36-763 Homework 5

Jining Qin

December 10, 2013

1

```
> ratings=read.csv("ratings.csv",header=T)
```

(a)

```
Analysis of Variance Table

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

Analysis of Variance Table

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

Analysis of Variance Table

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

```
Call:
lm(formula = Classical ~ Voice + Harmony + Instrument, 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 ***
Voicepar3rd      -0.41247    0.11271  -3.660 0.000258 ***
Voicepar5th      -0.37058    0.11264  -3.290 0.001016 **
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
Instrumentpiano   1.37359    0.11298  12.158  < 2e-16 ***
Instrumentstring  3.13312    0.11230  27.899  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.297 on 2485 degrees of freedom
  (27 observations deleted due to missingness)
Multiple R-squared:  0.255,        Adjusted R-squared:  0.2529
F-statistic: 121.5 on 7 and 2485 DF,  p-value: < 2.2e-16
```

All three variables are highly significant in linear regression model. Yet we see that in the factor variables $Harmony$ and $Voice$ only one of the levels has significant different mean from other levels. I re-coded these factor variables to examine the effect of the level of interest alone.

```
> ratings$Voice1=(ratings$Voice=="contrary")
> ratings$Harmony1=(ratings$Harmony=="I-V-VI")
> reg1a2=lm(Classical~Voice1+Harmony1+Instrument,data=ratings)
> summary(reg1a2)

Call:
lm(formula = Classical ~ Voice1 + Harmony1 + Instrument, data = ratings)

Residuals:
   Min     1Q Median     3Q    Max
-6.851 -1.720 -0.088  1.758 11.520

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       3.95492    0.08995  43.970  < 2e-16 ***
Voice1TRUE        0.39166    0.09753   4.016 6.10e-05 ***
```

```
Harmony1TRUE      0.76283    0.10625   7.180 9.17e-13 ***
Instrumentpiano   1.37369    0.11292  12.165  < 2e-16 ***
Instrumentstring  3.13310    0.11225  27.913  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.296 on 2488 degrees of freedom
  (27 observations deleted due to missingness)
Multiple R-squared:  0.2549,        Adjusted R-squared:  0.2537
F-statistic: 212.8 on 4 and 2488 DF,  p-value: < 2.2e-16
```

```
> BIC(reg1a2)

[1] 11259.92

> BIC(reg1a1)

[1] 11282.84
```

The new model has a lower BIC. And all levels interested in are significant.

(b)

(i)

$$Y_j = \alpha_{0i[j]} + X_j\beta_1 + \epsilon_j, \epsilon_j \sim N(0, \sigma^2)$$
$$\alpha_{0i} = \beta_0 + \eta_i, \eta_i \sim N(0, \tau^2)$$

Where Y_j is the classical rating for a particular entryl. $i[j]$ is the corresponding subject id for entry j. α_i is the random intercept for subject i. X_j is the vector of $Harmony$, $Voice$, $Instrument$(or their corresponding re-coded variables) for entry j.

(ii)

```
> reg1b1=lmer(Classical ~ Voice+Harmony+Instrument+(1|Subject),data=ratings)
> BIC(reg1b1)

[1] 10549.73

> BIC(reg1a1)

[1] 11282.84

> display(reg1b1)
```

```
lmer(formula = Classical ~ Voice + Harmony + Instrument + (1 |
    Subject), data = ratings)
                coef.est coef.se
(Intercept)      4.34     0.19
Voicepar3rd     -0.42     0.09
Voicepar5th     -0.37     0.09
HarmonyI-V-IV   -0.03     0.11
HarmonyI-V-VI    0.77     0.11
HarmonyIV-I-V    0.05     0.11
Instrumentpiano  1.38     0.09
Instrumentstring 3.13     0.09

Error terms:
 Groups   Name        Std.Dev.
 Subject  (Intercept) 1.30
 Residual             1.89
---
number of obs: 2493, groups: Subject, 70
AIC = 10491.5, DIC = 10426.2
deviance = 10448.9
```

The multilevel model has a much lower BIC compared to the linear regression model.

