36-763 Hierarchical Linear Model HW05

Tong Li

1. (a)

I use an individual fixed effect model as the baseline model, and see how do 'Instrument', 'Harmony' and 'Voice' improve the model fit.

First we fit the model including only individual indicators (factor variable 'Subject').

```
> fit.1a.baseline<-lm(Classical~Subject)
> summary(fit.1a.baseline)
```

```
Residual standard error: 2.328 on 2423 degrees of freedom
  (27 observations deleted due to missingness)
Multiple R-squared:  0.2536, Adjusted R-squared:  0.2324
F-statistic: 11.93 on 69 and 2423 DF,  p-value: < 2.2e-16
```

The result shows that the R-square is 0.2536, meaning that the individual indicators can explain about 25% of the variation in the classical ratings.

Then we include the factor variable 'Instrument' into the model.

```
> fit.1a.instru<-lm(Classical~Subject+Instrument)
> summary(fit.1a.instru)
```

```
Subject94          -0.47222     0.45481   -1.038 0.299238
Subject98          -0.47222     0.45481   -1.038 0.299238
Instrumentpiano     1.37636     0.09502   14.485  < 2e-16 ***
Instrumentstring    3.13148     0.09439   33.175  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.93 on 2421 degrees of freedom
  (27 observations deleted due to missingness)
Multiple R-squared:  0.4877, Adjusted R-squared:  0.4727
F-statistic: 32.46 on 71 and 2421 DF,  p-value: < 2.2e-16
```

It can be seen from the results that R-square is hugely increased from 0.2536 to 0.4877, meaning that 'Instrument' has brought considerable explanatory power to the model. Additionally, both instrument indicators are significant at 0.001 level. This means, with other factors equal, a stimuli of piano is associated with an average of 1.37 more points in classical rating compared to a stimuli of guitar; a stimuli of string is associated with an average of 3.13 more points in classical rating compared to that of guitar. Overall, the results indicate that 'Instrument' has significant influence on Classical ratings.

Similarly, we include the factor variable 'Harmony' into the model, to examine its influence on classical ratings.

```
> fit.1a.harmony<-lm(Classical~Subject+Harmony)
> summary(fit.1a.harmony)
```

```
Subject94      -0.47222    0.54327   -0.869 0.384809
Subject98      -0.47222    0.54327   -0.869 0.384809
HarmonyI-V-IV -0.02967    0.13055   -0.227 0.820224
HarmonyI-V-VI  0.77417    0.13055    5.930 3.46e-09 ***
HarmonyIV-I-V  0.05334    0.13044    0.409 0.682642
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.305 on 2420 degrees of freedom
  (27 observations deleted due to missingness)
Multiple R-squared:  0.2693,	Adjusted R-squared:  0.2476
F-statistic: 12.39 on 72 and 2420 DF,  p-value: < 2.2e-16
```

It can be seen from the result that 'Harmony' doesn't improve the model fit significantly. R-square only increases from 0.2536 to 0.2693. As to the coefficients of 'Harmony' indicators, only one of them (I-V-VI) is statistically significant. An I-V-VI stimuli is associated with a 0.77 more points in classical rating with everything else being equal. This is in line with the researchers' guess that I-V-VI might be frequently rated as classical due to people's familiarity with Pachelbel's Canon in D.

Finally we include 'Voice' into the model and examine how does it improve the model.

```
> fit.1a.voice<-lm(Classical~Subject+Voice)

> summary(fit.1a.voice)
```

```
Subject94    -0.47222    0.54715   -0.863 0.388194
Subject98    -0.47222    0.54715   -0.863 0.388194
Voicepar3rd -0.41346    0.11392   -3.629 0.000290 ***
Voicepar5th -0.37279    0.11385   -3.274 0.001074 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.321 on 2421 degrees of freedom
  (27 observations deleted due to missingness)
Multiple R-squared:  0.2585,	Adjusted R-squared:  0.2368
F-statistic: 11.89 on 71 and 2421 DF,  p-value: < 2.2e-16
```

Again, adding 'Voice' into the model doesn't improve the model fit. However, the coefficients of the two voice indicators are both significant. Stimuli of par3rd and par5th are both associated with a less classical rating compared to that of the contrary motion. This is in line with the researchers' expectation that contrary motion would be frequently rated as classical.

1. (b)

i. The model is as follows. Note that we denote participants as j, and each observation as i. We won't include any individual covariates in the intercept for now.

$$Classical = \alpha_{j[i]} + \beta_1 * Instrument + \beta_2 * Harmony + \beta_3 * Voice + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$$

$$\alpha_j = \alpha_0 + \eta_j, \eta_j \sim N(0, \tau^2)$$

ii. We use two methods to test whether the random effect is needed. The first method is to compare the DIC of the model with and without random effect. The second method is to check with simulation.

Method 1: Compare DIC

First we fit the original model without random effect in WinBUGS and get its DIC. The model takes the following form, where 'Subject' is a factor variable.

$$Classical \sim Subject + Instrument + Harmony + Voice$$

```
#### we first construct variables to store the numeric value of Subject and Instrument
Subject.num<-as.numeric(Subject)
Instrument.num<-as.numeric(Instrument)
Harmony.num<-as.numeric(Harmony)
Voice.num<-as.numeric(Voice)
#### calculate the n and J
n<-nrow(rating.data)    #2520
J<-length(unique(Subject))     #70
#### write the rube model
rube.lm.fixef<-"model {
for (i in 1:n) {
Classical[i]~dnorm(mu[i],sig2inv)
mu[i]<-b0[Subject.num[i]]+b1[Instrument.num[i]] + b2[Harmony.num[i]]+ b3[Voice.num[i]]
}
for (j in 1:J) {
b0[j]~dnorm(0,0.0001)
}
for (k in 1:3) {
b1[k]~ dnorm(0,0.0001)
}
for (p in 1:4) {
b2[p]~ dnorm(0,0.0001)
}
for (q in 1:3) {
b3[q]~ dnorm(0,0.0001)
```

```
}
sig2inv<-pow(sig,-2)
sig~dunif(0,100)
}"
data.list<-list(Classical=rating.data$Classical, Subject.num=Subject.num,
Instrument.num=Instrument.num, Harmony.num=Harmony.num, Voice.num=Voice.num, n=n,
J=J)
rube.lm.fixef.inits<-function() {
list(b0=rnorm(J), b1=rnorm(3), b2=rnorm(4), b3=rnorm(3), sig=runif(1,0,10))
}
rube(rube.lm.fixef, data.list, rube.lm.fixef.inits)
rube.lm.fixef.fit<-rube(rube.lm.fixef, data.list, rube.lm.fixef.inits,
parameters.to.save=c("b0","b1","b2","b3","sig"), n.chains=3)
rube.lm.fixef.fit
```

```
b0[9]    -7.42e-01 11.9364   -23.69    -8.73 -4.63e-01     7.98    21.84   1  1000
b0[10]   -3.17e-01 11.9473   -23.27    -8.33 -5.51e-02     8.53    21.55   1  1000
b1[1]     7.54e-02 46.3741   -88.00   -30.62 -8.30e-01    31.15    97.26   1   700
b1[2]     1.45e+00 46.3738   -86.69   -29.18  6.28e-01    32.62    98.72   1   690
b1[3]     3.21e+00 46.3751   -85.00   -27.47  2.28e+00    34.29   100.51   1   700
b2[1]     2.26e+00 42.9553   -87.08   -25.94  1.22e+00    31.10    84.45   1  1000
b2[2]     2.23e+00 42.9547   -87.03   -25.93  1.24e+00    30.98    84.39   1  1000
b2[3]     3.03e+00 42.9558   -86.19   -25.17  1.95e+00    31.79    85.10   1  1000
b2[4]     2.31e+00 42.9567   -86.99   -25.93  1.26e+00    31.15    84.52   1  1000
b3[1]     2.22e+00 47.2818   -90.53   -30.85  2.47e+00    34.20    92.79   1   710
b3[2]     1.80e+00 47.2819   -90.82   -31.22  1.95e+00    33.79    92.39   1   710
b3[3]     1.84e+00 47.2800   -90.84   -31.22  2.02e+00    33.77    92.38   1   710
deviance  1.03e+04 12.5069 10233.70 10247.39  1.03e+04 10263.90 10282.14   1  1000
sig       1.89e+00  0.0271     1.84     1.88  1.89e+00     1.91     1.95   1  1000

DIC = 10334.14
```

