STA490 Data Analysis

```r
library("knitr")
library("nlme")
library("tidyverse")
```

```
## -- Attaching packages --------------------------------------------------------------------------- tidyve
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
## -- Conflicts ------------------------------------------------------------------------------ tidyverse_co
## x dplyr::collapse() masks nlme::collapse()
## x dplyr::filter()   masks stats::filter()
## x dplyr::lag()      masks stats::lag()
```

```r
data.long = read.csv("EDA/data/cleaned_data/cleandata_long.csv")
data.wide = read.csv("EDA/data/cleaned_data/cleandata_wide.csv")
```

Summary

I wanted to investigate the effects of auditory distraction on the cognitive flexibility. In this experiment design, auditory distraction occurs in three levels: music with lyrics, music with no lyrices and control (quiet). Whereas the cognitive flexibility is represented by he difference between OnTime and OffTime, the higher the difference, the lower the cognitive flexibility.

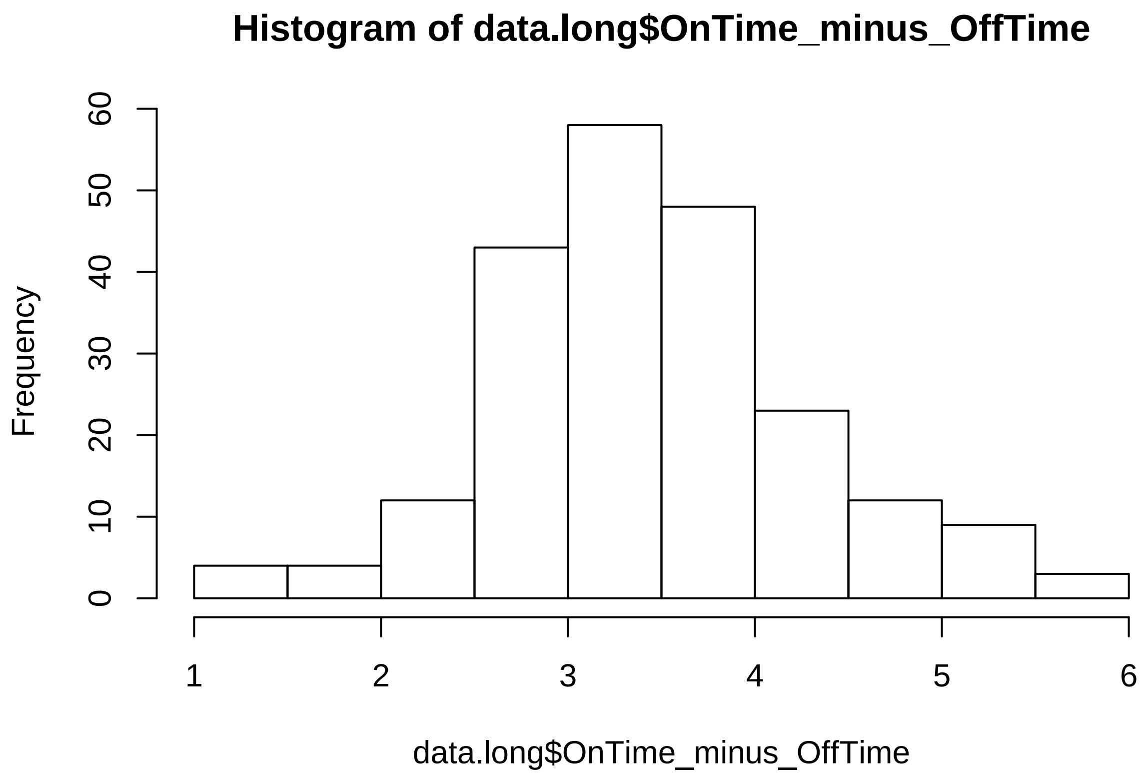
From Model 1, I have found that there is no difference in cognitive flexibility of the subjects when they are listening to either music with lyrics or without lyrics (classical). However, the cognitive flexibility of the subjects decreases when they are in a quiet setting.

