1 a 9/9
b 4/9
c 7/9

2 a 4/9
b 0/9
c 4/9

3 4/9

4 a 9/9
b 0/9
c 4/9

5 7/10

Total 52/100

HW 5

Tiffany Chen

December 16, 2015

1. a)

```
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 3.1.3
```

```
## Loading required package: Matrix
```

```
## Warning: package 'Matrix' was built under R version 3.1.2
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
library(lattice)
library(arm)
```

```
## Warning: package 'arm' was built under R version 3.1.3
```

```
## Loading required package: MASS
##
## arm (Version 1.8-6, built: 2015-7-7)
##
## Working directory is /Users/Tiffany/Documents/MSP/Multilevel Model/HW5
```

```
setwd("/Users/Tiffany/Documents/MSP/Multilevel Model/HW5")
data = read.csv('ratings.csv',header=TRUE)
ratings = data[which(!is.na(data$Classical)),]
ratings = data[which(!is.na(data$Popular)),]
head(ratings)
```

```
##   X Subject Harmony Instrument    Voice Selfdeclare OMSI X16.minus.17
## 1 1      15  I-IV-V     guitar contrary           5  734            5
## 2 2      15  I-IV-V     guitar   par3rd           5  734            5
## 3 3      15  I-IV-V     guitar   par5th           5  734            5
## 4 4      15  I-IV-V      piano contrary           5  734            5
## 5 5      15  I-IV-V      piano   par3rd           5  734            5
## 6 6      15  I-IV-V      piano   par5th           5  734            5
##   ConsInstr ConsNotes Instr.minus.Notes PachListen ClsListen KnowRob
## 1      4.33         5             -0.67          5         4       0
## 2      4.33         5             -0.67          5         4       0
## 3      4.33         5             -0.67          5         4       0
## 4      4.33         5             -0.67          5         4       0
## 5      4.33         5             -0.67          5         4       0
```

```
## 6      4.33         5             -0.67          5         4       0
##   KnowAxis X1990s2000s X1990s2000s.minus.1960s1970s CollegeMusic NoClass
## 1        0           5                            2            0       0
## 2        0           5                            2            0       0
## 3        0           5                            2            0       0
## 4        0           5                            2            0       0
## 5        0           5                            2            0       0
## 6        0           5                            2            0       0
##   APTheory Composing PianoPlay GuitarPlay X1stInstr X2ndInstr first12
## 1        0         4         1          5         4        NA  string
## 2        0         4         1          5         4        NA  string
## 3        0         4         1          5         4        NA  string
## 4        0         4         1          5         4        NA  string
## 5        0         4         1          5         4        NA  string
## 6        0         4         1          5         4        NA  string
##   Classical Popular
## 1         3       9
## 2         3       7
## 3         1       8
## 4         3       7
## 5         2       8
## 6         8       3
```

```
summary(ratings$Classical)
```

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

```
summary(ratings$Popular)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   4.000   5.000   5.381   7.000  19.000
```

```
lmfit1=lm(Classical~Instrument+Harmony+Voice,data=ratings)
summary(lmfit1)
```

```
##
## 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
## Multiple R-squared:  0.255,  Adjusted R-squared:  0.2529
## F-statistic: 121.5 on 7 and 2485 DF,  p-value: < 2.2e-16
```

```
lmfit2=lm(Classical~Harmony+Voice,data=ratings)
summary(lmfit2)
```

```
##
## Call:
## lm(formula = Classical ~ Harmony + Voice, data = ratings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6170 -2.2463  0.1549  2.1549 13.1024
##
## 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 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.632 on 2487 degrees of freedom
## Multiple R-squared:  0.02044,    Adjusted R-squared:  0.01847
## F-statistic: 10.38 on 5 and 2487 DF,  p-value: 7.201e-10
```

```
lmfit3=lm(Classical~Instrument+Voice,data=ratings)
summary(lmfit3)
```

```
##
## Call:
## lm(formula = Classical ~ Instrument + Voice, data = ratings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3011 -1.5407 -0.1233  1.7433 11.3299
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         4.5367     0.1038  43.724  < 2e-16 ***
## Instrumentpiano     1.3730     0.1141  12.035  < 2e-16 ***
## Instrumentstring    3.1334     0.1134  27.631  < 2e-16 ***
## Voicepar3rd        -0.4134     0.1138  -3.633 0.000286 ***
## Voicepar5th        -0.3690     0.1137  -3.244 0.001193 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.319 on 2488 degrees of freedom
## Multiple R-squared:  0.2395, Adjusted R-squared:  0.2383
## F-statistic: 195.9 on 4 and 2488 DF,  p-value: < 2.2e-16
```

```
lmfit4=lm(Classical~Instrument+Harmony,data=ratings)
summary(lmfit4)
```

```
##
## Call:
## lm(formula = Classical ~ Instrument + Harmony, data = ratings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9812 -1.8483 -0.0797  1.7874 11.7370
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       4.07966    0.11280  36.166  < 2e-16 ***
## Instrumentpiano   1.37327    0.11330  12.120  < 2e-16 ***
## Instrumentstring  3.13294    0.11262  27.818  < 2e-16 ***
## HarmonyI-V-IV    -0.03234    0.13045  -0.248    0.804
## HarmonyI-V-VI     0.76863    0.13045   5.892 4.33e-09 ***
## HarmonyIV-I-V     0.05045    0.13034   0.387    0.699
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.303 on 2487 degrees of freedom
## Multiple R-squared:  0.2502, Adjusted R-squared:  0.2487
## F-statistic: 165.9 on 5 and 2487 DF,  p-value: < 2.2e-16
```

```
lmerfit1=lmer(Classical~Instrument+Harmony+Voice+(1|Subject), data=ratings)
summary(lmerfit1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
##    Data: ratings
##
## REML criterion at convergence: 10471.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9808 -0.6389 -0.0178  0.6409  5.4802
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Subject  (Intercept) 1.702    1.305
##  Residual             3.581    1.892
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
```

```
##                  Estimate Std. Error t value
## (Intercept)       4.34374    0.18914   22.97
## Instrumentpiano   1.37705    0.09318   14.78
## Instrumentstring  3.13161    0.09257   33.83
## HarmonyI-V-IV    -0.03251    0.10718   -0.30
## HarmonyI-V-VI     0.77096    0.10718    7.19
## HarmonyIV-I-V     0.04989    0.10709    0.47
## Voicepar3rd      -0.41507    0.09287   -4.47
## Voicepar5th      -0.37439    0.09281   -4.03
##
## Correlation of Fixed Effects:
##             (Intr) Instrmntp Instrmnts HI-V-I HI-V-V HIV-I- Vcpr3r
## Instrumntpn -0.244
## Instrmntstr -0.245  0.498
## HrmnyI-V-IV -0.282  0.001    -0.001
## HrmnyI-V-VI -0.282  0.001    -0.001     0.499
## HrmnyIV-I-V -0.283 -0.001    -0.001     0.499  0.499
## Voicepar3rd -0.245 -0.001    -0.001    -0.002  0.001  0.002
## Voicepar5th -0.244 -0.001     0.000    -0.002 -0.003 -0.001  0.500
```