```
> exactRLRT(reg1b1)

        simulated finite sample distribution of RLRT.

        (p-value based on 10000 simulated values)

data:
RLRT = 763.3759, p-value < 2.2e-16
```

The RLRT method also agrees with the multilevel model.

(iii)

```
> reg1b2 = lmer(Classical ~ Instrument + Voice + (1 | Subject),
+                 data = ratings)
> anova(reg1b2,reg1b1)

Data: ratings
Models:
reg1b2: Classical ~ Instrument + Voice + (1 | Subject)
reg1b1: Classical ~ Voice + Harmony + Instrument + (1 | Subject)
       Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
reg1b2  7 10539 10580 -5262.4    10525
```

```
reg1b1 10 10469 10527 -5224.4    10449 75.931      3  2.288e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hence *Harmony* is significant and should be included.

```
> reg1b3 = lmer(Classical ~ Harmony + Voice + (1 | Subject),
+                data = ratings)
> anova(reg1b3,reg1b1)

Data: ratings
Models:
reg1b3: Classical ~ Harmony + Voice + (1 | Subject)
reg1b1: Classical ~ Voice + Harmony + Instrument + (1 | Subject)
       Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
reg1b3  8 11408 11455 -5696.2    11392
reg1b1 10 10469 10527 -5224.4    10449 943.59      2  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hence *Instrument* is significant and should be included.

```
> reg1b4 = lmer(Classical ~ Harmony + Instrument + (1 | Subject),
+                data = ratings)
> anova(reg1b4,reg1b1)

Data: ratings
Models:
reg1b4: Classical ~ Harmony + Instrument + (1 | Subject)
reg1b1: Classical ~ Voice + Harmony + Instrument + (1 | Subject)
       Df   AIC   BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
reg1b4  8 10489 10536 -5236.6    10473
reg1b1 10 10469 10527 -5224.4    10449 24.24      2   5.45e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hence *Voice* is significant and should be included.

(c)

(i)

```
> reg1c1=lmer(Classical ~ Harmony + Instrument + Voice + (1|Subject:Harmony) +
+              (1 | Subject:Instrument)+ (1|Subject:Voice), data = ratings)
```

Model	AIC	BIC	DIC
reg1a1	11467.34	11519.82	
reg1b1	10755.46	10813.78	10690.9
reg1c1	10307.76	10377.75	10247.9

reg1c1 turns out to be the best model.

(ii)

```
> display(reg1c1)

lmer(formula = Classical ~ Harmony + Instrument + Voice + (1 |
    Subject:Harmony) + (1 | Subject:Instrument) + (1 | Subject:Voice),
    data = ratings)
                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     0.06     0.14
Instrumentpiano   1.36     0.26
Instrumentstring  3.13     0.26
Voicepar3rd      -0.41     0.08
Voicepar5th      -0.37     0.08

Error terms:
 Groups              Name        Std.Dev.
 Subject:Harmony     (Intercept) 0.67
 Subject:Voice       (Intercept) 0.17
 Subject:Instrument (Intercept) 1.48
 Residual                        1.56
---
number of obs: 2493, groups: Subject:Harmony, 280; Subject:Voice, 210; Subject:Instrumen
AIC = 10075.5, DIC = 10015.5
deviance = 10033.5
```

From the summary table for reg1c1, we see that all three variables have levels with parameter estimate greater than twice their standard error. Hence all three variables(or at least the levels interested in) are significant. The variance of $Instrument$ is comparable to that of the residual while the other two are much smaller. This might indicate other variables should be included in the model.

(iii)

$$Y_j = X_j\beta + \sum_{k=1}^{3} \alpha_{h_k[i[j]]} + \epsilon_j, \epsilon_j \sim N(0, \sigma^2)$$

$$\alpha_{h_k} = \mu_{0k} + \eta_{h_k}, \eta_{h_k} \sim N(0, \tau_k), k = 1, 2, 3$$

Y_j is the classical rating for jth entry. $i[j]$ is the corresponding subject id for entry j. α_{h_k} is the random intercept for subject i combined with a level of $Harmony$(or $Voice$ or $Instrument$). X_j is the vector of $Harmony$, $Voice$, $Instrument$ for entry j.

