1 a 9/9
b 9/9
c 8/9

2 a 9/9
b 8/9
c 9/9

3 8/9

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

5 10/10

Total 95/100

36-763 Homework 5

Carmen Khoo

due Friday, December 18, 2015

Section 1: Three main experimental factors

Part (a): Linear Models

We want to examine the influence of the three main experimental factors, `Instrument`, `Harmony` and `Voice`. We take into consideration the hypotheses that the type of instrument should have the largest influence on rating and that the I-V-vi harmonic progression and contrary motion in voice leading are frequently associated with classical music.

Model 1: Harmony, Instrument, Voice

We first fit the full model which includes all our variables of interest. For `Harmony`, we can expect a 0.77 increase in the rating given that it is a I-V-vi progression compared to the baseline I-VI-V. There is an expected decrease of 0.031 in rating for the I-V-IV progression and an expected increase of 0.05 for the IV-I-V progression, both of which are relatively small changes in ratings.

For `Instrument`, we can expect a 1.37 increase in rating when the instrument is piano and a 3.13 increase when the it is a string quartet and this is unsurprising when compared to the baseline instrument, the electric guitar, since strings are often used in orchestras while the piano is versatile and commonly used in classical pieces throughout the ages.

For `Voice`, contrary motion is the baseline and as hypothesized, it is more likely to be associated with classical music since there is an expected decrease of 0.412 in the ratings when the stimulus is in parallel thirds and an expected decrease of 0.371 when it is in parallel fifths.

An ANOVA test informs us that each of the three variables are statistically significant predictors of Classical ratings. The diagnostic plots show that the regression assumptions of zero expectation, equal variance and normality are reasonably well-met. We now compare the full model to reduced models.

Model 2: Instrument, Voice

We remove `Harmony` from the model and examine the influence of `Instrument` and `Voice` on Classical ratings in the absence of harmonic progression. The estimated coefficients remain the same in terms of magnitude and direction for both the main factors. From the ANOVA test, both variables are significant predictors of Classical ratings. Regression assumptions seem reasonably well-met.

Model 3: Harmony, Voice

Without `Instrument` in the model, we find that, similar to Model 2, the magnitude and direction of the estimated coefficients remain the same. Our ANOVA test also tells us that both predictors are significant. The normality assumption appears mildly violated by a heavy upper tail.

Model 4: Harmony, Instrument

We obtain the same results in terms of magnitude and direction of the estimated coefficients as well as significance of variables from the ANOVA test. Regression assumptions are reasonably well-met.

We compare our models:

Model	AIC	BIC	Adjusted R^2
1 (Instrument, Harmony, Voice)	11230.45	11282.84	0.253
2 (Instrument, Voice)	11275.96	11310.89	0.238
3 (Harmony, Voice)	11908.94	11949.69	0.0184
4 (Harmony, Instrument)	11242.69	11283.43	0.249

Table 1: Comparison of Linear Regression Models

9 Since Models 2, 3 and 4 are reduced versions of Model 1 and therefore nested, we can compare them with the full model by their deviances. Our results are in the table below:

Models	Residual Df	p-value
1 vs 2	2485 vs 2488	< 0.0005
1 vs 3	2485 vs 2487	< 0.0005
1 vs 4	2485 vs 2487	< 0.0005

Table 2: Results from ANOVA

The p-values **reject the null hypothesis** which claims the reduced model fits the data well, and are **in favor of the full model** which includes all three main experimental factors.

Part (b)(i): Mathematical Notation

For this model, we fit a random intercept for each participant. We write this model as follows:

$$y_i = \alpha_{j[i]} + +\beta_{H[m]} + \beta_{I[m]} + \beta_{V[m]} + \epsilon_i$$

where m represents the level for the factors. The varying-intercept can be written as

$$\alpha_{j[i]} = \beta_0 + \eta_j$$

where $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ and $\eta_j \overset{iid}{\sim} N(0, \tau^2)$.

Part (b)(ii): Random Intercept per Subject

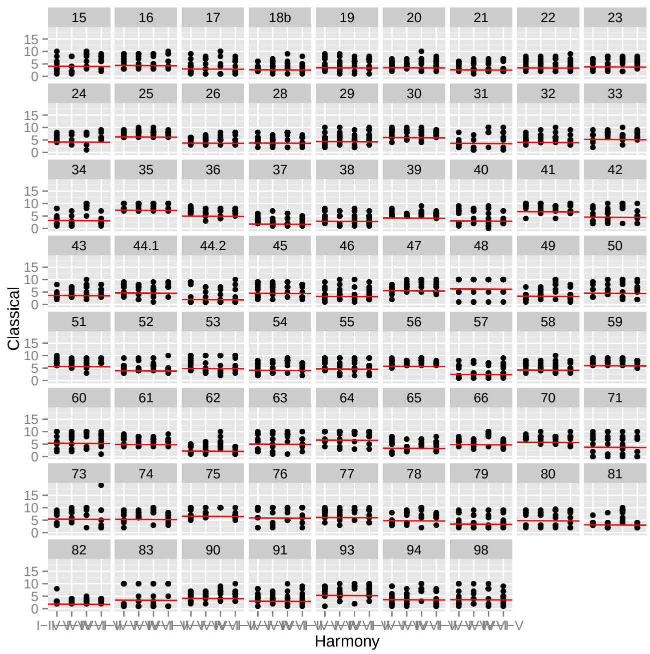
Model 5: Harmony, Instrument, Voice and Random Intercept per Subject