Instrument are the most important factor on classical ratings. When we remove Instrument from the linear regression, the adjusted R-square decreases sharply. Harmony and Voice are less important compared to the other two factors in determing classical ratings. When we remove Harmony or Voice from the linear regression model, the adjusted R-square only decreaes by a little bit. Among Instrument, whether the instrument is piano and whether the instrument is string instrument are both important since they have large fixed effects and the fixed effects are also statistically significant. The most important Harmony factor is HarmonyI-V-VI as it has the largest fixed effect and it is also statistically significant.

9

b) i)

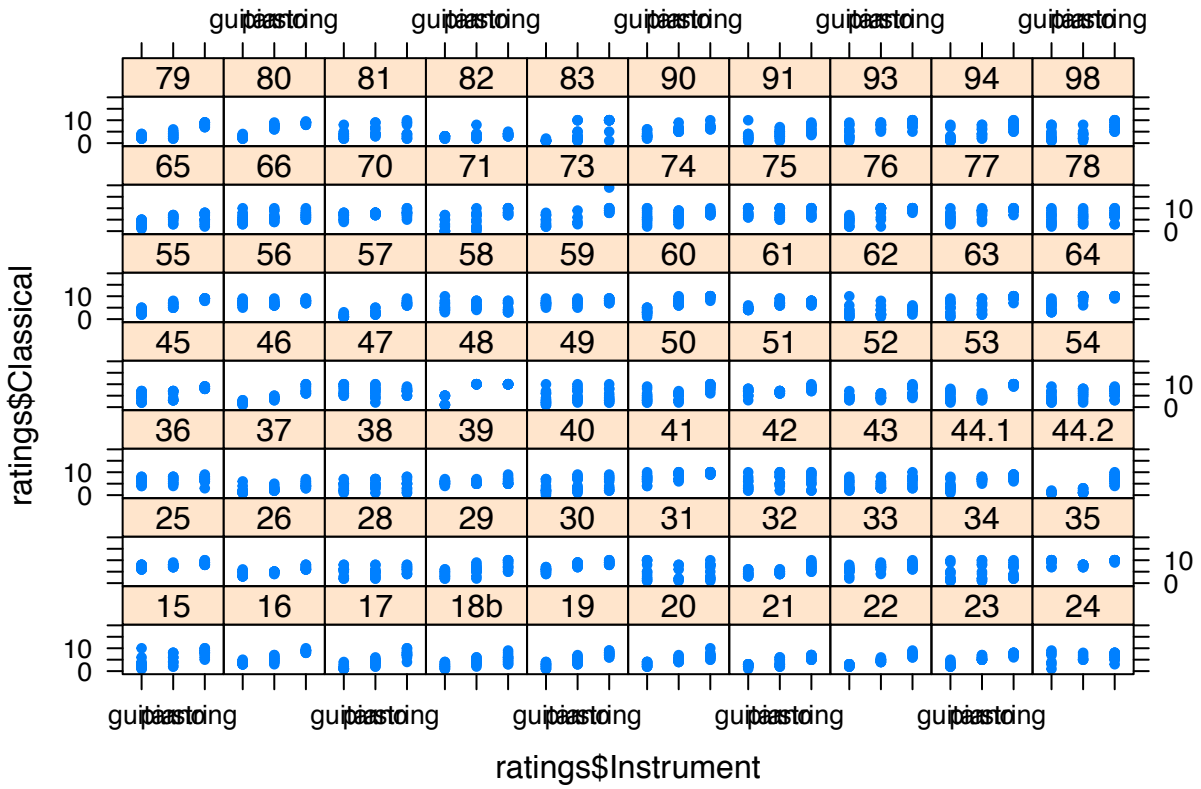
$Classical_n \sim \mathrm{N}(\beta_0 + \beta_{Instrument[i]} + \beta_{Voice[v]} + \beta_{Harmony[h]} + \alpha_{Subject[j]}, \sigma^2)$