2

(a)

The best model is reg1c1 in 1. After testing the significance of all variables using ANOVA procedure, I ended up with this model including three more variables.

```
> reg2a=update(reg1c1, . ~. + X16.minus.17+
+       X1990s2000s.minus.1960s1970s+ GuitarPlay)
> display(reg2a)

lmer(formula = Classical ~ Harmony + Instrument + Voice + (1 |
    Subject:Harmony) + (1 | Subject:Instrument) + (1 | Subject:Voice) +
    X16.minus.17 + X1990s2000s.minus.1960s1970s + GuitarPlay,
    data = ratings)
                             coef.est coef.se
(Intercept)                   4.33     0.26
HarmonyI-V-IV                -0.03     0.15
HarmonyI-V-VI                 0.76     0.15
HarmonyIV-I-V                 0.08     0.15
Instrumentpiano               1.48     0.26
Instrumentstring              3.26     0.26
Voicepar3rd                  -0.42     0.09
Voicepar5th                  -0.37     0.09
X16.minus.17                 -0.10     0.04
X1990s2000s.minus.1960s1970s  0.00     0.06
GuitarPlay                    0.09     0.08

Error terms:
 Groups              Name        Std.Dev.
 Subject:Harmony     (Intercept) 0.67
 Subject:Voice       (Intercept) 0.18
 Subject:Instrument (Intercept) 1.39
 Residual                        1.57
---
number of obs: 2325, groups: Subject:Harmony, 260; Subject:Voice, 195; Subject:Instrumen
AIC = 9420, DIC = 9331.2
deviance = 9360.6

> BIC(reg2a)

[1] 9506.232
```

(b)

```
> reg2b0 = lmer(Classical ~ (1 | Subject:Harmony) + (1 | Subject:Instrument) +
+          (1 | Subject:Voice) + Harmony + Instrument + Voice + X16.minus.17 +
```

```
+          X1990s2000s.minus.1960s1970s + GuitarPlay,data=ratings)
> reg2bh = lmer( Classical ~ (1 | Subject:Harmony) + Harmony + Instrument +
+                  Voice + X16.minus.17 + X1990s2000s.minus.1960s1970s +
+                  GuitarPlay, data = ratings)
> reg2bh1 = update(reg2b0, . ~ . - (1 | Subject:Harmony))
> exactRLRT(reg2bh, m0 = reg2bh1, mA = reg2b0)

        simulated finite sample distribution of RLRT.

        (p-value based on 10000 simulated values)

data:
RLRT = 94.5898, p-value < 2.2e-16
```

Hence we keep the random effect for $Harmony$.

```
> reg2bi = lmer( Classical ~ (1 | Subject:Instrument) + Harmony + Instrument +
+                  Voice + X16.minus.17 + X1990s2000s.minus.1960s1970s +
+                  GuitarPlay, data = ratings)
> reg2bi1 = update(reg2b0, . ~ . - (1 | Subject:Instrument))
> exactRLRT(reg2bi, m0 = reg2bi1, mA = reg2b0)

        simulated finite sample distribution of RLRT.

        (p-value based on 10000 simulated values)

data:
RLRT = 498.568, p-value < 2.2e-16
```

Hence we keep the random effect for $Instrument$.

```
> reg2bv = lmer( Classical ~ (1 | Subject:Voice) + Harmony + Instrument +
+                  Voice + X16.minus.17 + X1990s2000s.minus.1960s1970s +
+                  GuitarPlay, data = ratings)
> reg2bv1 = update(reg2b0, . ~ . - (1 | Subject:Voice))
> exactRLRT(reg2bv, m0 = reg2bv1, mA = reg2b0)

        simulated finite sample distribution of RLRT.

        (p-value based on 10000 simulated values)

data:
RLRT = 1.3049, p-value = 0.1195
```