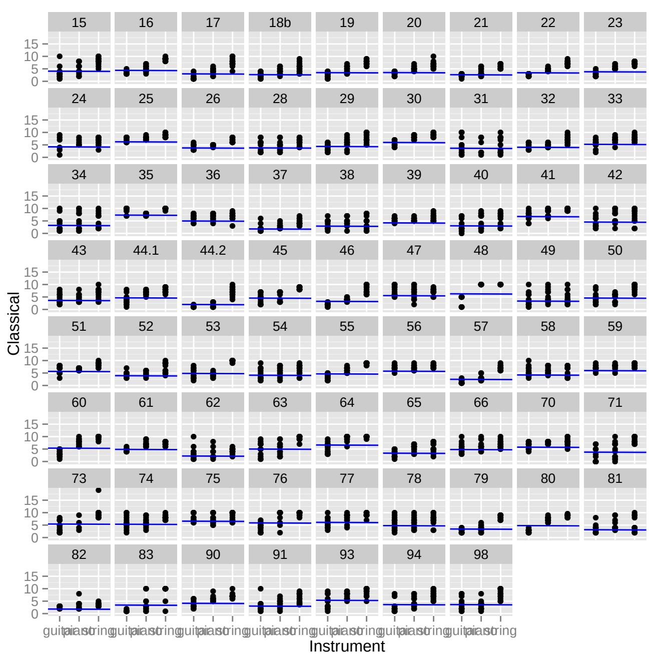
In order to determine whether the random intercept is required in the model, we compare the models with and without the random intercept by AIC and BIC.

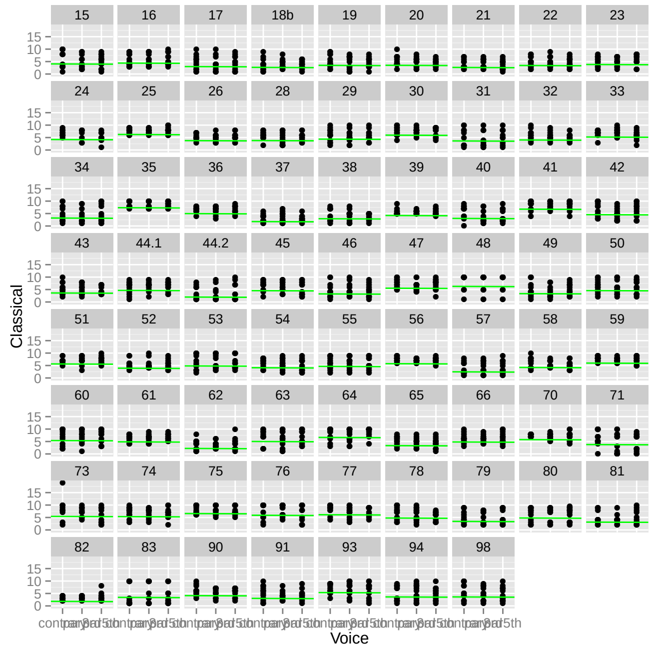
Model	AIC	BIC
1 (Without random intercept)	11230.45	11282.84
5 (With random intercept)	10491.5	10549.73

There is a large reduction of more than 700 in the AIC and BIC when we fit a varying-intercept model. Our results suggest that a model with a random intercept for each participant is more appropriate than one without.

We now plot Classical ratings with the estimated regression for the participants. Note that the unique subject ID starts at 15, has missing IDs (27, 67, 68, 69, 72, etc) and contains unexpected values, namely 18b, 44.1 and 44.2.







For all the factors, each subject has a slightly different intercept. Of the three faceted plots, `Harmony` shows greatest intercept-variation per subject.

We calculate the intraclass correlation coefficient as follows:

Groups	Variance	Std Dev
Subject	1.702	1.305
Residual	3.581	1.892

Table 3: Variance Component and Residual Variance

$$\begin{aligned}\rho &= \frac{\tau^2}{\tau^2+\sigma^2}\\ &= \frac{1.702}{1.702+3.581}\\ &= 0.322\end{aligned}$$

The random intercept accounts for about 32% of the overall variation and as such we would consider it justified to include a per-subject random intercept in our model. Finally, we perform a likelihood ratio test on our per-subject intercept, and obtain the following results:

	Chi-square	df	p-value
Subject	763	1	< 0.0005

Table 4: LRT for `Subject` Random Intercept

This is a method I used in class to motivate the random effects, but it is not really a test of the need for a random effect.

Based on the three methods above, we find that the random effect is needed in the model.

9

Part(b)(iii)

We reexamine the influence of our three primary experimental factors on Classical ratings based on the random-intercept model we fit in (b). There are no changes in the direction or magnitude of the effect of the factors on ratings compared what we observed when fitting the linear model.

We perform an exact likelihood ratio test to compare the linear model from the linear mixed model using simulated values. For this test, the model under the alternative hypothesis is the linear mixed model. We obtain an **observed likelihood ratio** of 763.58 and p-value $<$ **0.0005**. As such, we reject the null hypothesis in favor of the mixed model with the random intercept.

Our results above are further emphasized by the AIC and BIC:

Model	AIC	BIC
1	11230.45	11282.84
5	10491.51	10549.73

Table 5: M1 - Linear Model, M5 - Mixed Model

In both instances, the AIC and BIC support our findings that even though the estimated coefficients of the fixed effects are not very different between both models, the mixed effect model is a better fit.

Part (c)(i)

Model 6: Random Effects for Main Experimental Factors

We extend the table above to include the new model which has three new random effects term.

