.raneff6 <- lmer(Classical ~ Instrument + Harmony + Voice +
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Harmony), data = ratings, REML = F)
AIC(lmer.reduced1, lmer.raneff1, lmer.raneff2,
    lmer.raneff3, lmer.raneff4, lmer.raneff5, lmer.raneff6)
```

```
##                df      AIC
## lmer.reduced1 16 8550.614
## lmer.raneff1  15 9029.057
## lmer.raneff2  15 8612.401
## lmer.raneff3  15 8549.006
## lmer.raneff4  14 9075.487
## lmer.raneff5  14 8610.401
## lmer.raneff6  14 9027.057
```

```r
BIC(lmer.reduced1, lmer.raneff1, lmer.raneff2,
    lmer.raneff3, lmer.raneff4, lmer.raneff5, lmer.raneff6)
```

```
##                df      BIC
## lmer.reduced1 16 8641.077
## lmer.raneff1  15 9113.867
## lmer.raneff2  15 8697.211
## lmer.raneff3  15 8633.816
## lmer.raneff4  14 9154.642
## lmer.raneff5  14 8689.557
## lmer.raneff6  14 9106.213
```

From AIC, it is hard to tell whether the full model or the model without random effect of personal bias on classical ratings toward voice is better, as the difference is less than 3, which is the rule of thumb. In BIC, however, it is obvious that the model without the random effect of personal bias on classical ratings toward voice is better. Considering that we want a good model to find out and interpret the relationship instead of prediction, I tend to use the model without the random effect of personal bias on classical ratings toward voice. The random effect of subject and voice is not very influential, due to lack of variance in the random effect.

(c) Interpretation

The final model is a model with fixed effects including instrument, harmony, voice, and individual level fixed effects including measure of listener's ability to distinguish classical vs popular music, how concentrate on the notes while listening, how much listen to classical music and whether play piano. And the model has random effects for each person/harmony and person/instrument combination. It is listed below with the coefficients, standard deviations and the estimated variance components.

```r
display(lmer.raneff3)
```

```
## lmer(formula = Classical ~ Instrument + Harmony + Voice + X16.minus.17 +
##     ConsNotes + ClsListen + PianoPlay + (1 | Subject:Instrument) +
##     (1 | Subject:Harmony), data = ratings, REML = F)
##                  coef.est coef.se
## (Intercept)       3.81     0.33
## Instrumentpiano   1.45     0.28
## Instrumentstring  3.21     0.28
## HarmonyI-V-IV    -0.03     0.15
## HarmonyI-V-VI     0.82     0.15
## HarmonyIV-I-V     0.07     0.15
## Voicepar5th       0.02     0.08
## Voicecontrary     0.40     0.08
## X16.minus.17     -0.11     0.04
## ConsNotes        -0.11     0.06
## ClsListen         0.16     0.08
## PianoPlay         0.15     0.08
##
## Error terms:
##  Groups             Name        Std.Dev.
##  Subject:Harmony    (Intercept) 0.61
##  Subject:Instrument (Intercept) 1.44
##  Residual                       1.59
## ---
## number of obs: 2109, groups: Subject:Harmony, 236; Subject:Instrument, 177
## AIC = 8549, DIC = 8519
## deviance = 8519.0
```

All the interpretation below are keeping all the other variables in the model fixed. The coefficient of instrument:piano is 1.45, meaning that a stimuli in piano has 1.45 higher ratings in classical than a stimuli in guitar on average. Similarly, the coefficient of instrument:string is 3.21, meaning that a stimuli in string has 3.21 higher ratings in classical than a stimuli in guitar on average. The coefficient of harmony:I-V-IV is -0.03, but it is not significant, so we don't have enough evidence to say it has difference in classical ratings compared to harmony I-IV-V. It is the same with harmony IV-I-V, which has a coefficient 0.07. But for harmony I-V-VI, the coefficient is significant, meaning that harmony I-V-VI has 0.82 higher ratings in classical than a stimuli with harmony IV-I-V on average. Voice par5th don't have significant different in ratings of classical compared with par3rd, while contrary motion has 0.40 higher ratings in classical than a stimuli with voice par3rd on average.

When the subject's ability to distinguish popolar music and classical music increases by 1 point, he'll have 0.11 less point on classical on average. One unit increase in the level of concentration on the notes while listening decreases the classical ratings by 0.11. So more concentrated on notes when listening, the less ratings on classical would be given on average. A unit increase in how often the subject listens to classical music increases the classical rating by 0.16, meaning that people listening to classical more tend to rate music as classical more. Also, a unit increase in how often the subject plays piano increases the classical rating by 0.15. The standard deviation of the random effect on person/harmony combination is 0.61, and the standard deviation of the random effect on person/instrument combination is 1.44. The standard deviation of the model residuals is 1.59.

3. Musicians vs. Non-musicians

Look at the table of self-declare to decide how to dichotomize it.

