R Notebook

Contents

Introduction

In this study, we are interested in how auditory distraction affects cognitive flexibility, which is the ability of brain to shift between thinking about different problem.

For conducting this experiment, different music is regarded as different levels of auditory distraction. Candidates will perform stroop test when listening to different music and time used to complete On-run test and off-run test are recorded. In addition, other variables may affect stroop-test performance, like sleep time before taking the test and device used for test, are also recorded.

Method

Differences between stroop-On run and stroop-off run with same level of distraction is taken as response variable. Smaller values represent better cognitive performance. Since the research question is interested in how auditory distraction affect cognitive flexibility, distraction level is taken in as fixed effect. In addition, since each candidates contribute to 3 observations, observations from same candidates are apparently not independent, so candidates id is taken into the model as random effect.

Data Cleaning

Use cleaned data posted on Quercus.

Then look for outliers.

Use boxplot to check outlier, remove those candidates with more than 1 outlying response observation.

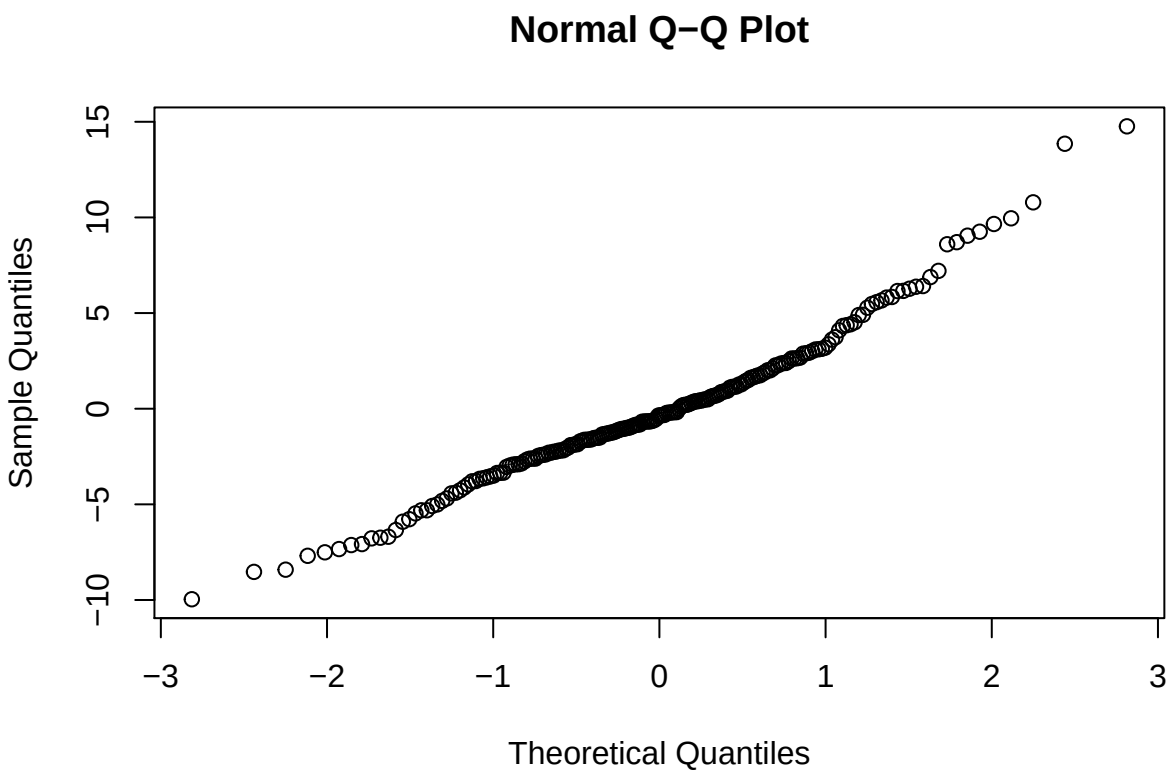
Procedure: find the minimum outlier value, sort values of **Ontime_minus_Offtime** in decreasing order, then read the data table and try to find if there are candidates with more than 1 outlying response observation. Remove all observations from those candidates.