Model	AIC	BIC
1	11230.45	11282.84
5	10491.5	10549.73
6	10075.5	10145.37

Table 6: AIC and BIC Comparison

To determine the usefulness of the three random effects, we perform a likelihood ratio test and our results are as follows:

	Chi-square	df	p-value
Subject:Harmony	101.93	1	< 0.0005
Subject:Instrument	560.68	1	< 0.0005
Subject:Voice	0.999	1	0.3

This is a method I used in class to motivate the random effects, but it is not really a test of the need for a random effect.

Table 7: LRT for Three Random Intercepts

Even though the random effect for the Subject-Voice combination is not significant on its own, the model as a whole appears to improve the fit of the data compared to models in 1(a) and 1(b).

Part (c)(ii)

As expected, there are no changes in the magnitude and direction of the estimated coefficients of the fixed effects since the estimates are the average changes.

Since we want to know the effect on the *fixed effects*, we can run an ANOVA test to compare the model with the subject-only random intercept with our current model with three new random effects.

Model	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
5	10	10468.86	10527.07	-5224.43	10448.86			
6	12	10057.53	10127.38	-5016.76	10033.53	415.33	2	< 0.0005

Table 8: ANOVA for M5 (null) vs M6 (alternative)

The results in the table above tell us to reject the null hypothesis in favor of the model with the three random intercepts. The table below shows the estimated variance components:

Groups	Variance	Std Dev
Subject:Harmony	0.43285	0.6579
Subject:Instrument	2.16929	1.4729
Subject:Voice	0.02473	0.1573
Residual	2.43721	1.5612

Table 9: Variance Components and Residual Variance

Of the three intercepts, the greatest variation is found within the subject-instrument combination, and the smallest within the subject-voice combination. The small variation is probably for reason for which the random intercept is not statistically significant under the likelihood ratio test. The estimated residual variance informs us of each subject's Classical ratings around the individual regression line for each subject. To verify, we calculate all three intraclass correlation coefficients.

$$\begin{aligned}\rho_H^2 &= \frac{\tau_H^2}{\tau_H^2 + \sigma^2} \\ &= \frac{0.4329}{0.4329 + 2.4372} \\ &= 0.15 \\ \rho_I^2 &= \frac{\tau_I^2}{\tau_I^2 + \sigma^2} \\ &= \frac{2.1693}{2.1693 + 2.4372} \\ &= 0.47 \\ \rho_V^2 &= \frac{\tau_V^2}{\tau_V^2 + \sigma^2} \\ &= \frac{0.02473}{0.02473 + 2.4372} \\ &= 0.01\end{aligned}$$

We consider correlations of greater than 0.05, or 5% to be somewhat important and we find that the subject-harmony random explains about 15% of the overall variation whereas the subject-instrument random effect accounts for nearly half the overall variation.

8

Part (c)(iii)

We write the model in mathematical notation as follows:

$$y_i = \alpha^H_{j[i],k[i]} + \alpha^I_{j[i],k[i]} + \alpha^V_{j[i],k[i]} + \epsilon_i$$

Right idea, but need different index letters for levels of the different experimental factors.

$$\alpha^H_{j[i],k[i]} = \beta_{H[m]} + \eta^H_{jk}$$

The index on each beta should be a level of the corresponding exper factor.

$$\alpha^I_{j[i],k[i]} = \beta_{I[m]} + \eta^I_{jk}$$

$$\alpha^V_{j[i],k[i]} = \beta_{V[m]} + \eta^V_{jk}$$

where

$$\epsilon \overset{iid}{\sim} N(0, \sigma^2)$$

$$\eta^H_{jk} \overset{iid}{\sim} N(0, \tau^2_H)$$

$$\eta^I_{jk} \overset{iid}{\sim} N(0, \tau^2_I)$$

$$\eta^V_{jk} \overset{iid}{\sim} N(0, \tau^2_V)$$

where m represents the level for the factors.

Section 2: Individual covariates

Part (a): Other Influential Covariates

We want to examine other covariates which may provide insight into Classical ratings. Upon perusal of the data set, we consider the variables `Selfdeclare` (whether participant is a musician), `OMSI` (score on test of musical knowledge) and `ClsListen` (how much classical music the participant listens to). We extend Model 1 and examine the results.

Model 7: Harmony, Instrument, Voice, Selfdeclare, OMSI, ClsListen

Model 8: Harmony, Instrument, Voice, Selfdeclare

Model 9: Harmony, Instrument, Voice, OMSI

Model 10: Harmony, Instrument, Voice, ClsListen

Model Comparison

Model	AIC	BIC	Adjusted R^2
1 (Instrument, Harmony, Voice)	11230.45	11282.84	0.253
7	9978.76	10106.62	0.286
8	10077.37	10176.34	0.264
9	10090.87	10166.54	0.253
10	9970.96	10063.94	0.262

Table 10: Comparison of Extended Linear Models

When we compare Models 7 through 10, which are the extended models, we find that Model 7 which incorporates all three new covariates is the best fit and the result is consistent by both AIC and BIC. Model 7 appears to be a better fit than Model 1 as well. However, performing an analysis of variance on each of the extended models tells us that the `OMSI` test score is not a significant predictor of Classical ratings, as shown in the table below.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Harmony	3	274.39	91.46	17.34	< 0.0005
Instrument	2	4127.13	2063.57	391.11	< 0.0005
Voice	2	85.64	42.82	8.12	0.0003
OMSI	1	1.40	1.40	0.26	0.6071
Residuals	2484	13106.08	5.28		

Table 11: ANOVA of Model 8

As such, we fit one more linear model to compare it with Models 1 and 7.