As shown in the above result screenshot, the DIC of this fixed effect model is 10334.14.

Next, we fit the random effect version of the model and get the DIC.

```
#### write the rube model
rube.lmer.ranef<-"model {
for (i in 1:n) {
Classical[i]~dnorm(mu[i],sig2inv)
mu[i]<-a0[Subject.num[i]]+b1[Instrument.num[i]] + b2[Harmony.num[i]]+ b3[Voice.num[i]]
}
for (j in 1:J) {
a0[j]~dnorm(b0,tau2inv)
}
b0~dnorm(0,0.0001)
for (k in 1:3) {
b1[k]~ dnorm(0,0.0001)
}
for (p in 1:4) {
b2[p]~ dnorm(0,0.0001)
```

```
}
for (q in 1:3) {
b3[q]~ dnorm(0,0.0001)
}
tau2inv<-pow(tau,-2)
tau~dunif(0,100)
sig2inv<-pow(sig,-2)
sig~dunif(0,100)
}"
data.list<-list(Classical=rating.data$Classical, Subject.num=Subject.num,
Instrument.num=Instrument.num, Harmony.num=Harmony.num, Voice.num=Voice.num, n=n,
J=J)
rube.lmer.ranef.inits<-function() {
list(b0=rnorm(1), b1=rnorm(3), b2=rnorm(4), b3=rnorm(3), a0=rnorm(J), sig=runif(1,0,10),
tau=runif(1,0,10))
}
rube(rube.lmer.ranef, data.list, rube.lmer.ranef.inits)
rube.lmer.ranef.fit<-rube(rube.lmer.ranef, data.list, rube.lmer.ranef.inits,
parameters.to.save=c("b0","b1","b2","b3","a0","tau","sig"), n.chains=3)
rube.lmer.ranef.fit
```

b0	3.858	4.6184	-5.42	-1.50	5.018	7.86	10.68	3.94	3
b1[1]	-0.518	45.9371	-88.87	-32.39	-2.721	31.55	84.03	1.00	1000
b1[2]	0.861	45.9388	-87.52	-31.06	-1.390	32.97	85.44	1.00	1000
b1[3]	2.616	45.9384	-85.78	-29.28	0.461	34.79	87.13	1.00	1000
b2[1]	1.695	41.6603	-77.38	-27.96	2.156	29.96	76.12	1.00	680
b2[2]	1.666	41.6552	-77.47	-27.96	2.207	29.92	76.19	1.00	690
b2[3]	2.464	41.6591	-76.69	-27.09	3.010	30.76	76.97	1.00	680
b2[4]	1.748	41.6568	-77.28	-27.91	2.217	29.97	76.21	1.00	690
b3[1]	-0.687	46.0521	-90.41	-30.99	-0.821	29.05	92.76	1.00	1000
b3[2]	-1.107	46.0537	-90.83	-31.41	-1.243	28.70	92.32	1.00	1000
b3[3]	-1.061	46.0531	-90.79	-31.38	-1.248	28.64	92.41	1.00	1000
deviance	10256.247	12.2991	10234.45	10247.83	10256.308	10264.22	10281.25	1.00	850
sig	1.893	0.0264	1.84	1.88	1.893	1.91	1.95	1.00	470
tau	1.326	0.1218	1.11	1.24	1.318	1.40	1.59	1.01	350

```
DIC = 10331.85
```

As shown in the above result screenshot, the DIC of this fixed effect model is 10331.85.

The difference of the DIC between the fixed effect model and random effect model is 2.29. By the rule of thumb, the difference is marginally interesting, which means that a random effect model may perform better.

Method 2: Simulation

To check whether a participant random effect is needed, we fit the original fixed effect model where subjects are included as factor variables, and simulate new data based on the fitted model. We examine the spread in the simulated data and compare it with the observed data. If the simulated data is more spread out, then we probably need a random effect model.