$\alpha_{subject[j]} \sim \mathrm{N}(0, \tau^2)$

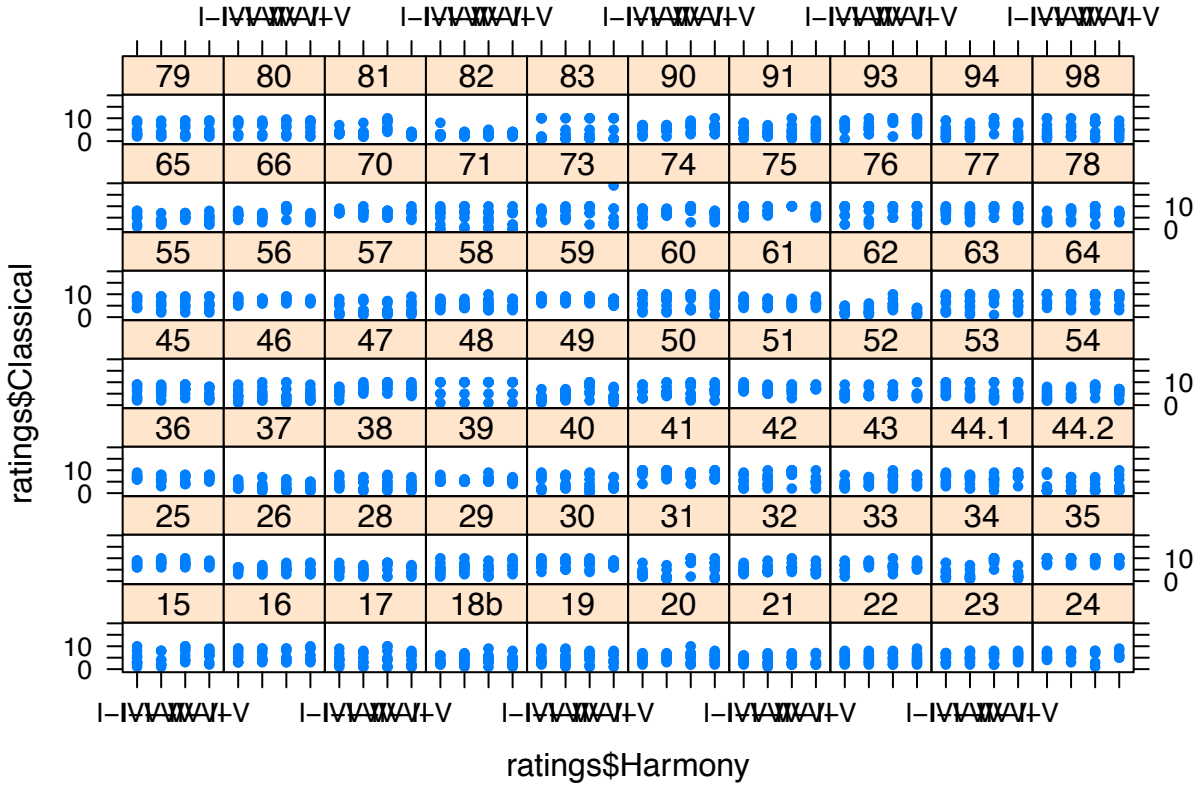
right idea, but indexing problematic. need to indicate obs number as well. If k is the observation number, then the subscripts sohudl be j[k], i[k], etc.

b) ii)