Model 11: Harmony, Instrument, Voice, Selfdeclare, ClsListen

Model	AIC	BIC
1	11230.45	11282.84
7	9978.76	10106.62
11	9967.05	10089.09

Table 12: Comparison of Extended Linear Models

The tables below show the results for Model 11.

	Estimates
(Intercept)	4.66
HarmonyI-V-IV	-0.03
HarmonyI-V-VI	0.78
HarmonyIV-I-V	0.05
Instrumentpiano	1.37
Instrumentstring	3.15
Voicepar3rd	-0.40
Voicepar5th	-0.36
factor(Selfdeclare)2	-0.99
factor(Selfdeclare)3	-0.56
factor(Selfdeclare)4	-0.92
factor(Selfdeclare)5	-2.02
factor(Selfdeclare)6	-1.41
factor(ClsListen)1	-0.03
factor(ClsListen)3	0.65
factor(ClsListen)4	1.40
factor(ClsListen)5	0.90

Table 13: Fixed Effects

We interpret the effects of the variables below (part (c)).

Groups	Variance	Std Dev
Subject:Harmony	0.44881	0.6699
Subject:Instrument	2.14111	1.4633
Subject:Voice	0.02729	0.1652
Residual	2.42233	1.5564

Table 14: Variance Components and Residual Variance

The variance for the random intercepts are all smaller than the residual variance. However, there is less variation for subject-harmony and subject-voice combinations than subject-instrument combinations since the size of its variance is close to that of the residual variance.

Part (b): Random Effects

Based on Model 11, we now check the random effects using the likelihood ratio test.

	Chi-square	df	p-value
Subject:Harmony	103.86	1	< 0.0005
Subject:Instrument	555.09	1	< 0.0005
Subject:Voice	0.957	1	0.3

Again, not a formal test of the random effects

8

Table 15: LRT for Three Random Intercepts

We observe changes in the values of the chi-square statistic, but not in the p-values.

Part (c): Interpretation

For the three main experimental factors, their effects on Classical ratings remain unchanged. For the variable `Selfdeclare`, Classical ratings decrease when each level at which a participant declares himself or herself to be a musician is compared to the baseline (not at all a musician). In other words, Classical ratings decrease
9 for when a participant considers himself or herself a musician compared to ratings when the participant is not at all a musician. However, there is no order to the effect in the sense that Classical ratings are not more likely to decrease for a level-6 musician than a level-5 musician.

For `ClsListen`, ratings typically increase as the frequency to which participants listen to classical music increases. The table above shows positive coefficients for levels 3 to 5 and a negative for level 1. Since level 2 is our baseline, we find that ratings decrease for someone who does not listen to classical music at all (level 1) compared to someone who listens to at least some classical music (level 2). As before, there is no uniform increase in the ratings when moving from one level to another.

Section 3: Musicians vs Non-musicians

We want to recode the variable `Selfdeclare` such that half the participants are categorized as musicians while half of them are not. Given that there are 6 levels, where 1 indicates that the participant is not at all a musician, we may consider grouping the first three levels into the "No" group and the latter three into the "Yes" group.

Level	1	2	3	4	5	6
No. of Participants	576	936	468	432	72	36

By the frequency table above, we instead consider grouping levels 1 and 2 into the "No" group and levels 3 to 6 into the "Yes" group so that the number of participants in the two groups would be more balanced. We create the indicator variable `Musician` which takes the value 1 if the participant declares himself or herself to be a musician (levels 3-6) and 0 otherwise (levels 1-2).

Musician	0=No	1=Yes
No. of Participants	1512	1008

There aren't that many participants in the data set so you must mean something else...

We fit various models: (1) with main effects of the `Musician` indicator, (2) interaction between Harmony and Musician, (3) interaction between Instrument and Musician and (4) interaction between Voice and Musician. We compare the models as follows:

Model	AIC	BIC
1	9973.74	10072.54
2	9957.44	10073.67
3	9974.52	10084.94
4	9980.59	10091.01

8 Table 16: Model Comparison by AIC & BIC

The AIC prefers Model 2 whereas the BIC informs us that Models 1 and 2 are comparable. We thus compare the models using ANOVA:

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
1	17	9955.24	10054.03	-4960.62	9921.24			
2	20	9935.64	10051.88	-4947.82	9895.64	25.59	3	< 0.0005

Table 17: ANOVA for M1 vs M2

We find that there is a significant interaction between harmonic progression and whether or not a participant considers himself or herself to be a musician.

	Estimates
(Intercept)	4.27
HarmonyI-V-IV	-0.05
HarmonyI-V-VI	0.29
HarmonyIV-I-V	0.02
Instrumentpiano	1.37
Instrumentstring	3.15
Voicepar3rd	-0.40
Voicepar5th	-0.36
factor(ClsListen)1	-0.00
factor(ClsListen)3	0.33
factor(ClsListen)4	-0.01
factor(ClsListen)5	0.71
factor(Musician)1	-0.42
HarmonyI-V-IV:factor(Musician)1	0.05
HarmonyI-V-VI:factor(Musician)1	1.21
HarmonyIV-I-V:factor(Musician)1	0.08

Table 18: Estimates from Model 2 (Harmony*Musician Interaction)

Since we are interested in examining the interaction between harmonic progression and whether or not a participant self-declares as a musician, we analyze the interaction terms. Note that the baseline harmonic progression is I-VI-V. If the participant is a musician, then we find that there is no change in Classical ratings when the harmony played is I-V-IV compared to when it is the I-VI-V progression. However, for the other harmonic progressions, Classical ratings are expected to increase. The expected increase in ratings for the I-V-vi progression is 1.5 whereas the expected increase for the IV-I-V progression is 0.1. Our results are probably unsurprising since we hypothesized that the I-V-vi progression is often rated as classical.

