

Multilevel models hw05

Yifei Zheng

December 18, 2015

1.(a)

First step is to read in the data and check that everything is OK.

```
library(lme4)

## Loading required package: Matrix

library(arm)

## Loading required package: MASS
##
## arm (Version 1.8-6, built: 2015-7-7)
##
## Working directory is C:/Users/Yifei Zheng/Desktop/Courses/Multilevel
Models/final project

ratings=read.csv("ratings.csv")
str(ratings)

## 'data.frame':    2520 obs. of  28 variables:
##  $ X                : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Subject          : Factor w/ 70 levels "15","16","17",...: 1
1 1 1 1 1 1 1 1 1 ...
##  $ Harmony          : Factor w/ 4 levels "I-IV-V","I-V-IV",...:
1 1 1 1 1 1 1 1 1 2 ...
##  $ Instrument       : Factor w/ 3 levels "guitar","piano",...: 1
1 1 2 2 2 3 3 3 1 ...
##  $ Voice            : Factor w/ 3 levels
"contrary","par3rd",...: 1 2 3 1 2 3 1 2 3 1 ...
##  $ Selfdeclare      : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ OMSI             : int  734 734 734 734 734 734 734 734 734
734 ...
##  $ X16.minus.17     : num  5 5 5 5 5 5 5 5 5 5 ...
##  $ ConsInstr        : num  4.33 4.33 4.33 4.33 4.33 4.33 4.33
4.33 4.33 4.33 ...
##  $ ConsNotes        : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ Instr.minus.Notes : num  -0.67 -0.67 -0.67 -0.67 -0.67 -0.67
-0.67 -0.67 -0.67 -0.67 ...
##  $ PachListen       : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ ClsListen        : int  4 4 4 4 4 4 4 4 4 4 ...
##  $ KnowRob          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ KnowAxis         : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
## $ X1990s2000s : int 5 5 5 5 5 5 5 5 5 5 ...
## $ X1990s2000s.minus.1960s1970s: int 2 2 2 2 2 2 2 2 2 2 ...
## $ CollegeMusic : int 0 0 0 0 0 0 0 0 0 0 ...
## $ NoClass : int 0 0 0 0 0 0 0 0 0 0 ...
## $ APTheory : int 0 0 0 0 0 0 0 0 0 0 ...
## $ Composing : int 4 4 4 4 4 4 4 4 4 4 ...
## $ PianoPlay : int 1 1 1 1 1 1 1 1 1 1 ...
## $ GuitarPlay : int 5 5 5 5 5 5 5 5 5 5 ...
## $ X1stInstr : int 4 4 4 4 4 4 4 4 4 4 ...
## $ X2ndInstr : int NA NA NA NA NA NA NA NA NA NA ...
## $ first12 : Factor w/ 3 levels "guitar","piano",...: 3
3 3 3 3 3 3 3 3 3 ...
## $ Classical : num 3 3 1 3 2 8 10 6 5 1 ...
## $ Popular : num 9 7 8 7 8 3 1 4 5 8 ...

table(ratings$Harmony)

##
## I-IV-V I-V-IV I-V-VI IV-I-V
## 630 630 630 630

table(ratings$Instrument)

##
## guitar piano string
## 840 840 840

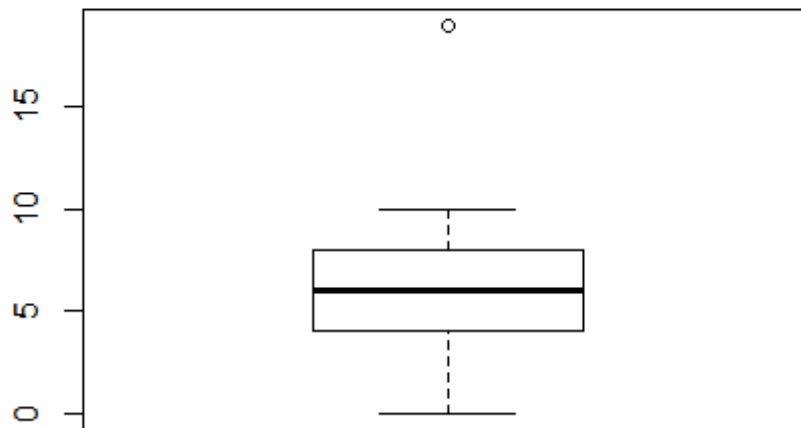
table(ratings$Voice)

##
## contrary par3rd par5th
## 840 840 840

summary(ratings$Classical)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
## 0.000 4.000 6.000 5.783 8.000 19.000 27
```