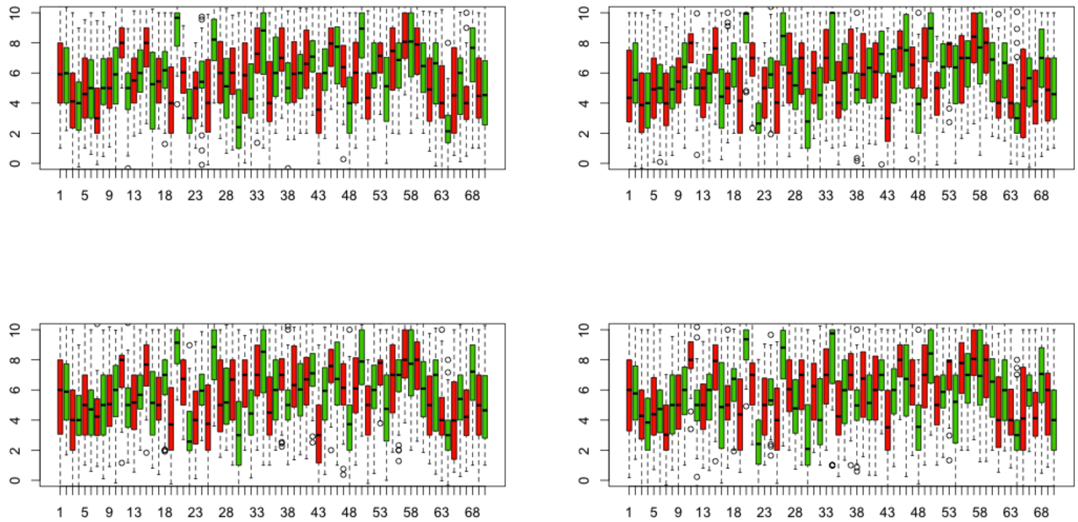
```
####First we get the estimated parameter values
attach(rube.lm.fixef.fit$sims.list)
```

```
b0.hat<-apply(b0,2,mean)
b1.hat<-apply(b1,2,mean)
b2.hat<-apply(b2,2,mean)
b3.hat<-apply(b3,2,mean)
sig.hat<-mean(sig)
detach()

#### write the rube model
rube.lm.fixef.new<-"model {
for (i in 1:n) {
Classical[i]~dnorm(mu[i],sig2inv)
mu[i]<-b0[Subject.num[i]]+b1[Instrument.num[i]] + b2[Harmony.num[i]]+ b3[Voice.num[i]]
}
for (j in 1:J) {
b0[j]~dnorm(0,0.0001)
}
for (k in 1:3) {
b1[k]~ dnorm(0,0.0001)
}
for (p in 1:4) {
b2[p]~ dnorm(0,0.0001)
}
for (q in 1:3) {
b3[q]~ dnorm(0,0.0001)
}
sig2inv<-pow(sig,-2)
sig~dunif(0,100)
for (i in 1:n) {
newClassical[i]~dnorm(mu[i],sig2inv)
}
}"
data.list<-list(Classical=rating.data$Classical, Subject.num=Subject.num,
Instrument.num=Instrument.num, Harmony.num=Harmony.num, Voice.num=Voice.num, n=n,
J=J)
rube.lm.fixef.new.inits<-function() {
list(b0=b0.hat, b1=b1.hat, b2=b2.hat, b3=b3.hat, sig=sig.hat)
}
rube(rube.lm.fixef.new, data.list, rube.lm.fixef.new.inits)
rube.lm.fixef.new.fit<-rube(rube.lm.fixef.new, data.list, rube.lm.fixef.new.inits,
parameters.to.save=c("newClassical"), n.iter=400,n.chains=1)

newClassical<-rube.lm.fixef.new.fit$sims.list$newClassical

(n.sims <- rube.lm.fixef.new.fit$n.keep)
```

Next we visualize the variability in ‘Classical’ ratings between participants.

Simulated data is in green, and observed data is in red.

```
boxes <- function(i,col=2:3,ylim=c(0,10)) {
tmp <- c(t(cbind(Classical,newClassical[i,])))
boxplot(split(tmp,rep(Subject.num,rep(2,n))),col=col,ylim=ylim)
}
par(mfrow=c(2,2))
samp <- sample(1:n.sims,4)

for (i in samp) boxes(i)
```



It can be seen that the green boxes are a little more spread out than red ones. This means that this fitted model tends to spread data out than what is happening in reality. Therefore we can shrink the data by making the participant rating intercept as a random effect.

iii. We first estimate the 'repeated-measures model' with only the random intercept for participants, and then add the three main experimental factors.

First, we estimate the 'repeated-measures model' with only the random intercept for participants.

```
#### write the rube model
rube.lmer.ranef<-"model {
for (i in 1:n) {

Classical[i]~dnorm(mu[i],sig2inv)
mu[i]<-a0[Subject.num[i]]
}
for (j in 1:J) {
a0[j]~dnorm(b0,tau2inv)
}
b0~dnorm(0,0.0001)
```

```
tau2inv<-pow(tau,-2)
tau~dunif(0,100)
sig2inv<-pow(sig,-2)
sig~dunif(0,100)
}"
data.list<-list(Classical=rating.data$Classical, Subject.num=Subject.num, n=n, J=J)
rube.lmer.ranef.inits<-function() {
list(b0=rnorm(1), a0=rnorm(J), sig=runif(1,0,10), tau=runif(1,0,10))
}
rube(rube.lmer.ranef, data.list, rube.lmer.ranef.inits)
rube.lmer.ranef.fit<-rube(rube.lmer.ranef, data.list, rube.lmer.ranef.inits,
parameters.to.save=c("b0","a0","tau","sig"), n.chains=3)

rube.lmer.ranef.fit
```

```
a0[9]        5.32  0.3690     4.61     5.07     5.32     5.57     6.09 1.00  720
a0[10]       6.06  0.4549     5.18     5.75     6.06     6.36     6.97 1.00  1000
b0           5.79  0.1588     5.48     5.68     5.79     5.89     6.10 1.00  1000
deviance 11289.10 11.9789 11268.61 11280.45 11288.82 11296.97 11313.26 1.00  1000
sig          2.33  0.0326     2.27     2.31     2.33     2.35     2.39 1.00   660
tau          1.30  0.1222     1.10     1.22     1.29     1.38     1.56 1.00  1000

DIC = 11360.97
```

Then we add ‘Instrument’ to the model

```
#### write the rube model
rube.lmer.ranef<-"model {
for (i in 1:n) {
Classical[i]~dnorm(mu[i],sig2inv)
mu[i]<-a0[Subject.num[i]]+b1[Instrument.num[i]]
}
for (j in 1:J) {
a0[j]~dnorm(b0,tau2inv)
}
b0~dnorm(0,0.0001)
for (k in 1:3) {
b1[k]~ dnorm(0,0.0001)
}
tau2inv<-pow(tau,-2)
tau~dunif(0,100)
sig2inv<-pow(sig,-2)
sig~dunif(0,100)
}"
data.list<-list(Classical=rating.data$Classical, Subject.num=Subject.num,
Instrument.num=Instrument.num, n=n, J=J)
rube.lmer.ranef.inits<-function() {
list(b0=rnorm(1), b1=rnorm(3), a0=rnorm(J), sig=runif(1,0,10), tau=runif(1,0,10))
```

```
}
rube(rube.lmer.ranef, data.list, rube.lmer.ranef.inits)
rube.lmer.ranef.fit<-rube(rube.lmer.ranef, data.list, rube.lmer.ranef.inits,
parameters.to.save=c("b0","b1","a0","tau","sig"), n.chains=3)

rube.lmer.ranef.fit
```

```
a0[10]      -8.97 10.0026 -2.60e+01   -17.45    -8.86     0.969     7.51 5.29     3
b0          -8.89  9.9939 -2.58e+01   -17.32    -8.86     1.009     7.48 5.34     3
b1[1]       13.16  9.9929 -3.27e+00     3.20    13.19    21.512    30.06 5.35     3
b1[2]       14.54  9.9920 -1.83e+00     4.62    14.53    22.941    31.42 5.35     3
b1[3]       16.29  9.9957 -2.66e-02     6.37    16.26    24.655    33.19 5.35     3
deviance 10352.35 12.1559  1.03e+04 10344.00 10351.71 10360.163 10377.15 1.00  1000
sig          1.93  0.0287  1.88e+00     1.91     1.93     1.949     1.98 1.00  1000
tau          1.33  0.1266  1.11e+00     1.24     1.32     1.408     1.59 1.00  1000

DIC = 10426.3
```