Section 4: Classical vs Popular

Part (a)

Model 1: Harmony, Instrument, Voice and Random Intercept per Participant

We fit the *three main experimental factors* with a *random intercept for each participant* to analyze the effects on Popular ratings. The estimated coefficients of the fixed effects are as follows:

	Estimate	Std. Error	df	t value	$\Pr(>\lvert t\rvert)$
(Intercept)	6.57768	0.18347	137.70	35.851	< 0.0005
HarmonyI-V-IV	-0.02495	0.10646	2416.20	-0.234	0.8147
HarmonyI-V-VI	-0.27228	0.10645	2416.10	-2.558	0.0106
HarmonyIV-I-V	-0.18616	0.10636	2416.10	-1.750	0.0802
Instrumentpiano	-0.94545	0.09255	2416.80	-10.216	< 0.0005
Instrumentstring	-2.60669	0.09194	2416.50	-28.352	< 0.0005
Voicepar3rd	0.17050	0.09224	2416.10	1.848	0.0647
Voicepar5th	0.16517	0.09218 2416.10000	1.792	0.0733	

Table 19: Model Summary for Fixed Effects

In comparison to the baseline harmonic progression I-VI-V, we expect a decrease in Popular ratings when the sound stimulus presented to the participant is one of the other three harmonic progressions. For instrument where the baseline is the electric guitar, we expect a decrease in Popular ratings when the instrument in use is either piano or strings. As for vocals, the expected increase in Popular ratings for when the voice leading is in parallel thirds is similar to the expected increase when the voice is in parallel fifths; both increase ratings in comparison to contrary motion. We also check the random effects:

Group	Chi-square	df	p-value
Subject	715	1	< 0.0005

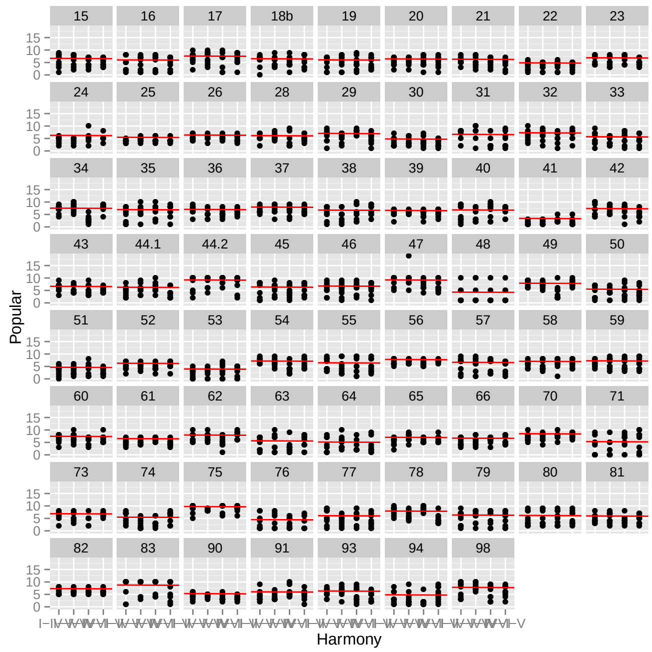
Groups	Variance	Std Dev
Subject	1.566	1.251
Residual	3.532	1.879