The maximum value of Classical is 19, which is obviously wrong, since the values of Classical are coded as 1-10. Let's check the boxplot for any more weird points:



It appears that only one observation had its Classical value mismeasured. Let's look at this observation in detail:

```
ratings[which.max(ratings$Classical),]
##           X Subject Harmony Instrument      Voice Selfdeclare OMSI
## 1978 1978      73  IV-I-V      string contrary          3 233
##           X16.minus.17 ConsInstr ConsNotes Instr.minus.Notes PachListen
## 1978           -1          5          5              0          5
##           ClsListen KnowRob KnowAxis X1990s2000s X1990s2000s.minus.1960s1970s
## 1978           3      NA          0          5                                3
##           CollegeMusic NoClass APTheory Composing PianoPlay GuitarPlay
## 1978           1          0          1          1          0          0
##           X1stInstr X2ndInstr first12 Classical Popular
## 1978           4          1    piano          19          6
```

The mismeasurement occurred in observation 1978. Since I have no idea what the original value of Classical was, I will remove this observation before conducting further analyses:

```
ratings=ratings[-1978,]
```

To examine the influence of the three main experimental factors (instrument, harmony and voice), I will first fit a linear model with Classical as the response and all three factors as predictors, and look at the coefficient estimates for the fixed effects to determine how particular levels of these factors affect ratings:

```
fit.1.full=lm(Classical~Harmony+Instrument+Voice,data=ratings)
summary(fit.1.full)
```

```
##
## Call:
## lm(formula = Classical ~ Harmony + Instrument + Voice, data = ratings)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.8673 -1.7092 -0.0252  1.7759  6.0942
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.33556    0.12924   33.547 < 2e-16 ***
## HarmonyI-V-IV  -0.03110    0.12944   -0.240 0.810132
## HarmonyI-V-VI   0.76909    0.12944    5.941 3.22e-09 ***
## HarmonyIV-I-V   0.03163    0.12939    0.244 0.806879
## Instrumentpiano  1.37360    0.11243   12.217 < 2e-16 ***
## Instrumentstring 3.11941    0.11179   27.904 < 2e-16 ***
## Voicepar3rd     -0.39867    0.11219   -3.553 0.000387 ***
## Voicepar5th     -0.35673    0.11212   -3.182 0.001483 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.286 on 2484 degrees of freedom
## (27 observations deleted due to missingness)
## Multiple R-squared:  0.2551, Adjusted R-squared:  0.253
## F-statistic: 121.6 on 7 and 2484 DF, p-value: < 2.2e-16
```

In terms of harmonic motion, using I-V-vi as the benchmark, the motions I-V-IV and IV-I-V do not have significant effects on classical rating since the coefficients are not significant. The motion I-V-VI increases the rating (sounds more like classical music to subjects). In terms of instrument, using electric guitar as the benchmark, it can be seen that both the piano and the string quartet are associated with a higher rating, with the string quartet having more impact than the piano since its coefficient is larger. In terms of voice leading, using contrary motion as the benchmark, the presence of parallel 3rds and 5ths are associated with lower ratings, since their coefficients are negative and significantly different from 0.