Candidates 31,35,50,72 are removed at this stage. Relevent boxplot are included in appendix.

In addition, **order** is qualitative value in the original dataset, convert to factor first.

Fitting initial model

Fit the initial and simplest model with only the interested fixed effect and the random effect to take a first look on possible model. Default reference group of distraction_level is classical, but there is a control group, so relevel **longdata$distraction_level** to make 'control' the reference group at first.



	MLE	Std.Error	DF	t-value	p-value
(Intercept)	7.328368	0.6282940	134	11.663915	0.0000000
distraction_levelclassical	-1.713985	0.7609499	134	-2.252429	0.0259213
distraction_levellyrics	-1.283147	0.7609499	134	-1.686244	0.0940758
σ	2.675015	NA	NA	NA	NA
τ	4.437062	NA	NA	NA	NA

	numDF	denDF	F-value	p-value
(Intercept)	1	134	198.575649	0.0000000
distraction_level	2	134	2.745806	0.0678109

For this initial model, residual is normal distributed and p-value is significant. Though this is not sufficient for verification of the model, we may stop here about this model since this model is only for taking a glance.

Backward Selecition

The initial model is too simple, we have more variables that may affect the response variable. It's hard to tell which variables are supposed to be included in the model and which variables are not. At this point, I decided to utlize backward selection technique I learnt from machine learning course, STA314.

Backward selection technique:

1.Start with all variables in the model

2.Remove the variable with the largest p-value, which means, the variable that is least statistically significant.

3.The new (p-1)-variable model is fit ,and the variable with the largest p-value is removed.

4.Repeat step 2&3 until a stopping rule is reached. For this analysis, I decided to stop when all remaining variables have a significant p-value.

In addition to independent variables given in the dataset, note that I also add the interaction terms of **factor(order)** and **distraction_level** into the fullmodel. Full model is shown below. More table and plot about full model can be found in appendix.

```
fit_full <- lme(OnTime_minus_OffTime ~ distraction_level + video_games + device + headphones + yrs_engli
            + start_time + factor(order)
            + distraction_level*factor(order),
            random = ~1|id, data = longdata)
```

Only final model are shown in method section. Intermediate runs are included in appendix.

Table 3: Anova table of the final model

	numDF	denDF	F-value	p-value
(Intercept)	1	134	228.187205	0.0000000
distraction_level	2	134	2.745810	0.0678106
video_games	1	63	4.974283	0.0292972
headphones	3	63	3.005653	0.0368492