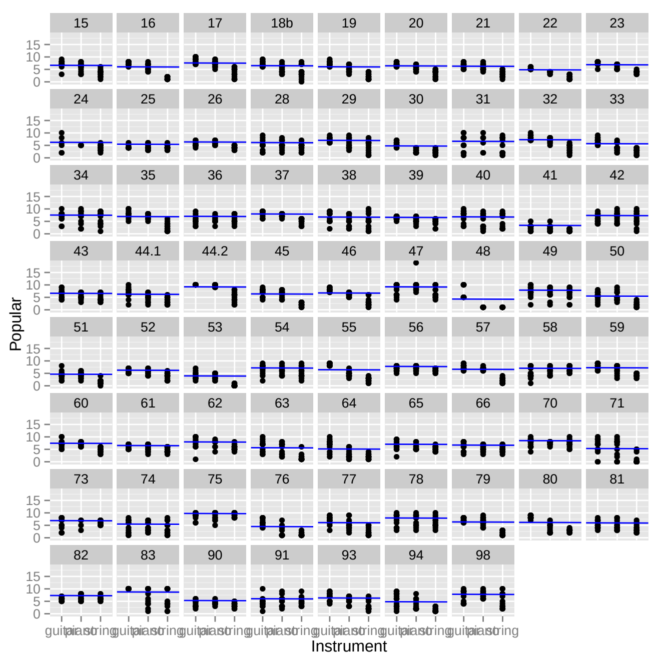
Table 20: Random Effects

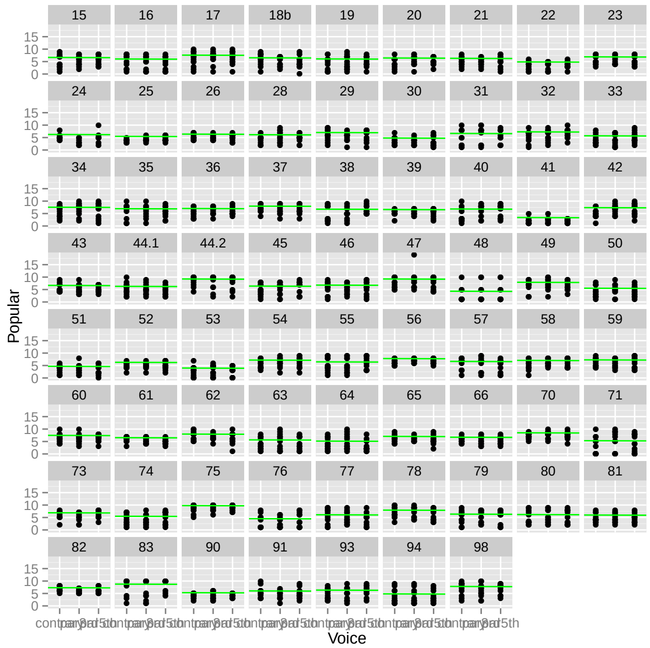
Based on the likelihood ratio test, the random effect of having varying-intercept by subject is statistically significant. Furthermore, we note that the standard deviation by Subject is about two-thirds the standard deviation of the residuals, suggesting that the random intercept is accounting for a reasonable amount of variation. To verify, we compute the intraclass correlation coefficient:

$$\begin{aligned} \rho &= \frac{\tau^2}{\tau^2 + \sigma^2} \\ &= \frac{1.566}{1.566 + 3.532} \\ &= 0.307 \end{aligned}$$

Given that $\rho = 0.307$, we maintain that the hierarchical structure is preferable over a model without the multilevel structure. The plots below illustrate the intercept variation for each of `Harmony`, `Instrument` and `Voice`. We observe that there are differences between the subjects. For instance, between subjects 47 and 48 in the plot of Classical ratings against Harmony. Another noticeable difference is when we examine the plot of Classical ratings against Instrument, where the intercept for subject 75 is greater than that of 74 and 76. If we observe subject 47 in the plot of Classical ratings against Voice, we find that his or her intercept is higher compared to subjects 46 and 48. As such, our observations support the notion that including varying-intercept per participant in the model is useful.







Model 2: Harmony, Instrument, Voice and Three Random Effects

This model takes into account three random effects similar to the ones included in the model for Classical ratings, namely the subject-harmony, subject-instrument and subject-voice combinations. We perform a likelihood ratio test to determine the respective significance of the random effects:

		Chi-square	df	p-value
	Subject:Harmony	88.13	1	< 0.0005
9	Subject: Instrument	498.11	1	< 0.0005
	Subject: Voice	1.25	1	0.3

Table 21: LRT for Random Effects

To compare the overall fit between the two models, we perform an analysis of variance:

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
1	10	10430.30	10488.51	-5205.15	10410.30			
2	12	10078.97	10148.82	-5027.48	10054.97	355.33	2	< 0.0005

Table 22: ANOVA for Model 1 vs Model 2

Our results are in favor of Model 2 which includes the three random effects. We now want to examine a model with additional covariates. Based on our modeling for Classical ratings, we consider including the indicator variable of whether the participant is a self-declared musician. However, we exclude the variable of the frequency to which the participants listen to classical music, since we are interested in ratings for popular music instead of classical music.

Unsurprisingly, there are negligible changes in the magnitude of the estimated coefficients for the fixed effects, but between the two, Model 2 is a better fit for the data than Model 1.

Part (b): Interpretation of Model 2 (3 Random Effects)

	Estimates
(Intercept)	6.58
HarmonyI-V-IV	-0.03
HarmonyI-V-VI	-0.27
HarmonyIV-I-V	-0.19
Instrumentpiano	-0.95
Instrumentstring	-2.61
Voicepar3rd	0.16
Voicepar5th	0.16

Table 23: Fixed Effects for Popular Music

For `Harmony`, Popular ratings decrease when each of the other levels are compared to the baseline, which is the I-VI-V harmonic progression.

When it comes to `Instrument`, ratings decrease when strings or piano is played compared to the baseline instrument, the electric guitar. Unsurprisingly, the decrease is greater for strings than for piano, since as aforementioned, strings are more likely to be affiliated with classical music whereas the piano is versatile. Moreover, Popular music often incorporates the guitar, typically the electric or the acoustic.

For `Voice`, the increase in ratings are similar in magnitude for parallel thirds and parallel fifths when compared to contrary motion. This result supports the notion that contrary motion is more closely associated with classical music.

Groups	Variance	Std Dev
Subject:Harmony	0.41144	0.6414
Subject:Instrument	1.99986	1.4142
Subject:Voice	0.03226	0.1796
Residual	2.49033	1.5781

Table 24: Variance Components and Residual Variance

$$\begin{aligned}\rho_H^2 &= \frac{\tau_H^2}{\tau_H^2 + \sigma^2} \\ &= \frac{0.4114}{0.4114 + 2.4903} \\ &= 0.14\end{aligned}$$

$$\begin{aligned}\rho_I^2 &= \frac{\tau_I^2}{\tau_I^2 + \sigma^2} \\ &= \frac{2.000}{2.000 + 2.4903} \\ &= 0.45\end{aligned}$$

9

$$\rho_V^2 = \frac{\tau_V^2}{\tau_V^2 + \sigma^2}$$
$$= \frac{0.03226}{0.03226 + 2.4903}$$
$$= 0.01$$

Our results are similar to the model for Classical ratings where the subject-voice intercept is not statistically significant compared to the subject-harmony and subject-instrument combinations.

Part (c)

We now include the indicator variable `Musician` which indicates whether or not a participant is a (self-declared) musician to the model. We compare the models from part (a) with this model.

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
1	10	10430.30	10488.51	-5205.15	10410.30			
2	12	10078.97	10148.82	-5027.48	10054.97	355.33	2	< 0.0005
3	13	10074.55	10150.23	-5024.27	10048.55	6.42	1	0.0113

Table 25: ANOVA for M1, M2 and M3

I don't know the definitions of these models.