```r
table(ratings$Selfdeclare)
```

```
##
##   1   2   3   4   5   6
## 576 936 468 432  72  36
```

The best cutoff to have two groups with similar amount of people is between 2 and 3. So define the people rating themselves 1-2 as non-musicians, and people rating themselves 3-6 as musicians. There are 42 non-musicians and 28 musicians. Add a new indicator variable called "musician".

```r
ratings$musician = as.numeric(ratings$Selfdeclare > 2)
table(ratings$musician)
```

```
##
##    0    1
## 1512 1008
```

Think of the potential influence of musician to the impact of other predictors on the classical ratings. Measure of listener's ability to distinguish classical vs popular music, how concentrate on the notes while listening, how much listen to classical music and whether play piano are not necessarily be influenced by the self-declare as musician, but it's worth to try.

Focus on the interactions of musician and the three main experimental variables. Fit a new full model with all the predictors together with the three interaction terms.

```r
lmer.int.full <- lmer(Classical ~ Instrument + Harmony + Voice+musician+
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Instrument+musician*Harmony+musician*Voice,
                    data = ratings, REML = F)
display(lmer.int.full)
```

```
## lmer(formula = Classical ~ Instrument + Harmony + Voice + musician +
##     X16.minus.17 + ConsNotes + ClsListen + PianoPlay + (1 | Subject:Instrument) +
##     (1 | Subject:Harmony) + musician * Instrument + musician *
##     Harmony + musician * Voice, data = ratings, REML = F)
##                            coef.est coef.se
## (Intercept)                 3.69     0.37
## Instrumentpiano             1.76     0.37
## Instrumentstring            3.61     0.37
## HarmonyI-V-IV              -0.04     0.18
## HarmonyI-V-VI               0.24     0.18
## HarmonyIV-I-V               0.06     0.18
## Voicepar5th                 0.09     0.11
## Voicecontrary               0.43     0.11
## musician                    0.23     0.47
## X16.minus.17               -0.11     0.04
## ConsNotes                  -0.11     0.06
## ClsListen                   0.16     0.08
## PianoPlay                   0.16     0.08
## Instrumentpiano:musician  -0.70     0.56
## Instrumentstring:musician -0.91     0.56
```

```
## HarmonyI-V-IV:musician     0.02     0.27
## HarmonyI-V-VI:musician     1.31     0.27
## HarmonyIV-I-V:musician     0.03     0.27
## Voicepar5th:musician      -0.16     0.17
## Voicecontrary:musician    -0.06     0.17
##
## Error terms:
##  Groups             Name        Std.Dev.
##  Subject:Harmony    (Intercept) 0.51
##  Subject:Instrument (Intercept) 1.45
##  Residual                       1.59
## ---
## number of obs: 2109, groups: Subject:Harmony, 236; Subject:Instrument, 177
## AIC = 8530.4, DIC = 8484.4
## deviance = 8484.4
```

Only the interaction term of harmony I-V-VI and musician is significant. The interaction terms also changed the coefficients of the predictors. Try to fit other models with only one or two interaction terms to examine whether the interaction terms are needed.