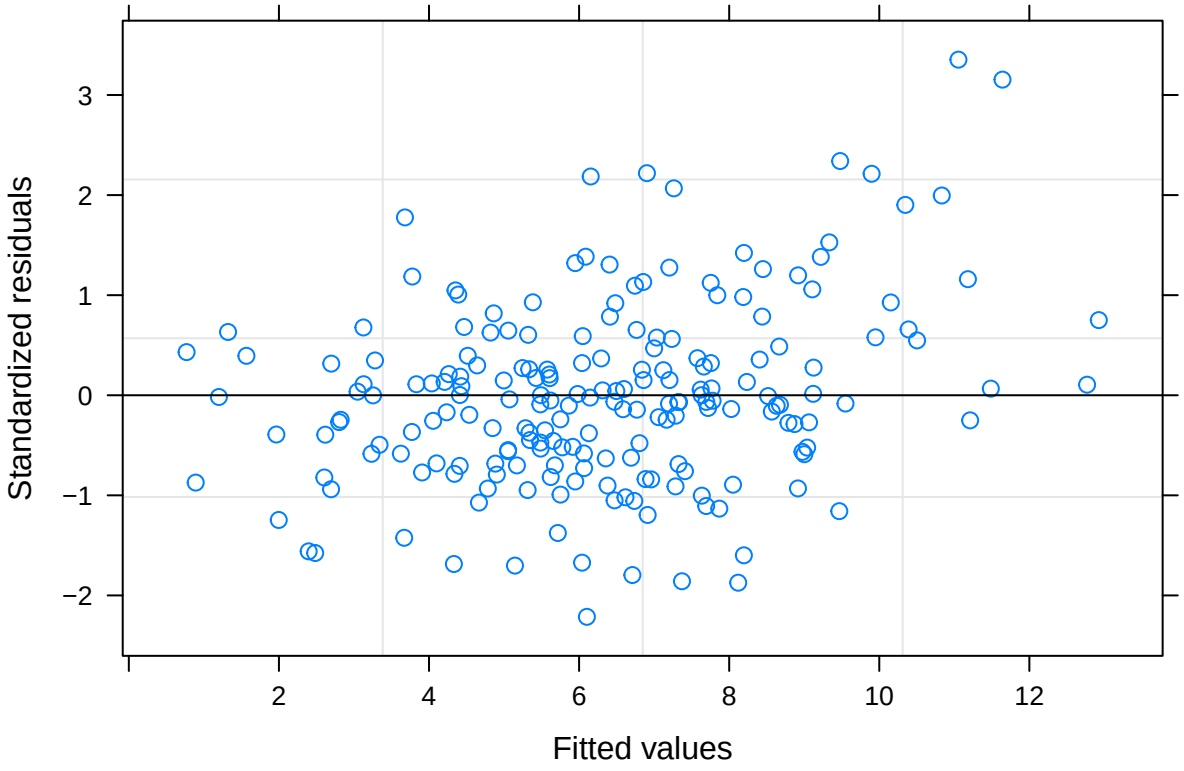
Table 4: lmeTable of the final model

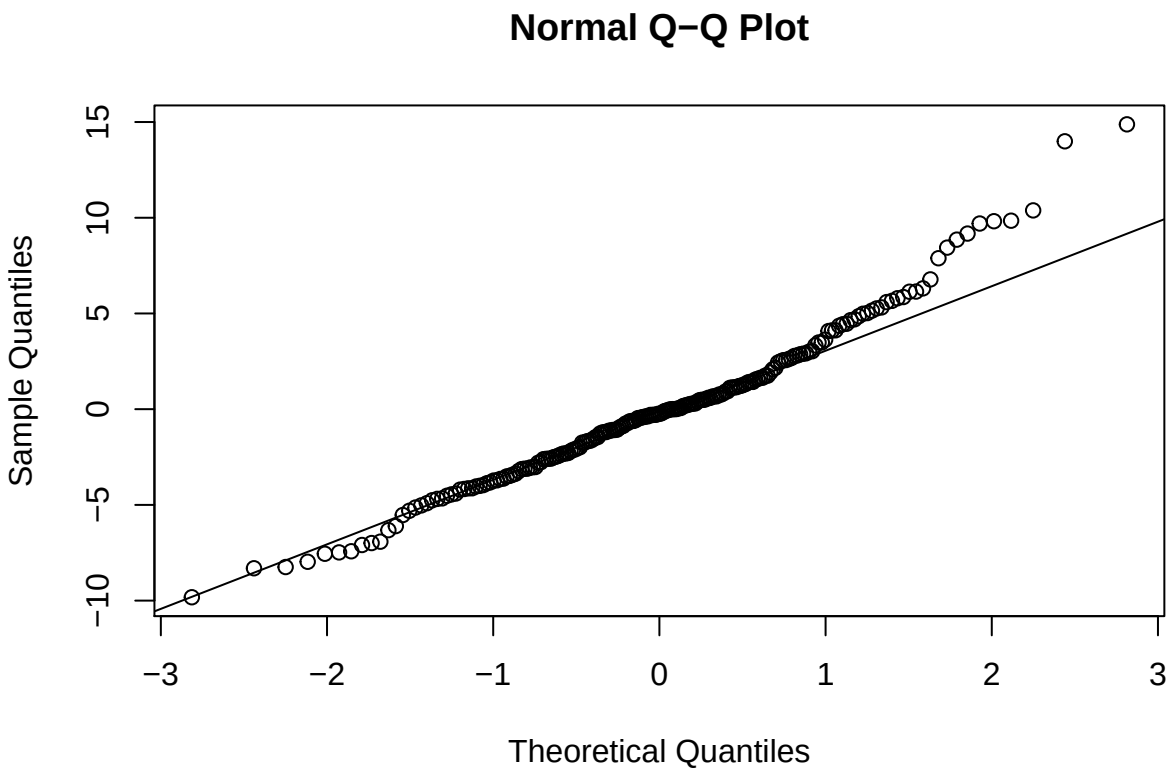
	MLE	Std.Error	DF	t-value	p-value
(Intercept)	5.581207	1.3678482	134	4.0802824	0.0000768
distraction_levelclassical	-1.713985	0.7609494	134	-2.2524301	0.0259212
distraction_levellyrics	-1.283147	0.7609494	134	-1.6862450	0.0940756
video_gamesYes	-1.689577	0.8893717	63	-1.8997427	0.0620464
headphonesIn-ear headphones; not noise cancelling	2.916917	1.3023429	63	2.2397457	0.0286448
headphonesOver-ear headphones; noise cancelling	1.181057	1.6407174	63	0.7198417	0.4742846
headphonesOver-ear headphones; not noise cancelling	4.535393	1.7417437	63	2.6039380	0.0114805
σ	2.318517	NA	NA	NA	NA
τ	4.437059	NA	NA	NA	NA