Of the three models, we find that Models 2 and 3 are better than Model 1. Between Models 2 and 3, the BIC informs us that Model 2 performs better but the other evidence such as AIC and deviance are in favor of Model 3.

I'm not seeing any checks for interaction of musician with other predictors....

7

	Estimates
(Intercept)	6.35
HarmonyI-V-IV	-0.03
HarmonyI-V-VI	-0.27
HarmonyIV-I-V	-0.19
Instrumentpiano	-0.95
Instrumentstring	-2.61
Voicepar3rd	0.16
Voicepar5th	0.16
factor(Musician)1	0.56

Table 26: Fixed Effects for Model 3

While the estimates of the other fixed effects remain unchanged, we find that Popular ratings are expected to increase by 0.56 points when the participant considers himself or herself to be a musician compared to when the participant does not consider himself or herself a musician. We now check the random effects:

Groups	Variance	Std Dev
Subject:Harmony	0.41080	0.6409
Subject:Instrument	1.94294	1.3939
Subject:Voice	0.03211	0.1792
Residual	2.49048	1.5781

Table 27: Variance Components and Residual Variance

Without recalculating the intraclass correlation coefficients, we can estimate by inspection that we would obtain similar results since the change in the variance of the random effects are negligible compared to the variance in Model 2.

Section 5: Brief writeup

We want to investigate the factors that influence the ratings for Classical music and the ratings for Popular music. Given that there were 70 participants in the experiment, it is reasonable to assume that they are similar to one another in some ways and different from one another in other ways. For our purposes, we aim to find out factors that may influence the way they rate the musical stimuli.

Influence of Harmony, Instrument & Voice on Classical vs Popular Ratings

Harmony

It is hypothesized that the I-V-vi harmonic progression is associated with classical music and our analysis does support the conjecture that *Classical* ratings *increase* for this specific harmonic progression whereas **Popular** ratings **decrease** when participants are presented with the same harmonic progression.

Instrument

10 Of the three primary experimental factors, it is believed that instrument should have the largest impact on ratings. For participants who have not been exposed to any musical knowledge, harmonic progressions and voice leading are subtle musical changes that are not immediately perceptible. However, instruments produce distinct sounds, especially the ones we have in the experiment, namely the piano, the string quartet and the electric guitar.

It is no surprise then that we find a decrease in Classical ratings when the electric guitar is the instrument in the musical stimuli represented but Popular ratings increase. On the other hand, Popular ratings decrease when the instrument is either the piano or the strings, but the magnitude of decrease is larger when participants hear the string quartet than when they hear the piano. Naturally, strings and piano increase Classical ratings.

Voice

We find that contrary notion does indeed increase Classical ratings but decrease Popular ratings. The increase in Popular ratings is the same when the voice lead is in either parallel thirds or parallel fifths whereas the decrease in Classical ratings differs by only a small amount when the voice lead is in either parallel thirds or parallel fifths. Our results support the notion that voice lead can be harder for the subjects to distinguish between the two, but contrary motion is more distinct that the other two categories of voice lead.

Random Effects

The sizes of the variance component for Subject-Harmony and Subject-Voice combinations are small, at least when compared to the overall variance. However, the size of the variance component for Subject-Instrument is almost the size of the residual variance. This informs us that the random effect is accounting for an important variation in ratings at the Subject-Instrument level for both Classical and Popular ratings.

Other Covariates

For Classical ratings, two additional covariates which help us understand ratings would be whether or not the participant regards himself or herself as a musician, and the frequency to which he or she listens to classical music. For Popular ratings, we find that ratings are influenced by whether or not the participant declares himself or herself to be a musician.

Appendix: R Code

```
library(foreign)
library(arm)
library(leaps)
library(lme4)
library(ggplot2)
library(RLRsim)
library(lmerTest)

# Load data set
ratings <- read.csv("ratings.csv")
attach(ratings)
```

Problem 1

```
##### Part 1(a) #####
# Full linear model
lm.full <- lm(Classical ~ Harmony + Instrument + Voice, data=ratings)

# Linear model without Harmony
lm.xharm <- lm(Classical ~ Instrument + Voice, data=ratings)

# Linear model without Instrument
lm.xinstr <- lm(Classical ~ Harmony + Voice, data=ratings)

# Linear model without Voice
lm.xvoice <- lm(Classical ~ Harmony + Instrument, data=ratings)

##### Part 1(b) #####
## (ii)
lmer1bii <- lmer(Classical ~ 1 + Harmony + Instrument + Voice +
                 (1|Subject), data=ratings)

facet.sub <- split(ratings, ratings$Subject)

# Extract random intercept for each subject
a0.bii <- fixef(lmer1bii)[1] + ranef(lmer1bii)$Subject[,1]

# Classical ratings vs Harmony
plot1biiH <- ggplot(ratings, aes(Harmony, Classical)) + geom_point() +
  facet_wrap(~ Subject)

for (j in 1:length(facet.sub)) {
  plot1biiH <- plot1biiH + geom_abline(data=facet.sub[[j]],
                                 intercept=a0.bii[j], slope=fixef(lmer1bii)[2],
                                 color="red")
}
plot1biiH

# Classical ratings vs Instrument
plot1biiI <- ggplot(ratings, aes(Instrument, Classical)) + geom_point() +
```

```
  facet_wrap(~ Subject)

for (j in 1:length(facet.sub)) {
  plot1biiI <- plot1biiI + geom_abline(data=facet.sub[[j]],
                                       intercept=a0.bii[j], slope=fixef(lmer1bii)[2],
                                       color="blue")
}
plot1biiI

# Classical ratings vs Voice
plot1biiV <- ggplot(ratings, aes(Voice, Classical)) + geom_point() +
  facet_wrap(~ Subject)

for (j in 1:length(facet.sub)) {
  plot1biiV <- plot1biiV + geom_abline(data=facet.sub[[j]],
                                       intercept=a0.bii[j], slope=fixef(lmer1bii)[2],
                                       color="green")
}
plot1biiV

##### Part 1(c) #####
## (i)
lmer1ci <- lmer(Classical ~ 1 + Harmony + Instrument + Voice +
                  (1|Subject:Harmony) + (1|Subject:Instrument) +
                  (1|Subject:Voice), data=ratings, REML=FALSE)
```