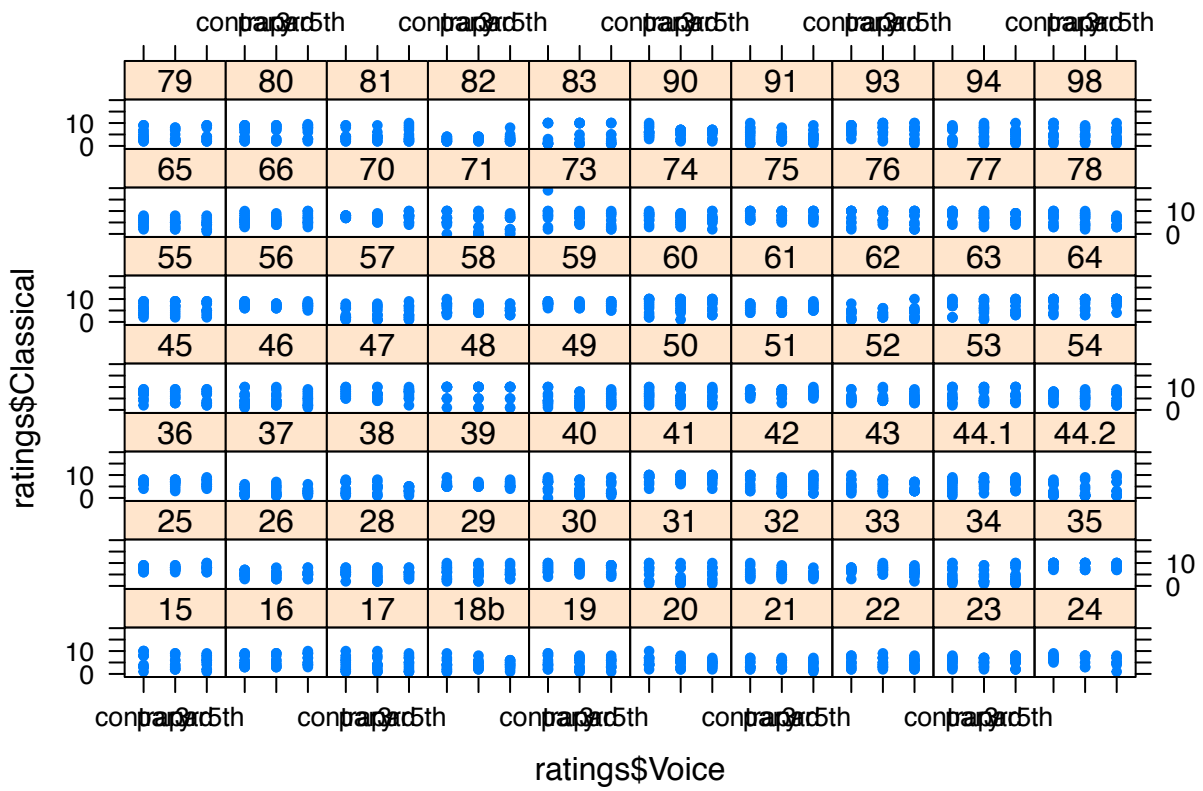
```
xyplot(ratings$Classical~ratings$Instrument|ratings$Subject, pch= 20)
```




```
xyplot(ratings$Classical~ratings$Harmony|ratings$Subject, pch= 20)
```



```
xyplot(ratings$Classical~ratings$Voice|ratings$Subject,pch=20)
```




```
AIC(lmerfit1)
```

```
## [1] 10491.51
```

```
AIC(lmfit1)
```

```
## [1] 11230.45
```

Method 1 - using AIC: The AIC of the random intercept model is much smaller than the conventional linear regression model. Therefore the random intercept model is necessary.

```
anova(lmerfit1, lmfit1)
```

```
## refitting model(s) with ML (instead of REML)

## Data: ratings
## Models:
## lmfit1: Classical ~ Instrument + Harmony + Voice
## lmerfit1: Classical ~ Instrument + Harmony + Voice + (1 | Subject)
##          Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmfit1    9 11230 11283 -5606.2    11212
## lmerfit1 10 10469 10527 -5224.4    10449 763.59      1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the anova analysis, the random intercept model is also necessary. Not a separate, valid method.

b) iii) Instrument is still the most important factor in this case since the coefficients on different levels of Instrument are statistically significant. The Voice and Harmony however are not as important since some β on levels of Voice and Harmony are not statistically significant. show me or refer to appx material that would show me.

c) i)

```
lmerfit2=lmer(Classical~Instrument+Harmony+Voice+(1|Subject:Instrument)+(1|Subject:Harmony)+(1|Subject:
summary(lmerfit2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Classical ~ Instrument + Harmony + Voice + (1 | Subject:Instrument) +
##     (1 | Subject:Harmony) + (1 | Subject:Voice)
##    Data: ratings
##
## REML criterion at convergence: 10051.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.3942 -0.5683 -0.0013  0.5446  5.7495
##
## Random effects:
##  Groups              Name        Variance Std.Dev.
##  Subject:Harmony     (Intercept) 0.44307  0.6656
##  Subject:Voice       (Intercept) 0.02809  0.1676
##  Subject:Instrument (Intercept) 2.19848  1.4827
##  Residual                        2.43753  1.5613
## Number of obs: 2493, groups:
## Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       4.34106    0.21435  20.252
## Instrumentpiano   1.36384    0.26232   5.199
## Instrumentstring  3.12836    0.26203  11.939
## HarmonyI-V-IV    -0.03023    0.14317  -0.211
## HarmonyI-V-VI     0.77063    0.14316   5.383
## HarmonyIV-I-V     0.05618    0.14310   0.393
## Voicepar3rd      -0.40699    0.08174  -4.979
## Voicepar5th      -0.37084    0.08168  -4.540
##
## Correlation of Fixed Effects:
##             (Intr) Instrmntp Instrmnts HI-V-I HI-V-V HIV-I- Vcpr3r
## Instrumntpn -0.611
## Instrmntstr -0.611  0.500
## HrmnyI-V-IV -0.333  0.000     0.000
## HrmnyI-V-VI -0.333  0.000     0.000     0.499
## HrmnyIV-I-V -0.333  0.000     0.000     0.500  0.500
## Voicepar3rd -0.190 -0.001     0.000    -0.002  0.001  0.002
## Voicepar5th -0.190 -0.001     0.000    -0.001 -0.002 -0.001  0.500
```