The DIC increases from 11360.67 to 10426.3, and the variance of residual is reduced from 2.33 to 1.93, indicating that adding 'Instrument' hugely improve the model fit. 'Instrument' helps a lot to explain the variation in data. This means that 'Instrument' is an important factor that influences people's classical rating. As shown in the result, a guitar usually leads to about 1.38 points (b1[2]-b1[1]) less in terms of Classical rating than a piano; a piano usually leads to about 1.75 points less than a string. The result is in line with that of problem (a).

We then add 'Harmony' into the model, and estimate it using rube.

```
#### write the rube model
rube.lmer.ranef<-"model {
for (i in 1:n) {
Classical[i]~dnorm(mu[i], sig2inv)
mu[i]<-a0[Subject.num[i]]+b1[Instrument.num[i]] +
b2[Harmony.num[i]]
}
for (j in 1:J) {
a0[j]~dnorm(b0,tau2inv)
}
b0~dnorm(0,0.0001)
for (k in 1:3) {
b1[k]~ dnorm(0,0.0001)
}
for (p in 1:4) {
b2[p]~ dnorm(0,0.0001)
}
tau2inv<-pow(tau,-2)
tau~dunif(0,100)
sig2inv<-pow(sig,-2)
sig~dunif(0,100)
```

```
}"
data.list<-list(Classical=rating.data$Classical, Subject.num=Subject.num,
Instrument.num=Instrument.num, Harmony.num=Harmony.num, n=n, J=J)
rube.lmer.ranef.inits<-function() {
list(b0=rnorm(1), b1=rnorm(3), b2=rnorm(4), a0=rnorm(J), sig=runif(1,0,10), tau=runif(1,0,10))
}
rube(rube.lmer.ranef, data.list, rube.lmer.ranef.inits)
rube.lmer.ranef.fit<-rube(rube.lmer.ranef, data.list, rube.lmer.ranef.inits,
parameters.to.save=c("b0","b1","b2","a0","tau","sig"), n.chains=3)
rube.lmer.ranef.fit
```

```
b0         2.79e-01  4.2988    -7.60    -2.52 -9.75e-02     3.59     7.79 3.03    4
b1[1]     -9.09e-01 39.2282   -75.87   -27.10 -1.74e+00    25.84    76.19 1.01  160
b1[2]      4.62e-01 39.2273   -74.45   -25.80 -3.93e-01    27.24    77.61 1.01  160
b1[3]      2.22e+00 39.2265   -72.68   -24.00  1.42e+00    29.03    79.36 1.01  160
b2[1]      4.71e+00 38.9991   -71.65   -22.72  5.91e+00    30.66    79.69 1.00 1000
b2[2]      4.68e+00 39.0016   -71.68   -22.67  5.88e+00    30.55    79.68 1.00 1000
b2[3]      5.48e+00 39.0036   -70.97   -21.85  6.66e+00    31.29    80.60 1.00 1000
b2[4]      4.76e+00 39.0024   -71.54   -22.56  5.99e+00    30.70    79.78 1.00 1000
deviance   1.03e+04 12.8709 10254.49 10269.65  1.03e+04 10286.48 10304.25 1.00  880
sig        1.90e+00  0.0269     1.85     1.88  1.90e+00     1.92     1.96 1.01  270
tau        1.33e+00  0.1172     1.11     1.25  1.32e+00     1.40     1.59 1.00 1000

DIC = 10361.23
```

It is shown that adding 'Harmony' decrease the model DIC from 10426.3 to 10361.23, indicating that 'Harmony' has influence on classical ratings. From the coefficient estimates, we see that b2[1], b2[2] and b2[4] are all close to each other; only b2[3] has an about 0.7 increment compared to the others. In the model, b2[3] is the coefficient for I-V-VI. This means that a stimuli of I-V-VI is more likely to get a higher classical rating.

Finally, we add 'Voice' to the model, which is the model we fit in (b. ii). The DIC value is 10331.85, smaller than 10361.23. This means that 'Voice' also have influence on classical ratings.

1. (c)

i. We estimate the model including all the three random effects as well as the three design factors (please refer to iii.) using WinBUGS

(An equivalent approach would be to estimate using lmer(Classical~(1|Subject:Instrument)+(1|Subject:Harmony)+(1|Subject:Voice) +Instrument+Harmony+Voice))

```
#### write the rube model
rube.lmer.ranef<-"model {
for (i in 1:n) {
Classical[i]~dnorm(mu[i], sig2inv)
mu[i]<-a_instr[Subject.num[i],Instrument.num[i]]+a_har[Subject.num[i],Harmony.num[i]]+a_voice[Subject.num[i],Voice.num[i]]+b1[Instrument.num[i]]+b2[Harmony.num[i]]+b3[Voice.num[i]]
```

```
}
for (j in 1:J) {
for (k in 1:3) {
a_instr[j,k]~dnorm(a1,tau1inv)
}}
for (j in 1:J) {
for (p in 1:4) {
a_har[j,p]~dnorm(a2,tau2inv)
}}
for (j in 1:J) {
for (q in 1:3) {
a_voice[j,q]~dnorm(a3,tau3inv)
}}
for (k in 1:3) {
b1[k]~dnorm(0,0.0001)
}
for (p in 1:4) {
b2[p]~dnorm(0,0.0001)
}
for(q in 1:3) {
b3[q]~dnorm(0,0.0001)
}
a1~dnorm(0,0.0001)
tau1inv<-pow(tau1,-2)
tau1~dunif(0,100)
a2~dnorm(0,0.0001)
tau2inv<-pow(tau2,-2)
tau2~dunif(0,100)
a3~dnorm(0,0.0001)
tau3inv<-pow(tau3,-2)
tau3~dunif(0,100)
sig2inv<-pow(sig,-2)
sig~dunif(0,100)
}"
data.list<-list(Classical=rating.data$Classical, Subject.num=Subject.num,
Instrument.num=Instrument.num, Harmony.num=Harmony.num,Voice.num=Voice.num, n=n,
J=J)
rube.lmer.ranef.inits<-function() {
list(a1=rnorm(1), a2=rnorm(1), a3=rnorm(1), b1=rnorm(3), b2=rnorm(4), b3=rnorm(3),
a_instr=matrix(rnorm(70*3),70,3), a_har=matrix(rnorm(70*4),70,4),
a_voice=matrix(rnorm(70*3),70,3), sig=runif(1,0,10),
tau1=runif(1,0,10),tau2=runif(1,0,10),tau3=runif(1,0,10))
}
rube(rube.lmer.ranef, data.list, rube.lmer.ranef.inits)
```

```
rube.lmer.ranef.fit<-rube(rube.lmer.ranef, data.list, rube.lmer.ranef.inits,
parameters.to.save=c("a1","a2","a3","b1","b2","b3","a_instr","a_har","a_voice","tau1","tau2","tau3
","sig"), n.chains=3)
rube.lmer.ranef.fit
```