Hence we don't keep the random effect for $Voice$.

```
> reg2b=lmer( Classical ~ (1 | Subject:Harmony) +(1 | Subject:Instrument) + Harmony +
+                  Voice + X16.minus.17 + X1990s2000s.minus.1960s1970s +
+                  GuitarPlay, data = ratings)
> display(reg2b)

lmer(formula = Classical ~ (1 | Subject:Harmony) + (1 | Subject:Instrument) +
    Harmony + Instrument + Voice + X16.minus.17 + X1990s2000s.minus.1960s1970s +
    GuitarPlay, data = ratings)
                             coef.est coef.se
(Intercept)                   4.33     0.26
HarmonyI-V-IV                -0.03     0.15
HarmonyI-V-VI                 0.76     0.15
HarmonyIV-I-V                 0.08     0.15
Instrumentpiano               1.47     0.26
Instrumentstring              3.26     0.26
Voicepar3rd                  -0.42     0.08
Voicepar5th                  -0.36     0.08
X16.minus.17                 -0.10     0.04
X1990s2000s.minus.1960s1970s  0.00     0.06
GuitarPlay                    0.09     0.08

Error terms:
 Groups              Name        Std.Dev.
 Subject:Harmony     (Intercept) 0.67
 Subject:Instrument (Intercept) 1.40
 Residual                        1.58
---
number of obs: 2325, groups: Subject:Harmony, 260; Subject:Instrument, 195
AIC = 9419.3, DIC = 9332
deviance = 9361.6
```

(c)

The parameter for each level of fixed effects is the change compared to baseline case of being in that certain level. The random effects are the corresponding group mean given subject and the factor variable(harmony or instrument).

3

```
> table(ratings$Selfdeclare)

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

```
> ratings$Selfdeclare1=(ratings$Selfdeclare>2)
> reg3=lmer(Classical ~ Selfdeclare1+Selfdeclare1:Harmony+Selfdeclare1:X16.minus.17 +
+          Selfdeclare1:X1990s2000s.minus.1960s1970s + Selfdeclare1:GuitarPlay+
+             Selfdeclare1:Instrument+Selfdeclare1:Voice+ X16.minus.17 +
+             X1990s2000s.minus.1960s1970s + GuitarPlay+Instrument+Harmony+
+             Voice+(1|Subject:Harmony)+(1|Subject:Instrument),data=ratings)
> BIC(reg3)

[1] 9557.078

> display(reg3)

lmer(formula = Classical ~ Selfdeclare1 + Selfdeclare1:Harmony +
    Selfdeclare1:X16.minus.17 + Selfdeclare1:X1990s2000s.minus.1960s1970s +
    Selfdeclare1:GuitarPlay + Selfdeclare1:Instrument + Selfdeclare1:Voice +
    X16.minus.17 + X1990s2000s.minus.1960s1970s + GuitarPlay +
    Instrument + Harmony + Voice + (1 | Subject:Harmony) + (1 |
    Subject:Instrument), data = ratings)
                                            coef.est coef.se
(Intercept)                                  4.13     0.31
Selfdeclare1TRUE                             0.61     0.61
X16.minus.17                                -0.02     0.05
X1990s2000s.minus.1960s1970s                 0.01     0.07
GuitarPlay                                   1.26     0.38
Instrumentpiano                              1.64     0.32
Instrumentstring                             3.46     0.32
HarmonyI-V-IV                               -0.04     0.18
HarmonyI-V-VI                                0.31     0.18
HarmonyIV-I-V                                0.03     0.18
Voicepar3rd                                 -0.40     0.10
Voicepar5th                                 -0.29     0.10
Selfdeclare1TRUE:HarmonyI-V-IV               0.03     0.29
Selfdeclare1TRUE:HarmonyI-V-VI               1.18     0.29
Selfdeclare1TRUE:HarmonyIV-I-V               0.14     0.29
Selfdeclare1TRUE:X16.minus.17               -0.22     0.09
Selfdeclare1TRUE:X1990s2000s.minus.1960s1970s -0.05     0.14
Selfdeclare1TRUE:GuitarPlay                 -1.18     0.39
Selfdeclare1TRUE:Instrumentpiano            -0.44     0.52
Selfdeclare1TRUE:Instrumentstring           -0.51     0.52
Selfdeclare1TRUE:Voicepar3rd                -0.05     0.17
Selfdeclare1TRUE:Voicepar5th                -0.19     0.17