```
AIC(lmerfit2)
```

```
## [1] 10075.51
```

The AIC of the model with all three new random effect terms is much smaller than the AIC of the model in 1a) and 1b). Furthermore, if we consider the significant of the coefficients of the models, more βs are statistically significant in this new model. So this model is better.

ii) The coefficients on different levels of Instrument and different levels of Voice are all statistically significant. However, two of the coefficients on Harmony are not statistically significant. In this case, Instrument and Voice are more important in Classical ratings compared to Harmony. again, show me.

7

iii

$Classical_n \sim \mathrm{N}(\beta_0 + \beta_{Instrument[i]} + \beta_{Voice[v]} + \beta_{Harmony[h]} + \alpha_{Subject[j]}, \sigma_y^2)$

$\alpha_{subject[j]} \sim \mathrm{N}(0, \tau^2)$

$\beta_{Instrument[i]} \sim \mathrm{N}(0, \sigma^2_{Instrument})$

$\beta_{Harmony[i]} \sim \mathrm{N}(0, \sigma^2_{Harmony})$

$\beta_{Voice[i]} \sim \mathrm{N}(0, \sigma^2_{Voice})$

similar comments to previous model... need to be more careful in indexing.

2.

a) If we include the covariates of all the three variables (Instrument, Harmony and Voice), out of 70 coefficients of alpha's, 34 of them are significant. The intercept is significant, all levels of Instrument and one level of Harmony is also statistically instrument. However, no level in Voice is statistically significant. The Rhat value are all 1 or very near to 1, which shows that the model is good.

4

what model did you start with? what fixed effects are you keeping? etc.

0 b) retest rand effects in model after variable selection in 2(a)

c) α are random effects for subjects. And β_0, $\beta_{Instrument}$, $\beta_{Harmony}$, β_{Voice} are all fixed effects. α are the effect of different people on the popular score given a piece of music. They are mostly around zero. β_0 is the overall intercept.$\beta_{Instrument}$ are the effect of each unique kind of instrument in the music on the popular score, when that piece of music is rated by people. $\beta_{Harmony}$ are the effect of each unique kind of harmony motion in the music on the popular score, when that piece of music is rated by people. β_{Voice} are the effect of each unique kind of voice leading in the music on the popular score, when that piece of music is rated by people. β_0 is 6.56, and most of the $\beta_{Instrument}$, β_{Voice} and $\beta_{Harmony}$'s are small compared to β_0 and are around zero.

4

how are the particular coefficients in your fitted model to be interpreted?

3.

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

you did not dichotomize selfdeclare. Also, you did not consider it as a main effect, or consider any of its interactions with the fixed effects selected in 2(a).

```
##
##   1   2   3   4   5   6
## 564 935 460 426  72  36
```

4

```
lmfit5=lm(Classical~Instrument+Voice+Harmony+Instrument:Selfdeclare+Voice:Selfdeclare+Harmony:Selfdecla
lmfit5both=step(lmfit5, direction="both",
                scope=list(lower = ~ 1,
                           upper = ~Instrument+Voice+Harmony+Instrument:Selfdeclare+
                             Voice:Selfdeclare+Harmony:Selfdeclare+Selfdeclare,
                           data=ratings),trace=1)
```

```
## Start:  AIC=4120.04
## Classical ~ Instrument + Voice + Harmony + Instrument:Selfdeclare +
##     Voice:Selfdeclare + Harmony:Selfdeclare + Selfdeclare
##
##                           Df Sum of Sq   RSS    AIC
## - Voice:Selfdeclare        2     0.317 12850 4116.1
## - Instrument:Selfdeclare   2    14.467 12864 4118.8
## <none>                                 12849 4120.0
## - Harmony:Selfdeclare      3   207.442 13057 4154.0
##
## Step:  AIC=4116.1
## Classical ~ Instrument + Voice + Harmony + Selfdeclare + Instrument:Selfdeclare +
##     Harmony:Selfdeclare
##
##                           Df Sum of Sq   RSS    AIC
## - Instrument:Selfdeclare   2    14.474 12864 4114.9
## <none>                                 12850 4116.1
## + Voice:Selfdeclare        2     0.317 12849 4120.0
## - Voice                    2    85.438 12935 4128.6
## - Harmony:Selfdeclare      3   207.510 13057 4150.0
##
## Step:  AIC=4114.91
## Classical ~ Instrument + Voice + Harmony + Selfdeclare + Harmony:Selfdeclare
##
##                           Df Sum of Sq   RSS    AIC
## <none>                                 12864 4114.9
## + Instrument:Selfdeclare   2      14.5 12850 4116.1
## + Voice:Selfdeclare        2       0.3 12864 4118.8
## - Voice                    2      85.4 12950 4127.4
## - Harmony:Selfdeclare      3     207.7 13072 4148.8
## - Instrument               2    4124.4 16989 4804.2
```