To determine which of the three main factors are most important to predicting rating, I fit three separate linear models, each with a factor missing. Then I used ANOVA to compare them with the full model. The magnitude of the F-test statistic is indicative of the importance of the missing factor in predicting rating.

```
fit.1.nHar=lm(Classical~Instrument+Voice,data=ratings)
fit.1.nVoi=lm(Classical~Harmony+Instrument,data=ratings)
fit.1.nIns=lm(Classical~Harmony+Voice,data=ratings)

anova(fit.1.full,fit.1.nHar)

## Analysis of Variance Table
##
## Model 1: Classical ~ Harmony + Instrument + Voice
```

```

## Model 2: Classical ~ Instrument + Voice
##   Res.Df  RSS Df Sum of Sq    F   Pr(>F)
## 1    2484 12975
## 2    2487 13252 -3    -277.16 17.687 2.33e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(fit.1.full,fit.1.nHar)

## Analysis of Variance Table
##
## Model 1: Classical ~ Harmony + Instrument + Voice
## Model 2: Classical ~ Instrument + Voice
##   Res.Df  RSS Df Sum of Sq    F   Pr(>F)
## 1    2484 12975
## 2    2487 13252 -3    -277.16 17.687 2.33e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(fit.1.full,fit.1.nHar)

## Analysis of Variance Table
##
## Model 1: Classical ~ Harmony + Instrument + Voice
## Model 2: Classical ~ Instrument + Voice
##   Res.Df  RSS Df Sum of Sq    F   Pr(>F)
## 1    2484 12975
## 2    2487 13252 -3    -277.16 17.687 2.33e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

It was found that the model missing Instrument had the largest F-test statistic, meaning that the variable instrument is the most important among the three factors.

To sum up, the instrument variable has the largest influence on rating. As for particular levels within the three factors, the harmonic motion I-V-VI, string quartet and contrary motion contribute the most to a higher rating.

- ii. The first way to test if the random effects are needed is to compare the AIC/BIC of the linear model and the random intercept model to see if there is a significantly large difference:

```

fit.ranint=lmer(Classical~Harmony+Instrument+Voice+(1|Subject),data=ratings,R
EML=F)
AIC(fit.1.full)

## [1] 11201.7

AIC(fit.ranint)

## [1] 10434.28

BIC(fit.1.full)

```

```
## [1] 11254.09
```

```
BIC(fit.ranint)
```

```
## [1] 10492.49
```

The AIC and BIC for the hierarchical model are much lower, indicating that the random effects are indeed needed.

Another way to test the random effects:

```
library(RLRsim)
```

```
## Warning: package 'RLRsim' was built under R version 3.2.3
```

```
exactLRT(fit.ranint,fit.1.full)
```

```
## No restrictions on fixed effects. REML-based inference preferable.
```

```
##
```

```
## simulated finite sample distribution of LRT. (p-value based on  
## 10000 simulated values)
```

```
##
```

```
## data:
```

```
## LRT = 769.42, p-value < 2.2e-16
```

The resulting p-value is essentially 0, indicating that the random intercept improves the fit of the original linear model.

iii. Since the random intercept model does not alter the slopes of the predictors, the three factors influences the classical rating in the same way as the linear model.

We already established that the random intercept model in part(b) is superior to the linear model in part(a) in terms of AIC/BIC. Now we compare the AIC/DIC between the model with the three new random effects and the random intercept model:

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

```
## lmer(formula = Classical ~ Harmony + Instrument + Voice + (1 |  
##      Subject:Harmony) + (1 | Subject:Instrument) + (1 | Subject:Voice),  
##      data = ratings, REML = F)  
##               coef.est coef.se  
## (Intercept)      4.34    0.21  
## HarmonyI-V-IV   -0.03    0.14  
## HarmonyI-V-VI    0.77    0.14  
## HarmonyIV-I-V    0.04    0.14  
## Instrumentpiano  1.36    0.26  
## Instrumentstring 3.12    0.26  
## Voicepar3rd      -0.39    0.08  
## Voicepar5th      -0.36    0.08  
##
```

```
## Error terms:
## Groups          Name          Std.Dev.
## Subject:Harmony (Intercept) 0.67
## Subject:Instrument (Intercept) 1.47
## Subject:Voice     (Intercept) 0.14
## Residual                  1.55
## ---
## number of obs: 2492, groups: Subject:Harmony, 280; Subject:Instrument,
210; Subject:Voice, 210
## AIC = 10015.5, DIC = 9991.5
## deviance = 9991.5

display(fit.ranint)