In the final model, all the remaining predictors have significant p-values.

Model Verification

Now need to test whether the model meet its assumptions.





Result

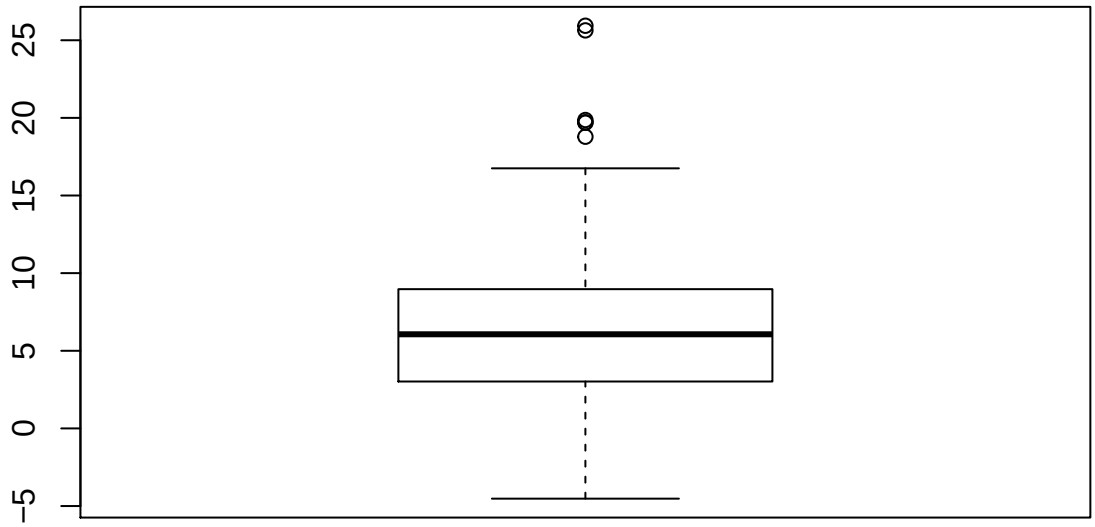
From the final model, we may conclude that in this experiment, distraction levels, video_games, which stands for whether the candidate plays video games often, and headphones, which stands for type of headphone candidates using when doing the test, have significant effect on cognitive flexiblity. More specifically, candidates who plays video games often will have a smaller **Onrun_minus_Offrun** values, which means they show better cognitive flexibility.

Discussion

Backward selection method from machine learning course is used for this analysis. Since this method only consider p-value for variable selection, variables that are in the final model may not be the best choice for understanding research question. In addition, not much interaction terms are introduced into the model, which means there might be better fit e

Appendix

Clean outliers




```
## [1] "minimum outlier: 18.787"
```

```
## [1] "minimum outlier: 18.787"
```

```
## [1] "minimum outlier: 18.787"
```