```
lmer.int.1 <- lmer(Classical ~ Instrument + Harmony + Voice+musician+
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Instrument,
                    data = ratings, REML = F)
lmer.int.2 <- lmer(Classical ~ Instrument + Harmony + Voice+musician+
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Harmony,
                    data = ratings, REML = F)
lmer.int.3 <- lmer(Classical ~ Instrument + Harmony + Voice+musician+
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Voice,
                    data = ratings, REML = F)
lmer.int.4 <- lmer(Classical ~ Instrument + Harmony + Voice+musician+
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Instrument+musician*Harmony,
                    data = ratings, REML = F)
lmer.int.5 <- lmer(Classical ~ Instrument + Harmony + Voice+musician+
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Instrument+musician*Voice,
                    data = ratings, REML = F)
lmer.int.6 <- lmer(Classical ~ Instrument + Harmony + Voice+musician+
                    X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Harmony+musician*Voice,
                    data = ratings, REML = F)
AIC(lmer.int.full, lmer.int.1, lmer.int.2, lmer.int.3,
    lmer.int.4, lmer.int.5, lmer.int.6, lmer.raneff3)
```

```
##                 df      AIC
```

```
## lmer.int.full 23 8530.363
## lmer.int.1    18 8552.087
## lmer.int.2    19 8526.106
## lmer.int.3    18 8554.033
## lmer.int.4    21 8527.296
## lmer.int.5    20 8555.138
## lmer.int.6    21 8529.173
## lmer.raneff3  15 8549.006
```

```
BIC(lmer.int.full, lmer.int.1, lmer.int.2, lmer.int.3,
    lmer.int.4, lmer.int.5, lmer.int.6, lmer.raneff3)
```

```
##               df      BIC
## lmer.int.full 23 8660.405
## lmer.int.1    18 8653.859
## lmer.int.2    19 8633.532
## lmer.int.3    18 8655.804
## lmer.int.4    21 8646.030
## lmer.int.5    20 8668.218
## lmer.int.6    21 8647.906
## lmer.raneff3  15 8633.816
```

According to AIC and BIC, the model with the only interactions between the dichotomized musician variable and harmony is the best. Look at the coefficients of the model(I didn't list it here, as it's similar with the full model). If the subject self-declare as a musician, he'll be more influenced on harmony I-V-VI. The difference on the influence of harmony I-V-VI on classical ratings between musicians and non-musicians can be as large as 1.31.

Also try interactions of whether a musician and other individual-level predictors.

```
lmer.int.ind <- lmer(Classical ~ Instrument + Harmony + Voice+musician+
                     X16.minus.17+ConsNotes+ClsListen+PianoPlay+
                     (1|Subject:Instrument) + (1|Subject:Harmony)+
                       musician*Harmony+musician* X16.minus.17+
                       musician*ConsNotes+musician*ClsListen+musician*PianoPlay,
                     data = ratings, REML = F)
BIC(lmer.int.ind, lmer.int.2)
```

```
##              df      BIC
## lmer.int.ind 23 8634.792
## lmer.int.2   19 8633.532
```

According to BIC, the individual level interactions are not needed. Therefore, the result we get is that only the interaction of musician and harmony is needed. If the subject self-declare as a musician, he'll be more influenced on harmony I-V-VI, and will give higher classical ratings than non-musicians when the stimuli is I-V-VI.

4. Classical vs. Popular

(a) Comment on the influence of three main experimental variables

Based on the analysis on the classical ratings, start the analysis of popular ratings with the similar model as the model in 1(a), 1(b) and 1(c), and found out that the model in 1(c) is the best, which has the three main experimental variables *Instrument*, *Harmony* and *Voice*, as well as the three random effects of person/instrument, person/harmony and person/voice combinations. As in 1(c), compare different models with or without each of the random effect and that variable. Comparing AIC and BIC, the result shows that all the three variables are needed, together with their random effects. I'll start with this model to do some analysis, including other individual variables and interaction terms. First examine the influence of instrument, harmony and voice on popular ratings.