## lmer(formula = Classical ~ Harmony + Instrument + Voice + (1 |
##      Subject), data = ratings, REML = F)
##              coef.est coef.se
## (Intercept)      4.34    0.19
## HarmonyI-V-IV    -0.03    0.11
## HarmonyI-V-VI     0.77    0.11
## HarmonyIV-I-V     0.03    0.11
## Instrumentpiano   1.37    0.09
## Instrumentstring  3.12    0.09
## Voicepar3rd       -0.40    0.09
## Voicepar5th       -0.36    0.09
##
## Error terms:
## Groups      Name          Std.Dev.
## Subject (Intercept) 1.29
## Residual              1.88
## ---
## number of obs: 2492, groups: Subject, 70
## AIC = 10434.3, DIC = 10414.3
## deviance = 10414.3
```

The AIC/DIC for the new model is much lower, so it is a better model than either the linear model in part(a) and the random intercept model in part(b).

Instead of an intercept that varies by subject, we now have a model with an intercept that varies by each combination of subject and particular levels of the three factors. The slope coefficients estimates remain unchanged, and even though the standard errors of the estimates have increased, the significance of the coefficient estimates are the same as before.

We can look at ANOVA's of the three reduced models:

```
fit.newran.nh=lmer(Classical~Instrument+Voice+(1|Subject:Instrument)+(1|Subje
ct:Voice),data=ratings,REML=F)
fit.newran.ni=lmer(Classical~Harmony+Voice+(1|Subject:Harmony)+(1|Subject:Voic
e),data=ratings,REML=F)
fit.newran.nv=lmer(Classical~Harmony+Instrument+(1|Subject:Harmony)+(1|Subjec
```

```

t:Instrument),data=ratings,REML=F)
anova(fit.newran,fit.newran.nh)

## Data: ratings
## Models:
## fit.newran.nh: Classical ~ Instrument + Voice + (1 | Subject:Instrument) +
(1 |
## fit.newran.nh:      Subject:Voice)
## fit.newran: Classical ~ Harmony + Instrument + Voice + (1 |
Subject:Harmony) +
## fit.newran:      (1 | Subject:Instrument) + (1 | Subject:Voice)
##           Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## fit.newran.nh  8 10211 10258 -5097.6  10195.2
## fit.newran    12 10016 10085 -4995.8   9991.5  203.72      4 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(fit.newran,fit.newran.ni)

## Data: ratings
## Models:
## fit.newran.ni: Classical ~ Harmony + Voice + (1 | Subject:Harmony) + (1 |
Subject:Voice)
## fit.newran: Classical ~ Harmony + Instrument + Voice + (1 |
Subject:Harmony) +
## fit.newran:      (1 | Subject:Instrument) + (1 | Subject:Voice)
##           Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## fit.newran.ni  9 11575 11627 -5778.3  11556.7
## fit.newran    12 10016 10085 -4995.8   9991.5  1565.1      3 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(fit.newran,fit.newran.nv)

## Data: ratings
## Models:
## fit.newran.nv: Classical ~ Harmony + Instrument + (1 | Subject:Harmony) +
(1 |
## fit.newran.nv:      Subject:Instrument)
## fit.newran: Classical ~ Harmony + Instrument + Voice + (1 |
Subject:Harmony) +
## fit.newran:      (1 | Subject:Instrument) + (1 | Subject:Voice)
##           Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## fit.newran.nv  9 10042 10095 -5012.2  10024.3
## fit.newran    12 10016 10085 -4995.8   9991.5  32.811      3 3.53e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

It appears that instrument is still the most important factor in predicting classical ratings.

Among the three estimates variance components, the variance between subject and instrument is the largest, which again confirms that instrument has the largest influence on

ratings. The estimated residual variance is larger than any of the three variance components.