Error terms:
 Groups             Name        Std.Dev.
 Subject:Harmony    (Intercept) 0.61
```

```
 Subject:Instrument (Intercept) 1.37
 Residual                       1.58
---
number of obs: 2325, groups: Subject:Harmony, 260; Subject:Instrument, 195
AIC = 9413.3, DIC = 9278.7
deviance = 9321.0
```

The variable $selfdeclare1$(equal to 1 if self-declare is greater than 2 and 0 otherwise) isn't significant itself, yet several interaction terms are significant. This implies that self-identification as a musician does change people's pattern of classical music controlling for the other variables in this model.

4

(a)

```
> reg4a1=lm(Popular ~ Voice+Harmony+Instrument,data=ratings)
> reg4av=lm(Popular ~ Harmony+Instrument,data=ratings)
> reg4ai=lm(Popular ~ Voice+Harmony,data=ratings)
> reg4ah=lm(Popular ~ Voice+Instrument,data=ratings)
> anova(reg4a1,reg4av)

Analysis of Variance Table

Model 1: Popular ~ Voice + Harmony + Instrument
Model 2: Popular ~ Harmony + Instrument
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1   2485 12656
2   2487 12672 -2   -15.263 1.4984 0.2237

> anova(reg4a1,reg4ai)

Analysis of Variance Table

Model 1: Popular ~ Voice + Harmony + Instrument
Model 2: Popular ~ Voice + Harmony
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1   2485 12656
2   2487 15580 -2   -2923.9 287.05 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(reg4a1,reg4ah)
```

```
Analysis of Variance Table

Model 1: Popular ~ Voice + Harmony + Instrument
Model 2: Popular ~ Voice + Instrument
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1   2485 12656
2   2488 12688 -3   -31.092 2.0349 0.1069
```

Only instrument is significant here and we leave other fixed effects out.

need to leave experimental factors in the model

```
> reg4a2=lmer(Popular ~ Instrument+(1|Subject:Voice)+(1|Subject:Harmony)+(1|Subject:In
> display(reg4a2)

lmer(formula = Popular ~ Instrument + (1 | Subject:Voice) + (1 |
    Subject:Harmony) + (1 | Subject:Instrument), data = ratings)
                coef.est coef.se
(Intercept)      6.57     0.18
Instrumentpiano -0.95     0.25
Instrumentstring -2.61    0.25

Error terms:
 Groups             Name        Std.Dev.
 Subject:Harmony    (Intercept) 0.65
 Subject:Instrument (Intercept) 1.41
 Subject:Voice      (Intercept) 0.20
 Residual                       1.58
---
number of obs: 2493, groups: Subject:Harmony, 280; Subject:Instrument, 210; Subject:Voic
AIC = 10083.9, DIC = 10060.4
deviance = 10065.2

> reg4a2v=update(reg4a2, . ~ .-(1|Subject:Voice))
> reg4a2i=update(reg4a2, . ~ .-(1|Subject:Instrument))
> reg4a2h=update(reg4a2, . ~ .-(1|Subject:Harmony))
> anova(reg4a2,reg4a2v)

Data: ratings
Models:
reg4a2v: Popular ~ Instrument + (1 | Subject:Harmony) + (1 | Subject:Instrument)
reg4a2: Popular ~ Instrument + (1 | Subject:Voice) + (1 | Subject:Harmony) +
reg4a2:     (1 | Subject:Instrument)
        Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
reg4a2v  6 10079 10114 -5033.5    10067
reg4a2   7 10079 10120 -5032.6    10065 1.8017      1     0.1795