```
lmer.type.pop <- lmer(Popular ~ Instrument + Harmony + Voice + (1|Subject:Instrument) + (1|Subject:Harmo
#display(lmer.type.pop)
#                  coef.est coef.se
#(Intercept)        6.74     0.21
#Instrumentpiano   -0.95     0.25
#Instrumentstring  -2.61     0.25
#HarmonyI-V-IV     -0.03     0.14
#HarmonyI-V-VI     -0.27     0.14
#HarmonyIV-I-V     -0.19     0.14
#Voicepar5th        0.00     0.08
#Voicecontrary     -0.16     0.08

#Error terms:
# Groups             Name        Std.Dev.
# Subject:Harmony    (Intercept) 0.63
# Subject:Voice      (Intercept) 0.17
# Subject:Instrument (Intercept) 1.40
# Residual                       1.58
```

It is similar as what we found about classical. The instrument has the largest imfluence on popular ratings. Piano tend to be less rated as popular than guitar, and string even less. In harmony, only I-V-VI is significant, and it indicates that the harmony I-V-VI is less likely to be rated as popular than I-IV-V. Voice contrary is the only significant one, meaning that contrary motion is less likely to be rated as popular than par3rd, while par5th doesn't have significant difference in popular ratings compared with par3rd. So the hypothesis that the instrument has largest influence on popular ratings is true. But harmony I-V-VI is less likely to be rated as popular than I-IV-V, which is the contrary of the hypothesis #2.

Comparing the standard deviations of the variance components and the residual variance, we can see that the largest among standard deviation among the three random effects is person/instrument. The standard deviation is 1.4, which is much larger than the rest two, meaning that different subjects have bigger variance on popular ratings with different instruments. The next is harmony, but the standard deviation of voice is small, indicating that there might not be a lot of difference in popular ratings with the type of voice. The random effect of person/voice might not be very important in the model, although included. Compared with the residual standard deviation which is 1.58, we can see that the variance in subject with instrument is close to it. This is further evidence that the random effect of person/instrument is needed and instrument is the most influential in popular ratings among the three variables.

(b) Individial covariates

Similar to the model for classical ratings, to determine which individual covariates to include in the model, start with a full model with all the variables that make sense to me. I choose the same variables as in model lmer.full2. CollegeMusic and Composing are turned into factors, and KnowRob and KnowAxis also turned into binary variables.

```
lmer.full.pop <- lmer(Popular ~ Instrument + Harmony + Voice + Selfdeclare+
                     OMSI+X16.minus.17+ConsInstr+ConsNotes+PachListen+ClsListen+
```

```
                      KnowRob1+KnowAxis1+X1990s2000s+X1990s2000s.minus.1960s1970s+
                      CollegeMusic+NoClass+APTheory+Composing+PianoPlay+GuitarPlay+
                      (1|Subject:Instrument) + (1|Subject:Harmony) +
                      (1|Subject:Voice), data = ratings, REML = F)
#display(lmer.full.pop) to see which variables are significant.
```

The significant variables in the model are x16.minus.17 and ConsNotes, and maybe we can also consider PachListen. Fit a reduced model. Actually I fit several reduced models with other covariates and even less covariates, and this one seems to be the best, given BIC and the number of covariates, and the coefficients and standard deviations.

```
lmer.reduced.pop <- lmer(Popular ~ Instrument + Harmony +
                      Voice+X16.minus.17+ConsNotes+PachListen+
                      (1|Subject:Instrument) + (1|Subject:Harmony) +
                      (1|Subject:Voice), data = ratings, REML = F)
#display(lmer.reduced.pop)
```

The model works well, and is a balance between simple model and low BIC. Use the model as a base to compare it with other models with and without one or two random effects to examine the influence of these random effects.