- (a) For this part I started out with the model in 1(c), and added in all the variables except Popular (will be used as the response later), subject ID (no use), Instr.minus.Notes (difference between two existing variables), X1stInstr and X2ndInstr (because both have a bunch of NA's), and first12 (told to ignore for this project):

```
fit.indcov=lmer(Classical~Harmony+Instrument+Voice+(1|Subject:Harmony)+(1|Subject:Instrument)+(1|Subject:Voice)+Selfdeclare+OMSI+X16.minus.17+ConsInstr+ConsNotes+PachListen+ClsListen+KnowRob+KnowAxis+X1990s2000s+X1990s2000s.minus.1960s1970s+CollegeMusic+NoClass+APTheory+Composing+PianoPlay+GuitarPlay,data=ratings,REML=F)
summary(fit.indcov)

## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula:
## Classical ~ Harmony + Instrument + Voice + (1 | Subject:Harmony) +
## (1 | Subject:Instrument) + (1 | Subject:Voice) + Selfdeclare +
## OM SI + X16.minus.17 + ConsInstr + ConsNotes + PachListen +
## ClsListen + KnowRob + KnowAxis + X1990s2000s +
## X1990s2000s.minus.1960s1970s +
## CollegeMusic + NoClass + APTheory + Composing + PianoPlay +
## GuitarPlay
## Data: ratings
##
##      AIC      BIC   logLik deviance df.resid
##  6232.9   6387.8  -3087.5   6174.9     1512
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.2461 -0.5671 -0.0030  0.5414  3.3577
##
## Random effects:
## Groups              Name            Variance Std.Dev.
## Subject:Harmony      (Intercept)  0.40860   0.6392
## Subject:Instrument    (Intercept)  1.34682   1.1605
## Subject:Voice         (Intercept)  0.03679   0.1918
## Residual              (Intercept)  2.48564   1.5766
## Number of obs: 1541, groups:
## Subject:Harmony, 172; Subject:Instrument, 129; Subject:Voice, 129
##
## Fixed effects:
##
##              Estimate Std. Error t value
## (Intercept)    2.6090160   1.2071285    2.161
## HarmonyI-V-IV  -0.0045957   0.1786308   -0.026
## HarmonyI-V-VI   0.8499313   0.1786865    4.757
## HarmonyIV-I-V   0.0597787   0.1785675    0.335
## Instrumentpiano  1.6475165   0.2690563    6.123
## Instrumentstring 3.5881585   0.2688637   13.346
```

```
## Voicepar3rd          -0.4035219  0.1067532 -3.780
## Voicepar5th          -0.3000683  0.1067533 -2.811
## Selfdeclare          -0.3621569  0.2520351 -1.437
## OMSI                 0.0004303  0.0011657  0.369
## X16.minus.17         -0.0882065  0.0576946 -1.529
## ConsInstr            0.0434989  0.1183524  0.368
## ConsNotes            -0.2114623  0.1060683 -1.994
## PachListen           0.1997872  0.1756654  1.137
## ClsListen            0.2410052  0.1212157  1.988
## KnowRob              0.0981312  0.0912782  1.075
## KnowAxis             0.0254281  0.0721753  0.352
## X1990s2000s          0.1375925  0.1402054  0.981
## X1990s2000s.minus.1960s1970s 0.0777157  0.1155459  0.673
## CollegeMusic         -0.1953863  0.4047870 -0.483
## NoClass              -0.0681067  0.1380982 -0.493
## APTheory             0.5214648  0.4184693  1.246
## Composing            0.2256237  0.1368813  1.648
## PianoPlay            0.3225209  0.0942471  3.422
## GuitarPlay           -0.1174963  0.1498692 -0.784

##
## Correlation matrix not shown by default, as p = 25 > 20.
## Use print(x, correlation=TRUE) or
## vcov(x) if you need it
```