```
summary(lmfit5both)
```

```
##
## Call:
## lm(formula = Classical ~ Instrument + Voice + Harmony + Selfdeclare +
##     Harmony:Selfdeclare, data = ratings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0492 -1.6500 -0.0038  1.6746 11.5834
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                5.05926    0.22859  22.132  < 2e-16 ***
## Instrumentpiano            1.37171    0.11202  12.245  < 2e-16 ***
## Instrumentstring           3.13184    0.11134  28.128  < 2e-16 ***
## Voicepar3rd               -0.41200    0.11174  -3.687 0.000232 ***
## Voicepar5th               -0.36966    0.11168  -3.310 0.000946 ***
## HarmonyI-V-IV             -0.16685    0.29709  -0.562 0.574444
## HarmonyI-V-VI             -0.72782    0.29688  -2.452 0.014293 *
## HarmonyIV-I-V             -0.20057    0.29674  -0.676 0.499169
## Selfdeclare               -0.29409    0.07728  -3.806 0.000145 ***
```

```
## HarmonyI-V-IV:Selfdeclare  0.05578    0.10950   0.509 0.610532
## HarmonyI-V-VI:Selfdeclare  0.61261    0.10944   5.598 2.41e-08 ***
## HarmonyIV-I-V:Selfdeclare  0.10277    0.10935   0.940 0.347433
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.277 on 2481 degrees of freedom
## Multiple R-squared:  0.2689, Adjusted R-squared:  0.2656
## F-statistic: 82.94 on 11 and 2481 DF,  p-value: < 2.2e-16
```

A conventional linear regression model is used to regress Classical on all the three factors as well as their interactions with self-declare status. And using model selection, we can see that the model selection only keeps the interaction term between Harmony and Selfdeclare. And only the interaction term between one level of Harmony and Selfdeclare is statistically significant.

4.

```
lmfit1pop=lm(Popular~Instrument+Harmony+Voice,data=ratings)
summary(lmfit1pop)
```

```
##
## Call:
## lm(formula = Popular ~ Instrument + Harmony + Voice, data = ratings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7218 -1.7026  0.2008  1.4691 13.2248
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        6.58263    0.12761  51.583   <2e-16 ***
## Instrumentpiano   -0.95200    0.11102  -8.575   <2e-16 ***
## Instrumentstring -2.61173    0.11035 -23.667   <2e-16 ***
## HarmonyI-V-IV     -0.02405    0.12782  -0.188   0.8508
## HarmonyI-V-VI     -0.26829    0.12782  -2.099   0.0359 *
## HarmonyIV-I-V     -0.18564    0.12772  -1.454   0.1462
## Voicepar3rd        0.16859    0.11075   1.522   0.1281
## Voicepar5th        0.16326    0.11068   1.475   0.1403
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.257 on 2485 degrees of freedom
## Multiple R-squared:  0.1901, Adjusted R-squared:  0.1878
## F-statistic: 83.32 on 7 and 2485 DF,  p-value: < 2.2e-16
```

```
lmfit2pop=lm(Popular~Harmony+Voice,data=ratings)
summary(lmfit2pop)
```

```
##
## Call:
## lm(formula = Popular ~ Harmony + Voice, data = ratings)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.5588 -1.5588 -0.1207  1.7157 13.4679
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.39169    0.12261  43.974   <2e-16 ***
## HarmonyI-V-IV -0.02673    0.14176  -0.189    0.850
## HarmonyI-V-VI -0.27098    0.14176  -1.912    0.056 .
## HarmonyIV-I-V -0.18755    0.14165  -1.324    0.186
## Voicepar3rd    0.16715    0.12283   1.361    0.174
## Voicepar5th    0.16355    0.12275   1.332    0.183
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.503 on 2487 degrees of freedom
## Multiple R-squared:  0.002983,   Adjusted R-squared:  0.0009782
## F-statistic: 1.488 on 5 and 2487 DF,  p-value: 0.1903
```

```
lmfit3pop=lm(Popular~Instrument+Voice,data=ratings)
summary(lmfit3pop)
```

```
##
## Call:
## lm(formula = Popular ~ Instrument + Voice, data = ratings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6259 -1.6326  0.1487  1.4888 13.3195
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         6.4632     0.1010  63.970   <2e-16 ***
## Instrumentpiano    -0.9520     0.1111  -8.570   <2e-16 ***
## Instrumentstring   -2.6119     0.1104 -23.654   <2e-16 ***
## Voicepar3rd         0.1694     0.1108   1.528    0.127
## Voicepar5th         0.1627     0.1108   1.469    0.142
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.258 on 2488 degrees of freedom
## Multiple R-squared:  0.1881, Adjusted R-squared:  0.1868
## F-statistic: 144.1 on 4 and 2488 DF,  p-value: < 2.2e-16
```