The results are as follows

```
> rube.lmer.ranef.fit
Rube Results:
Run by jags at 2013-12-09 08:13 and taking 31.28 secs
                mean      sd     2.5%      25%      50%       75%    97.5%  Rhat n.eff
a1             2.519  4.1785  -5.0930  -0.5531   3.3611  5.70e+00    8.598  4.28     3
a2            -2.142  3.1407  -7.7498  -4.8652  -0.9660  4.65e-01    1.459  4.57     3
a3            -1.123  1.6320  -3.7051  -3.2551  -0.1245  1.69e-02    0.549 15.09     3
a_har[1,1]    -2.637  3.1922  -8.5718  -5.5796  -1.5353 -4.52e-02    1.179  4.27     3
a_har[2,1]    -1.997  3.1757  -7.8065  -4.8825  -0.8306  6.41e-01    1.939  4.10     3
a_har[3,1]    -2.278  3.1685  -8.0908  -5.1890  -1.1264  3.09e-01    1.594  4.25     3
```

… … …

```
a_voice[8,2]  -1.107  1.6466  -3.7718  -3.2302  -0.1191  7.18e-02    0.625 11.88     3
a_voice[9,2]  -1.129  1.6396  -3.7736  -3.2456  -0.1431  3.34e-02    0.543 11.99     3
a_voice[10,2] -1.127  1.6385  -3.7842  -3.2481  -0.1277  4.79e-02    0.516 12.68     3
b1[1]         -0.616 45.6707 -91.1778 -29.8545  -0.0351  2.88e+01   87.360  1.01   190
b1[2]          0.749 45.6783 -89.6007 -28.3796   1.6072  3.03e+01   88.291  1.01   190
b1[3]          2.514 45.6726 -87.6349 -26.7729   3.4443  3.19e+01   90.239  1.01   190
b2[1]          2.811 42.4675 -77.3348 -26.7522   4.4214  3.16e+01   83.646  1.00   900
b2[2]          2.783 42.4754 -77.3040 -26.9497   4.3628  3.14e+01   83.552  1.00   910
b2[3]          3.584 42.4684 -76.6218 -26.1735   5.1138  3.24e+01   84.373  1.00   900
b2[4]          2.866 42.4656 -77.3098 -26.8562   4.5847  3.17e+01   83.657  1.00   900
b3[1]          2.893 44.5869 -84.1858 -26.3592   1.3201  3.26e+01   88.246  1.00   450
b3[2]          2.486 44.5852 -84.6660 -26.6804   0.9138  3.22e+01   87.900  1.00   450
b3[3]          2.522 44.5855 -84.5627 -26.6419   0.9185  3.23e+01   87.775  1.00   450
deviance    9301.195 33.8017 9237.8302 9278.2429 9300.8064 9.32e+03 9368.969 1.01  360
sig            1.562  0.0237   1.5163   1.5455   1.5630  1.58e+00    1.610  1.00  1000
tau1           1.494  0.0911   1.3232   1.4289   1.4932  1.55e+00    1.683  1.01   190
tau2           0.669  0.0565   0.5627   0.6277   0.6661  7.07e-01    0.783  1.01   240
tau3           0.143  0.0653   0.0457   0.0957   0.1289  1.86e-01    0.290  1.09    36

DIC = 9870.439
```

From the above result table, we can see that the model DIC is 9870.439, much smaller than that of any model in 1(a) and 1(b). Additionally, the residual variance sig=1.562, much smaller than any of that in previous models. This means that this model with all three random effect terms fit better than any of the previous models.

ii. In this model, we still see that 'Instrument' should have the biggest impact on 'Classical ratings', because the personal bias distribution regarding 'Instrument' is the most spread out one (tau1 is biggest, even as big as the residual variance). Besides the personal bias being accounted for in the model, we see that guitar is negatively associated with classical rating (b1[1]=-0.616), and the other two instruments are both positively associated ---- piano is associated with an average of 0.749 points and string with 2.514 points. Compare this result to what we have in the conventional regression in (a), we have got very similar results: treating guitar as a baseline, piano is associated with about 1.37 points more, and string is associated with about 3.1 points more. We can reach the same conclusion for ‘Harmony’ and ‘Voice’ by examining their coefficients.

The three estimated variance components are tau1=1.494, tau2=0.669 and tau3=0.143. We can see that the ‘personal bias’ regarding ‘Instrument’ varies the most, followed by that regarding ‘Harmony’ and then ‘Voice’. The estimated residual variance is 1.562, only a little bigger than the variance of ‘Instrument personal bias’ distribution.

iii. Write this model in mathematical terms

$$Classical = \alpha_{[j:Instr][i]} + \alpha_{[j:Har][i]} + \alpha_{[j:Voice][i]} + \beta_1 Instrument + \beta_2 Harmony + \beta_3 Voice + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$$

$$\alpha_{[j:Instr][i]} = \alpha_1 + \eta_{[j:Instr]}, \eta_{[j:Instr]} \sim N(0, {\tau_1}^2)$$

$$\alpha_{[j:Har][i]} = \alpha_2 + \eta_{[j:Har]}, \eta_{[j:Har]} \sim N(0, {\tau_2}^2)$$

$$\alpha_{[j:Voice][i]} = \alpha_3 + \eta_{[j:Voice]}, \eta_{[j:Voice]} \sim N(0, {\tau_3}^2)$$

2. (a)

To simplify the exploratory process, we use lmer() instead of WinBUGS to fit the model. We add the listed individual variable one by one into the model in 1(c) and see whether the coefficient of the variable is significant. By the rule of thumb, we will keep the variables whose t-value is larger than 2.