Problem 2

```
##### Part 2(a) #####
lmer.ext1 <- lmer(Classical ~ Harmony + Instrument + Voice +
                factor(Selfdeclare) + OMSI + factor(ClsListen) +
                  (1|Subject:Harmony) + (1|Subject:Instrument) +
                  (1|Subject:Voice),
              data=ratings)

lmer.ext2 <- lmer(Classical ~ Harmony + Instrument + Voice +
                factor(Selfdeclare) + (1|Subject:Harmony) +
                  (1|Subject:Instrument) + (1|Subject:Voice),
              data=ratings)

lmer.ext3 <- lmer(Classical ~ Harmony + Instrument + Voice +
                OMSI +
                  (1|Subject:Harmony) + (1|Subject:Instrument) +
                  (1|Subject:Voice), data=ratings)

lmer.ext4 <- lmer(Classical ~ Harmony + Instrument + Voice +
                factor(ClsListen) + (1|Subject:Harmony) +
                  (1|Subject:Instrument) + (1|Subject:Voice),
              data=ratings)

lmer.ext5 <- lmer(Classical ~ Harmony + Instrument + Voice +
                factor(Selfdeclare) + factor(ClsListen) +
```

```
                  (1|Subject:Harmony) + (1|Subject:Instrument) +
                  (1|Subject:Voice), data=ratings)

rand(lmer.ext5)
```

Problem 3

```
# Self-declare indicator
ratings$Musician <- ifelse(ratings$Selfdeclare %in% c(1,2), 0, 1)

lmer3.1 <- lmer(Classical ~ Harmony + Instrument + Voice +
                factor(ClsListen) + factor(Musician) +
                (1|Subject:Harmony) + (1|Subject:Instrument) +
                (1|Subject:Voice), data=ratings)

lmer3.2 <- lmer(Classical ~ Harmony + Instrument + Voice +
                 factor(ClsListen) + factor(Musician) +
                 Harmony*factor(Musician) +
                 (1|Subject:Harmony) + (1|Subject:Instrument) +
                 (1|Subject:Voice), data=ratings)

lmer3.3 <- lmer(Classical ~ Harmony + Instrument + Voice +
                  factor(ClsListen) + factor(Musician) +
                  Instrument*factor(Musician) +
                  (1|Subject:Harmony) + (1|Subject:Instrument) +
                  (1|Subject:Voice), data=ratings)

lmer3.4 <- lmer(Classical ~ Harmony + Instrument + Voice +
                  factor(ClsListen) + factor(Musician) +
                  Voice*factor(Musician) +
                  (1|Subject:Harmony) + (1|Subject:Instrument) +
                  (1|Subject:Voice), data=ratings)
```

Problem 4

```
lmer4a.1 <- lmer(Popular ~ Harmony + Instrument + Voice +
                (1|Subject), data=ratings)

lmer4a.2 <- lmer(Popular ~ Harmony + Instrument + Voice +
                   (1|Subject:Harmony) + (1|Subject:Instrument) +
                   (1|Subject:Voice), data=ratings)

# Extract random intercept for each subject (lmer4a.1)
a0.4a1 <- fixef(lmer4a.1)[1] + ranef(lmer4a.1)$Subject[,1]

# Popular ratings vs Harmony
plot4a1H <- ggplot(ratings, aes(Harmony, Popular)) + geom_point() +
  facet_wrap(~ Subject)

for (j in 1:length(facet.sub)) {
  plot4a1H <- plot4a1H + geom_abline(data=facet.sub[[j]],
                                      intercept=a0.4a1[j],
```

```
                                     slope=fixef(lmer4a.1)[2],
                                       color="red")
}
plot4a1H

# Popular ratings vs Instrument
plot4a1I <- ggplot(ratings, aes(Instrument, Popular)) + geom_point() +
  facet_wrap(~ Subject)

for (j in 1:length(facet.sub)) {
  plot4a1I <- plot4a1I + geom_abline(data=facet.sub[[j]],
                                       intercept=a0.4a1[j],
                                     slope=fixef(lmer4a.1)[2],
                                       color="blue")
}
plot4a1I


# Popular ratings vs Voice
plot4a1V <- ggplot(ratings, aes(Voice, Popular)) + geom_point() +
  facet_wrap(~ Subject)

for (j in 1:length(facet.sub)) {
  plot4a1V <- plot4a1V + geom_abline(data=facet.sub[[j]],
                                       intercept=a0.4a1[j],
                                     slope=fixef(lmer4a.1)[2],
                                       color="green")
}
plot4a1V

lmer4c <- lmer(Popular ~ Harmony + Instrument + Voice +
                 factor(Musician) + (1|Subject:Harmony) +
                 (1|Subject:Instrument) + (1|Subject:Voice),
               data=ratings)
```