Backward Selection full model

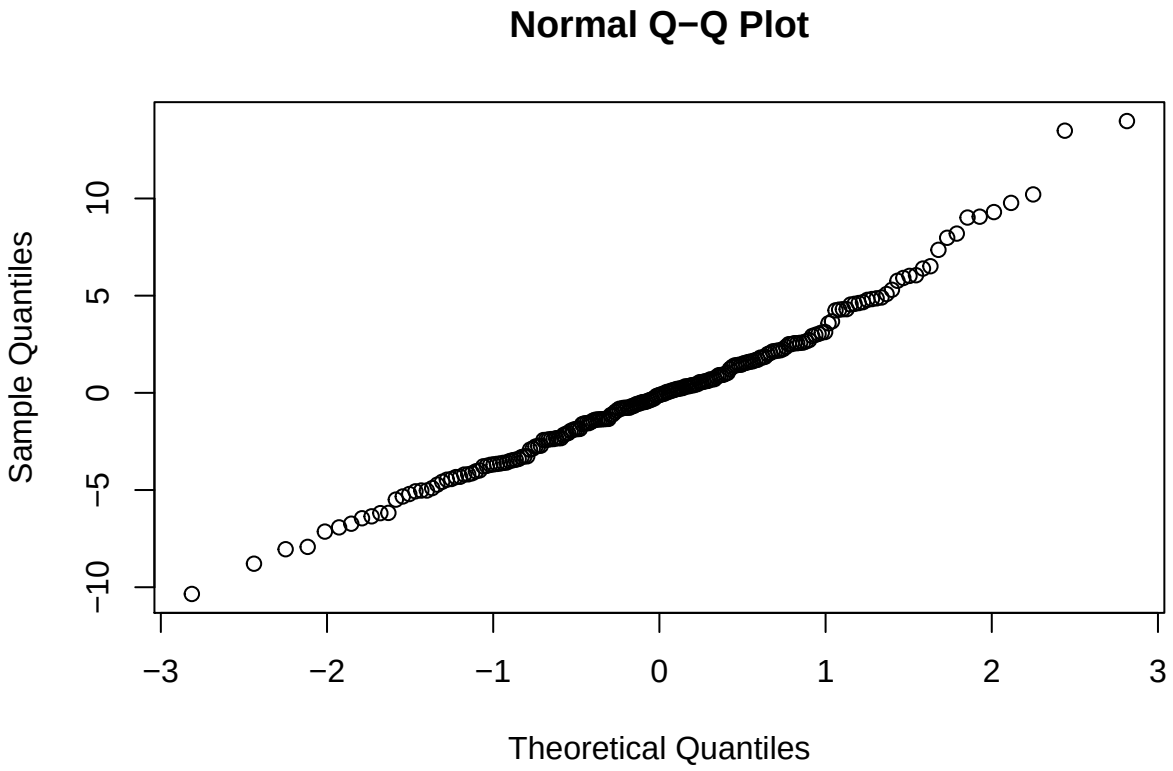


Table 5: lmerTable of initial fullmodel for backward selection

	MLE	Std.Error	DF	t-value	p-value
(Intercept)	8.0726321	3.6174129	125	2.2316037	0.0274230
distraction_levelclassical	1.8140239	2.1158970	125	0.8573309	0.3929028
distraction_levellyrics	-1.6953334	1.7056251	125	-0.9939660	0.3221596
video_gamesYes	-1.7444998	0.9187794	60	-1.8987147	0.0624139
deviceiPad tablet	-1.7769996	1.5943056	60	-1.1145916	0.2694703
deviceiPhone / iPod	-1.0829951	1.3422761	60	-0.8068348	0.4229484
headphonesIn-ear headphones; not noise cancelling	2.8392745	1.3930183	60	2.0382177	0.0459432
headphonesOver-ear headphones; noise cancelling	1.1635047	1.7440618	60	0.6671235	0.5072509
headphonesOver-ear headphones; not noise cancelling	5.1150170	1.9206206	60	2.6632106	0.0099227
yrs_english	0.0047428	0.1007765	60	0.0470625	0.9626197
sleep	-0.0845781	0.3501737	125	-0.2415320	0.8095386
start_timeevening	-1.4327885	1.0525135	125	-1.3613018	0.1758672
start_timemorning	0.8659226	1.3341589	125	0.6490400	0.5175032
factor(order)2	-1.1982595	1.9845656	125	-0.6037893	0.5470779
factor(order)3	-1.8389686	1.7766762	125	-1.0350612	0.3026386
distraction_levelclassical:factor(order)2	-3.2846087	3.1331046	125	-1.0483559	0.2964975
distraction_levellyrics:factor(order)2	1.3561070	2.7756230	125	0.4885775	0.6259969
distraction_levelclassical:factor(order)3	-2.4149133	2.8979421	125	-0.8333201	0.4062537
distraction_levellyrics:factor(order)3	1.4562272	2.7357533	125	0.5322948	0.5954667
σ	2.3990712	NA	NA	NA	NA