It turns out that only one variable has a t-value larger than 2, that is X16.minus.17. As shown in the below result screenshot, the coefficient of X16.minus.17 is estimated to be -0.09786, with a standard error 0.03804, and the t-value is -2.572.

```
> summary(lmer(Classical~(1|Subject:Instrument)+(1|Subject:Harmony)+(1|Subject:Voice)+Instrument+Harmony+Voice+X16.minus.17))
Linear mixed model fit by REML ['lmerMod']
Formula: Classical ~ (1 | Subject:Instrument) + (1 | Subject:Harmony) +     (1 | Subject:Voice) + Instrument + Harmony + Voice + X16.minus.17

REML criterion at convergence: 10049.67

Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.4424   0.6651
 Subject:Voice      (Intercept) 0.0281   0.1676
 Subject:Instrument (Intercept) 2.1327   1.4604
 Residual                       2.4377   1.5613
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210

Fixed effects:
                 Estimate Std. Error t value
(Intercept)       4.50915    0.22196  20.315
Instrumentpiano   1.36415    0.25871   5.273
Instrumentstring  3.12910    0.25842  12.108
HarmonyI-V-IV    -0.03069    0.14311  -0.214
HarmonyI-V-VI     0.77042    0.14309   5.384
HarmonyIV-I-V     0.05598    0.14304   0.391
Voicepar3rd      -0.40724    0.08175  -4.982
Voicepar5th      -0.37105    0.08168  -4.543
X16.minus.17     -0.09786    0.03804  -2.572
```

We estimate the same model again in WinBUGS, and the results are the same. Here b4 is the coefficient of X16.minus.17.

```
b3[3]        2.32e-01 47.7782  -92.3908 -3.24e+01    0.9128   32.2857   93.0410 1.00   920
b4          -9.69e-02  0.0376   -0.1702 -1.23e-01   -0.0964   -0.0717   -0.0219 1.00  1000
deviance     9.30e+03 35.1339 9232.4973  9.27e+03 9298.9046 9320.0685 9367.6891 1.02   130
sig          1.56e+00  0.0257    1.5136  1.55e+00    1.5621    1.5797    1.6149 1.01   330
tau1         1.47e+00  0.0872    1.3005  1.41e+00    1.4647    1.5195    1.6448 1.00  1000
tau2         6.70e-01  0.0572    0.5530  6.31e-01    0.6706    0.7042    0.7840 1.00   700
tau3         1.71e-01  0.0622    0.0668  1.28e-01    0.1627    0.2135    0.3029 1.10    29

DIC = 9906.859
```

By the rule of thumb, the coefficient is significant, indicating that X16.minus.17 is negatively associated with classical ratings. All the other individual variables are not statistically significant. Therefore, the final model takes the following form

Classical~(1|Subject:Instrument)+(1|Subject:Harmony)+(1|Subject:Voice)+Instrument+Harmony+Voice+X16.minus.17

2. (b)

We compare the random effects of the two models (with and without X16.minus.17). Overall, the random effects don't vary much. It worth to note that the random effects regarding 'Instrument' has the relatively biggest change after we add the individual covariate X16.minus.17 into the model. Its variance is reduced from 2.198 to 2.133. This indicates that X16.minus.17 actually captures some of the variation reflected in people's personal judgments regarding instruments.

```
> summary(lmer(Classical~(1|Subject:Instrument)+(1|Subject:Harmony)+(1|Subject:Voice)+Instrument+Harmony+Voice+X16.minus.17))
Linear mixed model fit by REML ['lmerMod']
Formula: Classical ~ (1 | Subject:Instrument) + (1 | Subject:Harmony) +     (1 | Subject:Voice) + Instrument + Harmony + Voice + X16.minus.17

REML criterion at convergence: 10049.67

Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.4424   0.6651
 Subject:Voice      (Intercept) 0.0281   0.1676
 Subject:Instrument (Intercept) 2.1327   1.4604
 Residual                       2.4377   1.5613
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210


> summary(lmer(Classical~(1|Subject:Instrument)+(1|Subject:Harmony)+(1|Subject:Voice)+Instrument+Harmony+Voice))
Linear mixed model fit by REML ['lmerMod']
Formula: Classical ~ (1 | Subject:Instrument) + (1 | Subject:Harmony) +     (1 | Subject:Voice) + Instrument + Harmony + Voice

REML criterion at convergence: 10051.51

Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.44307  0.6656
 Subject:Voice      (Intercept) 0.02809  0.1676
 Subject:Instrument (Intercept) 2.19850  1.4827
 Residual                       2.43753  1.5613
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210
```

2. (c)

People have intrinsic personal biases when rate music and such biases vary with the type of instruments, harmony and voice leading. Among them, the biases regarding instrument vary the most (variance of (1|Subject:Instrument) is 2.1327), followed by that regarding harmony (0.4424) and biases regarding voice has the least spread out distribution, with only a 0.0281 variance. This means that instrument type is the most important factor that leads to the variation in classical ratings.

As can be seen from the coefficients of 'instrument' indicators, music played by string on average earns the highest classical ratings, which is about 1.8 points more than that played by Piano and about 3.12 points more than that played by guitar. As to the type of harmony, we see that harmony type I-V-VI leads to the highest classical ratings among all the harmony types. On average, the difference between I-V-VI and other types is about 0.77 points. There exists a statistically significant difference between the voice type Contrary Motion and Parallel 3rd, 5th. A stimuli music of Contrary Motion tends to be rated 0.4 points higher in classical rating than the other two voice types do.

3.

We recode 'Selfdeclare' to a new variable 'Musician'. We group the cases 'Selfdeclare=1,2' to 'Musician=0', and the rest are 'Musician=1', so that we have about half people musicians and half non-musicians.

```
>Musician<-ifelse((Selfdeclare==1|Selfdeclare==2),0,1)
```

Now we interact the 'Musician' indicator with the three design factors 'Instrument', 'Harmony' and 'Voice', and examine how does this change the model estimates.

First we interact 'Musician' with 'Instrument'.

```
REML criterion at convergence: 10047.24

Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.44257  0.6653
 Subject:Voice      (Intercept) 0.02817  0.1678
 Subject:Instrument (Intercept) 2.13612  1.4615
 Residual                       2.43767  1.5613
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210

Fixed effects:
                          Estimate Std. Error t value
(Intercept)                4.27486    0.26604  16.069
Musician                   0.60908    0.38585   1.579
Instrumentpiano            1.61512    0.33410   4.834
Instrumentstring           3.41217    0.33394  10.218
HarmonyI-V-IV             -0.03079    0.14313  -0.215
HarmonyI-V-VI              0.77038    0.14311   5.383
HarmonyIV-I-V              0.05590    0.14306   0.391
Voicepar3rd               -0.40732    0.08176  -4.982
Voicepar5th               -0.37107    0.08169  -4.542
X16.minus.17              -0.10344    0.03890  -2.659
Musician:Instrumentpiano  -0.62679    0.52859  -1.186
Musician:Instrumentstring -0.70681    0.52783  -1.339
```

Although the two interaction estimates 'Musician:Instrumentpiano' and 'Musician:Instrumentstring' are not statistically significant at 0.05 level, their t-value are both larger than 1 and their signs are both negative. This indicates to some extent that musicians tend to be more conservative than non-musicians in terms of rating music played by piano and string music as classical. In others words, factor 'instrument' is less influential for musicians than for non-musicians.