```
AIC(lmer.reduced.pop, lmer.ran1.pop, lmer.ran2.pop, lmer.ran3.pop, lmer.ran4.pop, lmer.ran5.pop, lmer.ra
```

```
##                  df      AIC
## lmer.reduced.pop 15 8476.993
## lmer.ran1.pop    14 8858.921
## lmer.ran2.pop    14 8535.922
## lmer.ran3.pop    14 8475.909
## lmer.ran4.pop    13 8907.196
## lmer.ran5.pop    13 8533.999
## lmer.ran6.pop    13 8856.995
```

```
BIC(lmer.reduced.pop, lmer.ran1.pop, lmer.ran2.pop, lmer.ran3.pop, lmer.ran4.pop, lmer.ran5.pop, lmer.ra
```

```
##                  df      BIC
## lmer.reduced.pop 15 8561.544
## lmer.ran1.pop    14 8937.836
## lmer.ran2.pop    14 8614.837
## lmer.ran3.pop    14 8554.823
## lmer.ran4.pop    13 8980.474
## lmer.ran5.pop    13 8607.277
## lmer.ran6.pop    13 8930.273
```

I didn't list the models, but the results are discussed below. According to AIC and BIC, the model without the random effect of person/voice is better, which is consistent with what we found before in the analysis of variance components. The random effect of person/voice is dropped from the model. So the final model has these covariates:

Music-level fixed effects, including instrument, harmony, voice, and individual level fixed effects including measure of listener's ability to distinguish classical vs popular music, how concentrate on the notes while listening and how much listen to Pachelbel's Canon. And the model has random effects for each person/harmony and person/instrument combination.

Look at the coefficients and standard deviations.

```r
display(lmer.ran3.pop)
```

```
## lmer(formula = Popular ~ Instrument + Harmony + Voice + X16.minus.17 +
##     ConsNotes + PachListen + (1 | Subject:Instrument) + (1 |
##     Subject:Harmony), data = ratings, REML = F)
##                  coef.est coef.se
## (Intercept)       7.18     0.56
## Instrumentpiano  -0.94     0.27
## Instrumentstring -2.58     0.27
## HarmonyI-V-IV    -0.05     0.15
## HarmonyI-V-VI    -0.28     0.15
## HarmonyIV-I-V    -0.21     0.15
## Voicepar5th      -0.01     0.09
## Voicecontrary    -0.14     0.09
## X16.minus.17      0.08     0.04
## ConsNotes         0.09     0.06
## PachListen       -0.14     0.11
##
## Error terms:
##  Groups             Name        Std.Dev.
##  Subject:Harmony    (Intercept) 0.62
##  Subject:Instrument (Intercept) 1.37
##  Residual                       1.63
## ---
## number of obs: 2073, groups: Subject:Harmony, 232; Subject:Instrument, 174
## AIC = 8475.9, DIC = 8447.9
## deviance = 8447.9
```

All the interpretation below are keeping all the other variables in the model fixed. The coefficient of instrument:piano is -0.94, meaning that a stimuli in piano has 0.94 lower ratings in popular than a stimuli in guitar on average. Similarly, the coefficient of instrument:string is -2.58, meaning that a stimuli in string has 2.58 lower ratings in popular than a stimuli in guitar on average. The coefficient of harmony:I-V-IV is -0.05, but it is not significant, so we don't have enough evidence to say it has difference in popular ratings compared to harmony I-IV-V. It is the same with harmony IV-I-V, which has a coefficient -0.21. But for harmony I-V-VI, the coefficient is about significant, meaning that harmony I-V-VI has 0.28 lower ratings in popular than a stimuli with harmony IV-I-V on average. Voice par5th don't have significant different in ratings of popular compared with par3rd, while contrary motion has 0.14 lower ratings in popular than a stimuli with voice par3rd on average.

When the subject's ability to distinguish popolar music and classical music increases by 1 point, he'll have 0.08 more points on popular on average. One unit increase in the level of concentration on the notes while listening decreases the popular ratings by 0.09. So more concentrated on notes when listening, the less ratings on popular would be given on average. A unit increase in how often the subject listens to Pachelbel's Canon decreases the popular rating by 0.14, meaning that people listening to Pachelbel's Canon more tend to rate music as popular less. The standard deviation of the random effect on person/harmony combination is 0.62, and the standard deviation of the random effect on person/instrument combination is 1.37. The standard deviation of the model residuals is 1.63.

(c) Musicians vs Non-musicians

Focus on the interactions of musician and the three main experimental variables. Fit a new full model with all the predictors together with the three interaction terms.