	MLE	Std.Error	DF	t-value	p-value
τ	4.4210479	NA	NA	NA	NA

Table 6: Anova table of initial full model for backward selection

	numDF	denDF	F-value	p-value
(Intercept)	1	125	221.9994486	0.0000000
distraction_level	2	125	2.7657342	0.0667843
video_games	1	60	4.8393955	0.0316795
device	2	60	1.7945040	0.1750153
headphones	3	60	2.0940597	0.1104666
yrs_english	1	60	0.0004598	0.9829635
sleep	1	125	0.0486605	0.8257706
start_time	2	125	1.8673489	0.1588180
factor(order)	2	125	1.8778850	0.1572016
distraction_level:factor(order)	4	125	0.6564130	0.6234390

Backward Selection intermediate

```
## [1] "remove yrs_english"
```

Table 7: Anova table for model after iteration 1 of backward selection

	numDF	denDF	F-value	p-value
(Intercept)	1	125	225.8771964	0.0000000
distraction_level	2	125	2.7661367	0.0667586
video_games	1	61	4.9239271	0.0302174
device	2	61	1.8258493	0.1697743
headphones	3	61	2.1306374	0.1055556
sleep	1	125	0.0504543	0.8226414
start_time	2	125	1.9022766	0.1535235
factor(order)	2	125	1.8787841	0.1570645
distraction_level:factor(order)	4	125	0.6559295	0.6237765

```
## [1] "remove sleep"
```

Table 8: Anova table for model after iteration 2 of backward selection

	numDF	denDF	F-value	p-value
(Intercept)	1	126	229.3687136	0.0000000
distraction_level	2	126	2.7671790	0.0666614
video_games	1	61	5.0000391	0.0290154
device	2	61	1.8540725	0.1653149
headphones	3	61	2.1635720	0.1014780
start_time	2	126	1.9483654	0.1467765
factor(order)	2	126	1.8859656	0.1559394
distraction_level:factor(order)	4	126	0.6412957	0.6340209

```
## [1] "remove interaction term"
```

Table 9: Anova table for model after iteration 3 of backward selection

	numDF	denDF	F-value	p-value
(Intercept)	1	130	235.028520	0.0000000
distraction_level	2	130	2.767283	0.0665370
video_games	1	61	5.123418	0.0271742
device	2	61	1.899823	0.1583415
headphones	3	61	2.216959	0.0951989
start_time	2	130	1.999780	0.1395079
factor(order)	2	130	1.886983	0.1556553

```
## [1] "remove device"
```

Table 10: Anova table for model after iteration 4 of backward selection

	numDF	denDF	F-value	p-value
(Intercept)	1	130	236.713505	0.0000000
distraction_level	2	130	2.769243	0.0664120
video_games	1	63	5.160149	0.0265360
headphones	3	63	3.117961	0.0322270
start_time	2	130	1.875024	0.1574748
factor(order)	2	130	1.873787	0.1576644

```
## [1] "remove factor(order)"
```

Table 11: Anova table for model after iteration 5 of backward selection

	numDF	denDF	F-value	p-value
(Intercept)	1	132	236.484870	0.0000000
distraction_level	2	132	2.734698	0.0685933
video_games	1	63	5.155165	0.0266063
headphones	3	63	3.114949	0.0323430
start_time	2	132	1.874309	0.1575223

```
## [1] "remove start_time"
```

Table 12: Anova table for model after iteration 6 of backward selection

	numDF	denDF	F-value	p-value
(Intercept)	1	134	228.187205	0.0000000
distraction_level	2	134	2.745810	0.0678106
video_games	1	63	4.974283	0.0292972
headphones	3	63	3.005653	0.0368492