```
lmfit4pop=lm(Popular~Instrument+Harmony,data=ratings)
summary(lmfit4)
```

```
##
## Call:
## lm(formula = Classical ~ Instrument + Harmony, data = ratings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -6.9812 -1.8483 -0.0797  1.7874 11.7370
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       4.07966    0.11280  36.166  < 2e-16 ***
## Instrumentpiano   1.37327    0.11330  12.120  < 2e-16 ***
## Instrumentstring  3.13294    0.11262  27.818  < 2e-16 ***
## HarmonyI-V-IV    -0.03234    0.13045  -0.248    0.804
## HarmonyI-V-VI     0.76863    0.13045   5.892 4.33e-09 ***
## HarmonyIV-I-V     0.05045    0.13034   0.387    0.699
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.303 on 2487 degrees of freedom
## Multiple R-squared:  0.2502, Adjusted R-squared:  0.2487
## F-statistic: 165.9 on 5 and 2487 DF,  p-value: < 2.2e-16
```

```
lmerfit1pop=lmer(Popular~Instrument+Harmony+Voice+(1|Subject), data=ratings)
summary(lmerfit1pop)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Popular ~ Instrument + Harmony + Voice + (1 | Subject)
##    Data: ratings
##
## REML criterion at convergence: 10433.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7257 -0.6433  0.0396  0.6617  5.6143
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Subject  (Intercept) 1.566    1.251
##  Residual             3.532    1.879
## Number of obs: 2493, groups:  Subject, 70
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)       6.57768    0.18347   35.85
## Instrumentpiano  -0.94545    0.09255  -10.22
## Instrumentstring -2.60669    0.09194  -28.35
## HarmonyI-V-IV    -0.02495    0.10646   -0.23
## HarmonyI-V-VI    -0.27228    0.10645   -2.56
## HarmonyIV-I-V    -0.18616    0.10636   -1.75
## Voicepar3rd       0.17050    0.09224    1.85
## Voicepar5th       0.16517    0.09218    1.79
##
## Correlation of Fixed Effects:
##             (Intr) Instrmntp Instrmnts HI-V-I HI-V-V HIV-I- Vcpr3r
## Instrumntpn -0.250
## Instrmntstr -0.251  0.498
## HrmnyI-V-IV -0.289  0.001    -0.001
## HrmnyI-V-VI -0.289  0.001    -0.001     0.499
## HrmnyIV-I-V -0.289 -0.001    -0.001     0.499  0.499
```

```
## Voicepar3rd -0.251 -0.001    -0.001    -0.002  0.001  0.002
## Voicepar5th -0.250 -0.001     0.000    -0.002 -0.003 -0.001  0.500
```

Using conventional linear regression, we can see that Instrument are the most important factor in Popular ratings. In the full model where we include all the three factors, only Instrument are statistically significant and one of the Harmony level is statististically significant. Voice is not statistically significant at all. Moreover, when we remove Instrument from the linear regression, Harmony and Voice are not statistically significant at all and the adjusted R square is near 0. If we look at the varying intercept model, all the coefficients on Instrument are statistically significant and only one coefficient on Harmony is statistically significant. So we can say that Instrument is the most important factor in determing ratings on Popular.

b) All the α are random effects for subjects. And β_0, $\beta_{Instrument}$, $\beta_{Harmony}$, β_{Voice} are all fixed effects. α are the effect of different people on the popular score given a piece of music. They are mostly around zero. β_0 is the overall intercept.$\beta_{Instrument}$ are the effect of each unique kind of instrument in the music on the popular score, when that piece of music is rated by people. $\beta_{Harmony}$ are the effect of each unique kind of harmony motion in the music on the popular score, when that piece of music is rated by people. β_{Voice} are the effect of each unique kind of voice leading in the music on the popular score, when that piece of music is rated by people. β_0 is 6.56, and most of the $\beta_{Instrument}$, β_{Voice} and $\beta_{Harmony}$'s are small compared to β_0 and are around zero.

no interpretatior of the particular estimated coefficients in your model

c)

```
lmfit5pop=lm(Popular~Instrument+Voice+Harmony+Instrument:Selfdeclare+Voice:Selfdeclare+Harmony:Selfdecla
lmfit5popboth=step(lmfit5pop, direction="both",
                   scope=list(lower = ~ 1,
                              upper = ~Instrument+Voice+Harmony+Instrument:Selfdeclare+
                                Voice:Selfdeclare+Harmony:Selfdeclare+Selfdeclare,
                              data=ratings),trace=1)
```