```r
lmer.pop.full <- lmer(Popular ~ Instrument + Harmony + Voice+musician+
                    X16.minus.17+ConsNotes+PachListen+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Instrument+musician*Harmony+musician*Voice,
                    data = ratings, REML = F)
#display(lmer.pop.full)
#                           coef.est coef.se
#Instrumentpiano:musician    0.36     0.54
#Instrumentstring:musician   0.69     0.54
#HarmonyI-V-IV:musician      0.27     0.29
#HarmonyI-V-VI:musician     -0.79     0.29
#HarmonyIV-I-V:musician      0.11     0.29
#Voicepar5th:musician        0.28     0.18
#Voicecontrary:musician      0.12     0.18
```

The interaction term of harmony I-V-VI and musician is the only significant one. The interaction terms also changed the coefficients of the predictors. Try to fit other models with only one or two interaction terms to examine whether the interaction terms are needed.

```r
lmer.pop.1 <- lmer(Popular ~ Instrument + Harmony + Voice
                      +X16.minus.17+ConsNotes+PachListen+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Instrument,
                    data = ratings, REML = F)
lmer.pop.2 <- lmer(Popular ~ Instrument + Harmony + Voice
                      +X16.minus.17+ConsNotes+PachListen+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Harmony,
                    data = ratings, REML = F)
lmer.pop.3 <- lmer(Popular ~ Instrument + Harmony + Voice
                      +X16.minus.17+ConsNotes+PachListen+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Voice,
                    data = ratings, REML = F)
lmer.pop.4 <- lmer(Popular ~ Instrument + Harmony + Voice
                      +X16.minus.17+ConsNotes+PachListen+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Instrument+musician*Harmony,
                    data = ratings, REML = F)
lmer.pop.5 <- lmer(Popular ~ Instrument + Harmony + Voice
                      +X16.minus.17+ConsNotes+PachListen+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Instrument+musician*Voice,
                    data = ratings, REML = F)
lmer.pop.6 <- lmer(Popular ~ Instrument + Harmony + Voice
                      +X16.minus.17+ConsNotes+PachListen+
                    (1|Subject:Instrument) + (1|Subject:Harmony)+
                      musician*Harmony+musician*Voice,
                    data = ratings, REML = F)
AIC(lmer.pop.full, lmer.pop.1, lmer.pop.2,
    lmer.pop.3, lmer.pop.4, lmer.pop.5, lmer.pop.6, lmer.ran3.pop)
```

```
##                df      AIC
```

```
## lmer.pop.full 22 8470.167
## lmer.pop.1    17 8477.464
## lmer.pop.2    18 8466.266
## lmer.pop.3    17 8476.661
## lmer.pop.4    20 8468.630
## lmer.pop.5    19 8478.995
## lmer.pop.6    20 8467.804
## lmer.ran3.pop 14 8475.909
```

```r
BIC(lmer.pop.full, lmer.pop.1, lmer.pop.2,
    lmer.pop.3, lmer.pop.4, lmer.pop.5, lmer.pop.6, lmer.ran3.pop)
```