Since almost none of the covariates added have coefficients significantly different from 0, I tried removing some covariates with very small T-statistics. The result is a reduced model:

```
fit.indcov.reduce=lmer(Classical~Harmony+Instrument+Voice+(1|Subject:Harmony)
+(1|Subject:Instrument)+(1|Subject:Voice)+Selfdeclare+X16.minus.17+ConsNotes+
ClsListen+APTheory+PianoPlay,data=ratings,REML=F)
summary(fit.indcov.reduce)

## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula:
## Classical ~ Harmony + Instrument + Voice + (1 | Subject:Harmony) +
## (1 | Subject:Instrument) + (1 | Subject:Voice) + Selfdeclare +
## X16.minus.17 + ConsNotes + ClsListen + APTheory + PianoPlay
## Data: ratings
##
##      AIC      BIC   logLik deviance df.resid
## 7807.9   7908.0  -3885.9   7771.9     1910
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.2888 -0.5740  0.0068  0.5498  3.3493
##
## Random effects:
##      Groups              Name            Variance Std.Dev.
## Subject:Harmony (Intercept)  0.39490     0.6284
## Subject:Instrument (Intercept) 1.93168     1.3898
```

```

## Subject:Voice      (Intercept) 0.02335  0.1528
## Residual              2.49527  1.5796
## Number of obs: 1928, groups:
## Subject:Harmony, 216; Subject:Instrument, 162; Subject:Voice, 162
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    4.58559    0.39770  11.530
## HarmonyI-V-IV  -0.02775    0.15805   -0.176
## HarmonyI-V-VI   0.83591    0.15809   5.287
## HarmonyIV-I-V   0.05236    0.15806   0.331
## Instrumentpiano  1.50415    0.28185   5.337
## Instrumentstring 3.27177    0.28153  11.621
## Voicepar3rd     -0.39412    0.09297  -4.239
## Voicepar5th     -0.38110    0.09295  -4.100
## Selfdeclare     -0.32031    0.13745  -2.330
## X16.minus.17    -0.07167    0.04433  -1.617
## ConsNotes       -0.06661    0.07201  -0.925
## ClsListen        0.20030    0.09028   2.219
## APTheory         0.63505    0.31996   1.985
## PianoPlay        0.15992    0.08720   1.834
##
## Correlation of Fixed Effects:
##              (Intr) HI-V-I HI-V-V HIV-I- Instrmntp Instrmnts Vcpr3r Vcpr5t
## HrmnyI-V-IV -0.199
## HrmnyI-V-VI -0.199  0.500
## HrmnyIV-I-V -0.199  0.500  0.500
## Instrumntpn -0.355  0.000  0.000  0.000
## Instrmntstr -0.354  0.000  0.000  0.000  0.499
## Voicepar3rd -0.117  0.000  0.001  0.000  0.000  0.000
## Voicepar5th -0.116 -0.001 -0.001 -0.002  0.000  0.000  0.501
## Selfdeclare -0.478  0.000  0.000  0.000  0.001  0.000  0.000  0.000
## X16.mins.17 -0.156  0.000  0.000  0.000  0.000  0.000  0.000  0.000
## ConsNotes   -0.289  0.000  0.000  0.000  0.001  0.000  0.000  0.000
## ClsListen   -0.216  0.000  0.000  0.000  0.001  0.000 -0.001 -0.001
## APTheory     0.044  0.001  0.000  0.000  0.003  0.000  0.000 -0.001
## PianoPlay    0.188  0.000  0.000  0.000 -0.003  0.000  0.000  0.001
##
## Slfdcl X16..1 CnsNts ClsLst APThry
## HrmnyI-V-IV
## HrmnyI-V-VI
## HrmnyIV-I-V
## Instrumntpn
## Instrmntstr
## Voicepar3rd
## Voicepar5th
## Selfdeclare
## X16.mins.17 -0.295
## ConsNotes   -0.272  0.285
## ClsListen   -0.322  0.193  0.077

```

```
## APTheory      -0.129  0.083 -0.113 -0.038
## PianoPlay     -0.304 -0.051 -0.052 -0.151 -0.248
```

It was found that the AIC/BIC increased by a large amount, so I chose to use the "full" model as the final model.

(b) Not sure what to do here.

(c) For variables coded from 0-5, such as PachListen, ClsListen, etc. the coefficient estimate is the expected change in ratings when the variable moves up to the next level. For binary variables, such as College Music, the coefficient is the difference in rating for a subject with and without the variable. For numeric variables, such as OMSI, the coefficient is the change in ratings for each one unit increase in the variable.

3. First we look at the distribution of Selfdeclare:

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

```
##
##  1  2  3  4  5  6
## 576 936 467 432 72 36
```

Since there are 2519 observations in the data, the first two values of Selfdeclare give us a little more than half of the observations (1512), so we can recode the variable to take on the value of 0 when Selfdeclare=1 or 2, and 1 otherwise:

```
ratings$Selfdeclare[ratings$Selfdeclare<3]=0
ratings$Selfdeclare[ratings$Selfdeclare>=3]=1
```