Then we interact 'Musician' with 'Harmony'

```
Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.35861  0.5988
 Subject:Voice      (Intercept) 0.02759  0.1661
 Subject:Instrument (Intercept) 2.19218  1.4806
 Residual                       2.43761  1.5613
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210

Fixed effects:
                     Estimate Std. Error t value
(Intercept)           4.58641    0.24483  18.733
Instrumentpiano       1.36452    0.26198   5.209
Instrumentstring      3.12918    0.26169  11.958
Musician             -0.17062    0.28994  -0.588
HarmonyI-V-IV        -0.04954    0.17353  -0.285
HarmonyI-V-VI         0.28283    0.17349   1.630
HarmonyIV-I-V         0.02559    0.17344   0.148
Voicepar3rd          -0.40700    0.08165  -4.984
Voicepar5th          -0.37023    0.08159  -4.538
X16.minus.17         -0.10338    0.03885  -2.661
Musician:HarmonyI-V-IV  0.04741  0.27457   0.173
Musician:HarmonyI-V-VI  1.21991  0.27455   4.443
Musician:HarmonyIV-I-V  0.07529  0.27441   0.274
```

It can be seen from the above results that, after adding the interaction terms, 'HarmonyI-V-VI' is no longer significant, but the interaction term 'Musician:HarmonyI-V-VI' is very significant and is positive. This means that musicians and non-musicians do treat Harmony I-V-VI differently. Musicians tend to give music of I-V-VI type a high classical rating while non-musicians don't.

Finally, we interact 'Musician' with 'Voice'. The interaction terms are not significant. There is no obvious evidence that musicians and non-musicians treat 'Voice' differently.

```
Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.44281  0.6654
 Subject:Voice      (Intercept) 0.03078  0.1754
 Subject:Instrument (Intercept) 2.13586  1.4615
 Residual                       2.43750  1.5613
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210

Fixed effects:
                     Estimate Std. Error t value
(Intercept)           4.43291    0.23964  18.499
Instrumentpiano       1.36463    0.25889   5.271
Instrumentstring      3.12923    0.25860  12.101
HarmonyI-V-IV        -0.03053    0.14315  -0.213
HarmonyI-V-VI         0.77039    0.14313   5.382
HarmonyIV-I-V         0.05615    0.14308   0.392
Musician              0.21448    0.25588   0.838
Voicepar3rd          -0.39867    0.10604  -3.759
Voicepar5th          -0.32058    0.10594  -3.026
X16.minus.17         -0.10344    0.03892  -2.658
Musician:Voicepar3rd -0.02176    0.16788  -0.130
Musician:Voicepar5th -0.12690    0.16777  -0.756
```

4. (a)

We fit the similar hierarchical linear model on 'Popular Ratings' in WinBUGS. We first estimate a model with only the three random effects, and then include the three design factors one by one, to see how they improve the model fit.

The DIC of the three random effects model is DIC=10036.85. Adding 'Instrument', the model DIC=9953.106; adding 'Harmony', the model DIC=9968.583; adding 'Voice', the model DIC=9984.148. This indicates that 'Instrument' has the biggest influence among the three desing factors on the variability of popular rating.

4. (b)

```
> summary(lmer(Popular~(1|Subject:Instrument)+(1|Subject:Harmony)+
(1|Subject:Voice)+Instrument+Harmony+Voice+X16.minus.17))
Linear mixed model fit by REML ['lmerMod']
Formula: Popular ~ (1 | Subject:Instrument) + (1 | Subject:Harmony) +     (1 | Subject:Voice) + Instrument +
Harmony + Voice + X16.minus.17

REML criterion at convergence: 10073.56

Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.41100  0.6411
 Subject:Voice      (Intercept) 0.03206  0.1790
 Subject:Instrument (Intercept) 1.96194  1.4007
 Residual                       2.49060  1.5782
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210

Fixed effects:
                Estimate Std. Error t value
(Intercept)      6.44644    0.21520  29.956
Instrumentpiano -0.94923    0.24936  -3.807
Instrumentstring -2.60650   0.24906 -10.465
HarmonyI-V-IV   -0.02517    0.14054  -0.179
HarmonyI-V-VI   -0.27138    0.14053  -1.931
HarmonyIV-I-V   -0.18528    0.14047  -1.319
Voicepar3rd      0.16402    0.08320   1.971
Voicepar5th      0.16224    0.08314   1.951
X16.minus.17     0.07772    0.03672   2.117
```

how did you arrive at this model?

Similar with the case of 'classical rating', instrument type is the most important factor that leads to the variation in people's intrinsic bias regarding whether a music is popular. The fixed effects are mostly opposite with that in the 'classical rating' case. As can be seen from the coefficients of 'instrument' indicators, music played by string on average earns the lowest popular ratings, which is about 1.7 points less than that played by piano and about 2.6 points less than that played by guitar. As to the type of harmony, we see that harmony type I-V-VI leads to the lowest popular ratings among all the harmony types. On average, the difference between I-V-VI and other types is about 0.27 points. There exists a difference between the voice type Contrary Motion and Parallel 3rd, 5th. A stimuli music of Contrary Motion tends to be rated 0.16 points lower in popular rating than the other two voice types do.

4. (c)

Now we interact the 'Musician' indicator with the three design factors 'Instrument', 'Harmony' and 'Voice', and examine how does this change the model estimates.