> anova(reg4a2,reg4a2i)
```

```
Data: ratings
Models:
reg4a2i: Popular ~ Instrument + (1 | Subject:Voice) + (1 | Subject:Harmony)
reg4a2: Popular ~ Instrument + (1 | Subject:Voice) + (1 | Subject:Harmony) +
reg4a2:     (1 | Subject:Instrument)
        Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
reg4a2i  6 10571 10606 -5279.3    10559
reg4a2   7 10079 10120 -5032.6    10065 493.39      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> anova(reg4a2,reg4a2h)

Data: ratings
Models:
reg4a2h: Popular ~ Instrument + (1 | Subject:Voice) + (1 | Subject:Instrument)
reg4a2: Popular ~ Instrument + (1 | Subject:Voice) + (1 | Subject:Harmony) +
reg4a2:     (1 | Subject:Instrument)
        Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
reg4a2h  6 10169 10204 -5078.5    10157
reg4a2   7 10079 10120 -5032.6    10065 91.913      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> reg4a0=lmer(Popular ~ Instrument+(1|Subject:Harmony)+(1|Subject:Instrument),data=rat
> display(reg4a0)

lmer(formula = Popular ~ Instrument + (1 | Subject:Harmony) +
    (1 | Subject:Instrument), data = ratings)
                coef.est coef.se
(Intercept)      6.57     0.18
Instrumentpiano -0.95     0.25
Instrumentstring -2.61     0.25

Error terms:
 Groups             Name        Std.Dev.
 Subject:Harmony    (Intercept) 0.65
 Subject:Instrument (Intercept) 1.42
 Residual                       1.59
---
number of obs: 2493, groups: Subject:Harmony, 280; Subject:Instrument, 210
AIC = 10083.7, DIC = 10062.3
deviance = 10067.0
```

The random effects of *Instrument* and *Harmony* are significant, which we would leave in.

(b)

```
> reg4b0=lmer(Popular ~ Instrument+(1|Subject:Harmony)+(1|Subject:Instrument),data=rat
```

By adding single variables and comparing BIC, I ended up with the following model:

need to be sure no cases were dropped because of NA's, which would change the sample size and invalidate BIC comparisons.

```
> reg4b1=lmer(Popular ~ (1 | Subject:Harmony) + (1 | Subject:Instrument)+
+         Harmony + Instrument + Voice  +OMSI+
+         X1990s2000s.minus.1960s1970s,data = ratings)
> display(reg4b1)

lmer(formula = Popular ~ (1 | Subject:Harmony) + (1 | Subject:Instrument) +
    Harmony + Instrument + Voice + OMSI + X1990s2000s.minus.1960s1970s,
    data = ratings)
                             coef.est coef.se
(Intercept)                   6.31     0.27
HarmonyI-V-IV                -0.03     0.15
HarmonyI-V-VI                -0.29     0.15
HarmonyIV-I-V                -0.23     0.15
Instrumentpiano              -0.99     0.26
Instrumentstring             -2.63     0.26
Voicepar3rd                   0.19     0.08
Voicepar5th                   0.18     0.08
OMSI                          0.00     0.00
X1990s2000s.minus.1960s1970s  0.01     0.06

Error terms:
 Groups              Name        Std.Dev.
 Subject:Harmony     (Intercept) 0.65
 Subject:Instrument (Intercept) 1.39
 Residual                        1.59
---
number of obs: 2325, groups: Subject:Harmony, 260; Subject:Instrument, 195
AIC = 9454.8, DIC = 9358.2
deviance = 9393.5
```

One level of harmony is significant. Instrument is significant and voice is significant in the model.

(c)

```
> reg4c=lmer(Popular ~ (1 | Subject:Harmony) + (1 | Subject:Instrument)+
+         Selfdeclare1:Harmony + Selfdeclare1:Instrument + Selfdeclare1:Voice +
+         Selfdeclare1 +Selfdeclare1:OMSI+Selfdeclare1:X1990s2000s.minus.1960s1970s
+         + Harmony + Instrument + Voice + Selfdeclare1 +OMSI+
+         X1990s2000s.minus.1960s1970s     ,data = ratings)
> display(reg4c)
```