I suspected that people who claimed themselves to be musicians might concentrate more on the notes when listening to the music provided in the experiment, so I included an interaction term between Selfdeclare and ConsNotes, and refit the model:

```
fit.indcov.new=lmer(Classical~Harmony+Instrument+Voice+(1|Subject:Harmony)+(1|Subject:Instrument)+(1|Subject:Voice)+Selfdeclare+ConsInstr+OMSI+X16.minus.17+Selfdeclare:ConsNotes+ConsNotes+PachListen+ClsListen+KnowRob+KnowAxis+X1990s2000s+X1990s2000s.minus.1960s1970s+CollegeMusic+NoClass+APTheory+Composing+PianoPlay+GuitarPlay,data=ratings,REML=F)
summary(fit.indcov.new)
```

```
## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula:
## Classical ~ Harmony + Instrument + Voice + (1 | Subject:Harmony) +
## (1 | Subject:Instrument) + (1 | Subject:Voice) + Selfdeclare +
## ConsInstr + OMSI + X16.minus.17 + Selfdeclare:ConsNotes +
## ConsNotes + PachListen + ClsListen + KnowRob + KnowAxis +
## X1990s2000s + X1990s2000s.minus.1960s1970s + CollegeMusic +
## NoClass + APTheory + Composing + PianoPlay + GuitarPlay
## Data: ratings
##
##      AIC      BIC    logLik deviance df.resid
## 6231.8   6392.0   -3085.9   6171.8     1511
```

```

##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.2392 -0.5770 -0.0052  0.5351  3.3606
##
## Random effects:
##      Groups             Name             Variance Std.Dev.
## Subject:Harmony      (Intercept)  0.40517   0.6365
## Subject:Instrument  (Intercept)  1.31354   1.1461
## Subject:Voice        (Intercept)  0.03589   0.1894
## Residual                                2.48667   1.5769
## Number of obs: 1541, groups:
## Subject:Harmony, 172; Subject:Instrument, 129; Subject:Voice, 129
##
## Fixed effects:
##
##              Estimate Std. Error t value
## (Intercept)      3.199984    1.244417    2.571
## HarmonyI-V-IV     -0.004623    0.178199   -0.026
## HarmonyI-V-VI      0.849887    0.178255    4.768
## HarmonyIV-I-V      0.059773    0.178136    0.336
## Instrumentpiano    1.647323    0.266171    6.189
## Instrumentstring    3.588139    0.265977   13.490
## Voicepar3rd       -0.403544    0.106575   -3.786
## Voicepar5th       -0.300072    0.106575   -2.816
## Selfdeclare       -1.004643    0.814661   -1.233
## ConsInstr          0.057880    0.117467    0.493
## OMSI               -0.000808    0.001135   -0.712
## X16.minus.17       -0.133027    0.058474   -2.275
## ConsNotes          -0.423078    0.118786   -3.562
## PachListen         0.115741    0.179543    0.645
## ClsListen          0.032353    0.118988    0.272
## KnowRob            0.080944    0.092026    0.880
## KnowAxis           0.084423    0.077734    1.086
## X1990s2000s        0.152403    0.143413    1.063
## X1990s2000s.minus.1960s1970s -0.052858    0.108629   -0.487
## CollegeMusic       -0.080067    0.423118   -0.189
## NoClass            -0.176511    0.147226   -1.199
## APTheory           0.857422    0.433877    1.976
## Composing          0.360919    0.157337    2.294
## PianoPlay          0.307439    0.095002    3.236
## GuitarPlay        -0.146634    0.153983   -0.952
## Selfdeclare:ConsNotes 0.401852    0.190814    2.106
##
## Correlation matrix not shown by default, as p = 26 > 20.
## Use print(x, correlation=TRUE) or
## vcov(x) if you need it

```

The coefficient for interaction term is significant, so the secondary hypothesis of the researchers that people who self-identify as musicians are indeed influenced by different things than non-musicians.

Two hierarchical models

$$y_i = \alpha_{j[i]} + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$$

$$\alpha_j = \beta_0 + \eta_j, \eta_j \sim N(0, \tau^2)$$

$$y_i = \alpha_{j_1[i]} + \alpha_{j_2[i]} + \alpha_{j_3[i]} + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$$

$$\alpha_{j_1} = \beta_{01} + \eta_{j_1}, \eta_{j_1} \sim N(0, \tau_1^2)$$

$$\alpha_{j_2} = \beta_{02} + \eta_{j_2}, \eta_{j_2} \sim N(0, \tau_2^2)$$

$$\alpha_{j_3} = \beta_{03} + \eta_{j_3}, \eta_{j_3} \sim N(0, \tau_3^2)$$