First, we show the result of the original model

```
>summary(lmer(Popular~(1|Subject:Instrument)+(1|Subject:Harmony)+(1|Subject:Voice)+Instrument+Harmony+Voice+X16.minus.17))
```

```
REML criterion at convergence: 10073.56

Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.41100  0.6411
 Subject:Voice      (Intercept) 0.03206  0.1790
 Subject:Instrument (Intercept) 1.96194  1.4007
 Residual                       2.49060  1.5782
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210

Fixed effects:
                 Estimate Std. Error t value
(Intercept)       6.44644    0.21520  29.956
Instrumentpiano  -0.94923    0.24936  -3.807
Instrumentstring -2.60650    0.24906 -10.465
HarmonyI-V-IV    -0.02517    0.14054  -0.179
HarmonyI-V-VI    -0.27138    0.14053  -1.931
HarmonyIV-I-V    -0.18528    0.14047  -1.319
Voicepar3rd       0.16402    0.08320   1.971
Voicepar5th       0.16224    0.08314   1.951
X16.minus.17      0.07772    0.03672   2.117
```

Then we interact ‘Musician’ with ‘Instrument’.

```
Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.41045  0.6407
 Subject:Voice      (Intercept) 0.03189  0.1786
 Subject:Instrument (Intercept) 1.92586  1.3878
 Residual                       2.49073  1.5782
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210

Fixed effects:
                          Estimate Std. Error t value
(Intercept)                6.42952    0.25558  25.156
Musician                   0.11264    0.36882   0.305
Instrumentpiano           -1.12248    0.31910  -3.518
Instrumentstring          -2.87966    0.31893  -9.029
HarmonyI-V-IV             -0.02523    0.14049  -0.180
HarmonyI-V-VI             -0.27139    0.14047  -1.932
HarmonyIV-I-V             -0.18531    0.14042  -1.320
Voicepar3rd                0.16399    0.08318   1.972
Voicepar5th                0.16212    0.08311   1.951
X16.minus.17               0.06135    0.03725   1.647
Musician:Instrumentpiano   0.43532    0.50489   0.862
Musician:Instrumentstring  0.68324    0.50409   1.355
```

The two coefficients of instrument indicators are still significant, and none of the interaction terms are significant. Therefore we believe there’s no obvious evidence that musicians and non-musicians treat instrument differently.

Next, we interact ‘Musician’ with ‘Harmony’

```
Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.37632  0.6134
 Subject:Voice      (Intercept) 0.03186  0.1785
 Subject:Instrument (Intercept) 1.94636  1.3951
 Residual                       2.49023  1.5780
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210

Fixed effects:
                      Estimate Std. Error t value
(Intercept)            6.23688    0.23703  26.313
Instrumentpiano       -0.94820    0.24847  -3.816
Instrumentstring      -2.60606    0.24816 -10.502
Musician               0.59486    0.28370   2.097
HarmonyI-V-IV         -0.11064    0.17674  -0.626
HarmonyI-V-VI          0.01209    0.17671   0.068
HarmonyIV-I-V         -0.20880    0.17665  -1.182
Voicepar3rd            0.16371    0.08316   1.969
Voicepar5th            0.16156    0.08310   1.944
X16.minus.17           0.06126    0.03721   1.646
Musician:HarmonyI-V-IV  0.21346   0.27965   0.763
Musician:HarmonyI-V-VI -0.70942   0.27963  -2.537
Musician:HarmonyIV-I-V  0.05827   0.27949   0.208
```

We see that, after adding the interaction term, 'HarmonyI-V-VI' is no longer significant, while the interaction term 'Musician:HarmonyI-V-VI' is significant and negative. This means that musicians and non-musicians treat Harmony I-V-VI differently when rate whether a music is popular or not. Musicians tend to give music of I-V-VI type a low popular rating while non-musicians don't.

Finally we interact 'Musician' with 'Voice'. Still none of the interaction terms is significant, so we believe musicians and non-musicians are similar in terms of the influence of 'Voice' on 'popular ratings'.

```
Random effects:
 Groups             Name        Variance Std.Dev.
 Subject:Harmony    (Intercept) 0.41090  0.6410
 Subject:Voice      (Intercept) 0.03395  0.1842
 Subject:Instrument (Intercept) 1.92271  1.3866
 Residual                       2.49061  1.5782
Number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrument, 210

Fixed effects:
                     Estimate Std. Error t value
(Intercept)           6.28827    0.23080  27.246
Instrumentpiano      -0.94842    0.24710  -3.838
Instrumentstring     -2.60618    0.24680 -10.560
HarmonyI-V-IV        -0.02554    0.14053  -0.182
HarmonyI-V-VI        -0.27151    0.14052  -1.932
HarmonyIV-I-V        -0.18562    0.14046  -1.322
Musician              0.46580    0.24704   1.886
Voicepar3rd           0.18394    0.10775   1.707
Voicepar5th           0.11869    0.10765   1.103
X16.minus.17          0.06133    0.03724   1.647
Musician:Voicepar3rd -0.05019    0.17058  -0.294
Musician:Voicepar5th  0.10886    0.17046   0.639
```

what about also interacting musician with x16.minus.17?

5.

Overall, among the three experimental factors, we find that the 'Instrument' is the most important factor that influences people's judgment on whether a music piece is classical or popular, followed next by 'Harmony'. 'Voice' has the least influence. People who declare themselves as musicians take into account 'Instrument' and 'Harmony' differently compared with those who declare themselves as non-musicians.

In our analysis, instead of a standard repeated measures model, we use a model with three random effects that captures people's intrinsic bias regarding to different types of the three factors. In both 'classical rating' and 'popular rating', the random effect of 'instrument' turns out to have the largest variance (about 1.5), followed by that of 'harmony' (about 0.2) and 'voice' (about 0.03). This indicates that the personal bias distribution regarding 'instrument type' is the most spread out one. Therefore we believe that instrument type is the most important factor that leads to the variation in people's intrinsic bias regarding whether a piece of music is classical or popular.

Besides the personal bias reflected in random effects, we gain more insights in interpreting the fixed effect coefficient estimates. Music played by guitar tends to be rated as the least classical music; in contrast, music played by piano and string tends to be associated with respectively 1.4 and 3.2 more points in classical rating compared to that played by guitar. As to 'harmony', only I-V-VI is shown to be a significant type that impact people's judgement. An I-V-VI stimuli is associated with a 0.77 more points in classical rating with everything else being equal. This is in line with the researchers' guess that I-V-VI might be frequently rated as classical due to people's familiarity with Pachelbel's Canon in D. For 'voice', we find that stimuli of par3rd and par5th are both associated with a less classical rating compared to that of the contrary motion. This is in line with the researchers' expectation that contrary motion would be frequently rated as classical. As to the 'popular rating' case, the fixed effects are mostly opposite with that in the 'classical rating' case. Music played by guitar is rated high 'popular rating' and string is rated least. I-V-VI is also rated significantly lower than other 'harmony' types. Voice types don't have significant influence on 'popular ratings'.

We further examine the difference between self-declared musicians and non-musicians in considering the three experimental factors. The results show that musicians rely their judgement less on instruments than non-musicians do. In others words, factor 'instrument' is less influential for musicians than for non-musicians. Additionally, musicians tend to be more likely to identify I-V-VI music as classical music, while non-musicians don't have this tendency. Musicians and non-musicians don't exhibit significant difference in treating 'voice' type.

Finally we find that the more capable the listener is to distinguish classical vs popular music, the more likely he will give high classical rating.

4: 16
5: 20

36