```
## Start:  AIC=4007.13
## Popular ~ Instrument + Voice + Harmony + Instrument:Selfdeclare +
##     Voice:Selfdeclare + Harmony:Selfdeclare + Selfdeclare
##
##                           Df Sum of Sq   RSS    AIC
## - Voice:Selfdeclare        2     1.411 12282 4003.4
## - Instrument:Selfdeclare   2     5.015 12285 4004.1
## <none>                                 12280 4007.1
## - Harmony:Selfdeclare      3   126.073 12406 4026.6
##
## Step:  AIC=4003.42
## Popular ~ Instrument + Voice + Harmony + Selfdeclare + Instrument:Selfdeclare +
##     Harmony:Selfdeclare
##
##                           Df Sum of Sq   RSS    AIC
## - Instrument:Selfdeclare   2     5.023 12287 4000.4
## - Voice                    2    15.130 12297 4002.5
## <none>                                 12282 4003.4
## + Voice:Selfdeclare        2     1.411 12280 4007.1
## - Harmony:Selfdeclare      3   126.061 12408 4022.9
##
## Step:  AIC=4000.44
## Popular ~ Instrument + Voice + Harmony + Selfdeclare + Harmony:Selfdeclare
##
```

```
##                           Df Sum of Sq   RSS    AIC
## - Voice                    2     15.11 12302 3999.5
## <none>                                 12287 4000.4
## + Instrument:Selfdeclare   2      5.02 12282 4003.4
## + Voice:Selfdeclare        2      1.42 12285 4004.1
## - Harmony:Selfdeclare      3    126.11 12413 4019.9
## - Instrument               2   2918.29 15205 4527.7
##
## Step:  AIC=3999.5
## Popular ~ Instrument + Harmony + Selfdeclare + Harmony:Selfdeclare
##
##                           Df Sum of Sq   RSS    AIC
## <none>                                 12302 3999.5
## + Voice                    2     15.11 12287 4000.4
## + Instrument:Selfdeclare   2      5.00 12297 4002.5
## - Harmony:Selfdeclare      3    126.14 12428 4018.9
## - Instrument               2   2918.18 15220 4526.2
```

```
summary(lmfit5popboth)
```

```
##
## Call:
## lm(formula = Popular ~ Instrument + Harmony + Selfdeclare + Harmony:Selfdeclare,
##     data = ratings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.5418 -1.5948  0.0667  1.5536 12.5759
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 5.871512   0.214460  27.378  < 2e-16 ***
## Instrumentpiano            -0.947057   0.109500  -8.649  < 2e-16 ***
## Instrumentstring           -2.608501   0.108840 -23.966  < 2e-16 ***
## HarmonyI-V-IV              -0.312494   0.290411  -1.076 0.282013
## HarmonyI-V-VI               0.683485   0.290204   2.355 0.018591 *
## HarmonyIV-I-V              -0.167464   0.290067  -0.577 0.563770
## Selfdeclare                 0.335157   0.075541   4.437 9.53e-06 ***
## HarmonyI-V-IV:Selfdeclare   0.117885   0.107035   1.101 0.270842
## HarmonyI-V-VI:Selfdeclare  -0.389448   0.106976  -3.641 0.000278 ***
## HarmonyIV-I-V:Selfdeclare  -0.007795   0.106895  -0.073 0.941871
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.226 on 2483 degrees of freedom
## Multiple R-squared:  0.2128, Adjusted R-squared:  0.2099
## F-statistic: 74.57 on 9 and 2483 DF,  p-value: < 2.2e-16
```

Using model selection, we can see that the final model only has the interaction terms between Selfdeclare and Harmony and only one coefficient of the interaction term is statistically significant.

same sorts of difficulties as with #3

Professional Summary

Tiffany Chen

To examine the influence of the three main experiemental factor (Instrument, Harmony and Voice), on the Classical ratings and Popular ratings of music, we used both conventional linear regression and multilevel regression model. As I mentioned in 1a) and 4a), if we use conventional linear regression, Instrument is the most important and statistically significant factor in determining the Classical and Popular ratings of music. The coefficients on Instrumentpiano and Instrumentstring are both statistically significant when we regress the Classical or Popular ratings on Instrument, Harmony and Voice. If we remove Instrument from our linear regression models, the adjusted R square drop drastically. For Classical rating, Voice is also relatively important as both Voicepar3rd and Voicepar5th are statistically significant. But removing Voice from the regression does not make a big different on the adjusted R squared. But for Popular rating, Voice is not important at all since the coefficients are statistically insignificant and removing Voice does not change the R squared very much. The HarmonyI-V-VI appears to be statistically significant in both Classical ratings and Popular ratings. Furthermore, there is a significant interaction effect between HarmonyI-V-VI and the subject's self declaration of whether he or she knows music.

what direction were these effects?

7

To adjust for the random effects between each participant, we also utilize a multilevel model. The multilevel model also shows that Instrument is the most important factor in determining both Classical and Popular ratings. The coefficients on different levels of Instrument and different levels of Voice are all statistically significant in the case of Classical ratings. However, two of the coefficients on Harmony are not statistically significant. The random effects are mostly centered around 0 and the β's on the variables are very small.

what does this say that a researcher in the psychology of music might find useful?

However, when I include the covariates of all the 3 variables, as shown in 2a), 34 out of 70 α's are statistically significant. The intercept is significant when we include the covariates of all the 3 variables. Moreover all levels of Instrument and one level of Harmony is also statistically instrument. However, no level in Voice is statistically significant in this case.