```
##               df      BIC
## lmer.pop.full 22 8594.176
## lmer.pop.1    17 8573.288
## lmer.pop.2    18 8567.727
## lmer.pop.3    17 8572.485
## lmer.pop.4    20 8581.365
## lmer.pop.5    19 8586.094
## lmer.pop.6    20 8580.539
## lmer.ran3.pop 14 8554.823
```

According to AIC and BIC, the model with the only interactions between the dichotomized musician variable and harmony is the best. Look at the coefficients of the model(I didn't list it here, as it's similar with the full model). If the subject self-declare as a musician, he'll be more influenced on harmony I-V-VI. The difference on the influence of harmony I-V-VI on popular ratings between musicians and non-musicians can be as large as 0.79. Musicians are less likely to rate a stimuli with I-V-VI as popular compared to non-musicians.

A Summary of Relationship between Ratings of Music and Instrument, Harmony and Voice

Introduction

Listeners' identification of music as "classical" or "popular" can differ with the instrument, harmonic motion and voice leading of the music. To examine the influence of different types of instrument, harmony and voice on ratings, personal bias in ratings should also be taken into account. Hypothesizes to test include that instrument should have a largest influence on ratings, the specific harmony I-V-vi might both be frequently rates as classical and popular, and that contrary motion would be frequently rated as classical. Data are collected through an experiment by Ivan Jimenez and Vincent Rossi from University of Pittsburgh. They presented 36 musical stimuli to 70 listeners, recruited from the population of undergraduates at the University of Pittsburgh, and asked the listeners to rate the music on how classical it sounds and how popular it sounds. The two scales are independent. Personal history concerning musical knowledge and experience are also collected.

Methods

Both linear models and hierarchical models are used to fit the data. Interactions between whether the person self-declared as musician and the musical features are considered. The final model for classical ratings is a hierarchical model with instrument, harmony, voice, and individual level covariates including measure of listener's ability to distinguish classical vs popular music, how concentrate on the notes while listening, how much listen to classical music and whether play piano. And the model has random effects for each person/harmony and person/instrument combination. The final model for popular ratings is a similar hierarchical model except that individual level covariates include measure of listener's ability to distinguish classical vs popular music, how concentrate on the notes while listening, how much listen to Pachelbel's Canon. The statistical package R was used. The analysis of variance, Akaike and Bayesian Information Criteria are used to compare models.

Results

From the two models, a stimuli played with piano has 1.45 higher ratings in classical and 0.94 lower ratings in popular compared with guitar, and a stimuli played with string has 3.21 higher ratings in classical and 2.58 lower ratings in popular compared with guitar on average. We don't have enough evidence to say harmony I-V-IV or IV-I-V have difference in classical ratings or popular ratings compared to harmony I-IV-V. But harmony I-V-VI has 0.82 higher ratings in classical and 0.28 lower ratings in popular than a stimuli with harmony IV-I-V on average. For voice, contrary motion has 0.40 higher ratings in classical and 0.14 lower ratings in popular than a stimuli with voice par3rd on average, while there's no significant difference between voice par3rd and par5th.

When the listener's more able to distinguish popular and classical music, he is less likely to rate the music as classical and more probable to have higher rates on popular. The more concentrated on notes when listening the listener is, the less ratings on classical and popular would be given on average. People who listen to classical music more tend to rate music as classical more. More often the subject plays piano, the higher the classical ratings are. People listening to Pachelbel's Canon more tend to rate music as popular less.

In the model for classical ratings, the standard deviation of the random effect on person/harmony is 0.61, and the t on person/instrument combination is 1.44. The standard deviation of the model residuals is 1.59. In the model for popular ratings, the random effects included are the same, and the corresponding standard deviations are 0.62, 1.37 and 1.63. This is not a standard repeated measures model, as the random effects are based on two categorical variables. Each variance component shows the variance of personal bias on different types of instruments, harmony and voices. Other variance components that might be needed are for each individual with different types of music features, and each type of instrument, harmony, voice on different people. Also, it is reasonable to think that there might be correlation between the variance components.

Discussion

In summary, concerning the hypothesizes, instrument does have the largest influence on rating among the three experimental factors. The specific harmonic I-V-VI is frequently rated as classical, but not popular. And contrary motion voice would also be frequently rated as classical.