```
lmer(formula = Popular ~ (1 | Subject:Harmony) + (1 | Subject:Instrument) +
    Selfdeclare1:Harmony + Selfdeclare1:Instrument + Selfdeclare1:Voice +
    Selfdeclare1 + Selfdeclare1:OMSI + Selfdeclare1:X1990s2000s.minus.1960s1970s +
    Harmony + Instrument + Voice + Selfdeclare1 + OMSI + X1990s2000s.minus.1960s1970s,
    data = ratings)
                                             coef.est coef.se
(Intercept)                                   6.44     0.38
Selfdeclare1TRUE                             -0.18     0.63
HarmonyI-V-IV                                -0.10     0.18
HarmonyI-V-VI                                 0.00     0.18
HarmonyIV-I-V                                -0.22     0.18
Instrumentpiano                              -1.15     0.33
Instrumentstring                             -2.90     0.33
Voicepar3rd                                   0.22     0.10
Voicepar5th                                   0.12     0.10
OMSI                                          0.00     0.00
X1990s2000s.minus.1960s1970s                  0.00     0.07
Selfdeclare1TRUE:HarmonyI-V-IV                0.18     0.30
Selfdeclare1TRUE:HarmonyI-V-VI               -0.77     0.30
Selfdeclare1TRUE:HarmonyIV-I-V               -0.02     0.30
Selfdeclare1TRUE:Instrumentpiano              0.42     0.53
Selfdeclare1TRUE:Instrumentstring             0.70     0.53
Selfdeclare1TRUE:Voicepar3rd                 -0.09     0.17
Selfdeclare1TRUE:Voicepar5th                  0.16     0.17
Selfdeclare1TRUE:OMSI                         0.00     0.00
Selfdeclare1TRUE:X1990s2000s.minus.1960s1970s 0.03     0.13

Error terms:
 Groups              Name        Std.Dev.
 Subject:Harmony     (Intercept) 0.62
 Subject:Instrument (Intercept) 1.41
 Residual                        1.59
---
number of obs: 2325, groups: Subject:Harmony, 260; Subject:Instrument, 195
AIC = 9477.2, DIC = 9322.8
deviance = 9377.0
```

Still, selfdeclare itself isn't significant yet quite some interaction terms are significant. Meaning change in features of the music have different influence on a person's rating of the popularity of that piece of music whether or not this person identify himself as a musician.

5

5.1 Using classical as dependent variable

I started with multiple linear regressions and included all three fixed effects in it. Analysis of variance confirms that all three effects are significant, while from the regression table (in (1)) we see that certain levels of the variables(conveniently the levels mentioned in his main hypotheses) are significant. Recoding these levels in the linear regression model leads to a model with better BIC. I then added the random intercept for each person, which turned out to be helpful to the model. Then I included the "personal biases" in the model and found out that all three variables combined with person should be included. The variance component isn't a standard variance model since the residual variance is still higher than two random effects. We hence need more covariates in the model. With further anova testing I included $GuitarPlay$, $X16.minus.17$, $X1990s2000s.minus.1960s1970s$ in the model. And judging from the variance components, variance of the residual is roughly comparable to that of random effects.

Also, after recoding the self-declare variable into two levels with roughly equal subjects, we find that the variable itself isn't significant yet some of its iteractions with other variables are. which indicates that self-identified as musician produces another layer of bias in rating a piece of music as classical. To be specific, the interaction of self-declare above 2 and Harmony level "IV-V-VI", $X16.minus.17$ and $PlayGuitar$ are significant.

5.2 Using popular as dependent variable

As before, I started with linear models and found only instrument is significant. I then included the personal random intercept and found the model improved in that its BIC fell. Then the three "personal bias" random effects are tested and found instrument and harmony significant. Variance of the residual is comparable to one of the random effects and bigger than the other one, implying need for more variables. By further testing I added $OMIS$,$X1990s2000s.minus.1960s1970s$ to the model.

The recoded $selfdeclare$ variable is still not significant itself. Yet its interaction with $Harmony$ level "I-V-VI" is significant.

4: 16
5: 20

36