Research Question: The effects of auditory distraction on the cognitive flexibility.

The model we consider for subject i (i=1,. . . . .,72) at the three levels of auditory distractions (control, lyrics and classical).

For model 1, the random effect is assumed to follow a normal distribution with independent of the error term. Model 1 assumes that the intercepts differ randomly across the 72 subjects, but the Difference changes with the levels of distractions.

```r
# Model 1
data.long = read.csv("EDA/data/cleaned_data/cleandata_long.csv")
data.long$OnTime_minus_OffTime = (data.long$OnTime_minus_OffTime -
                                    min(data.long$OnTime_minus_OffTime) + 1)^(1/2)
hist(data.long$OnTime_minus_OffTime)
```



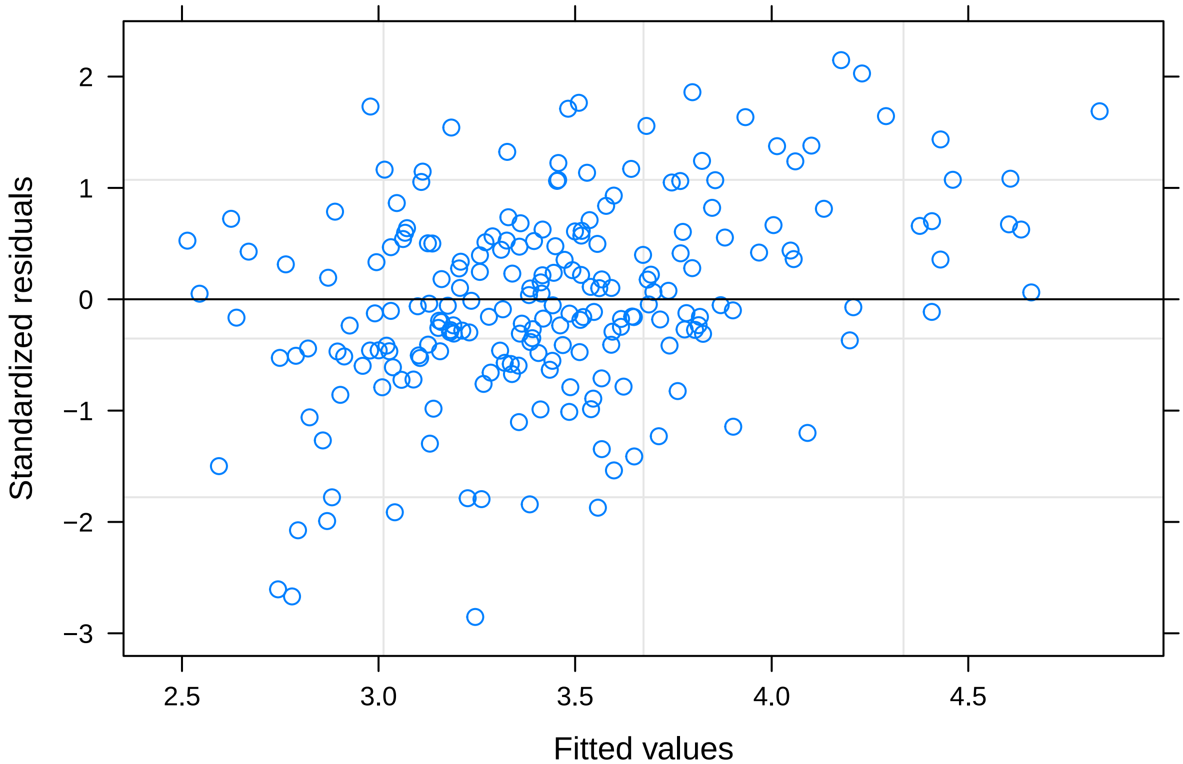

```r
mod1 = nlme::lme(OnTime_minus_OffTime ~ distraction_level,random = ~1|id,data = data.long)
```

1.1 Verify model assumptions and fit

There are a few model assumptions I want to check: 1. Homogeneity of variance 2. Normality of error term 3. Normality of random effects 4. Independence/Constant Correlation

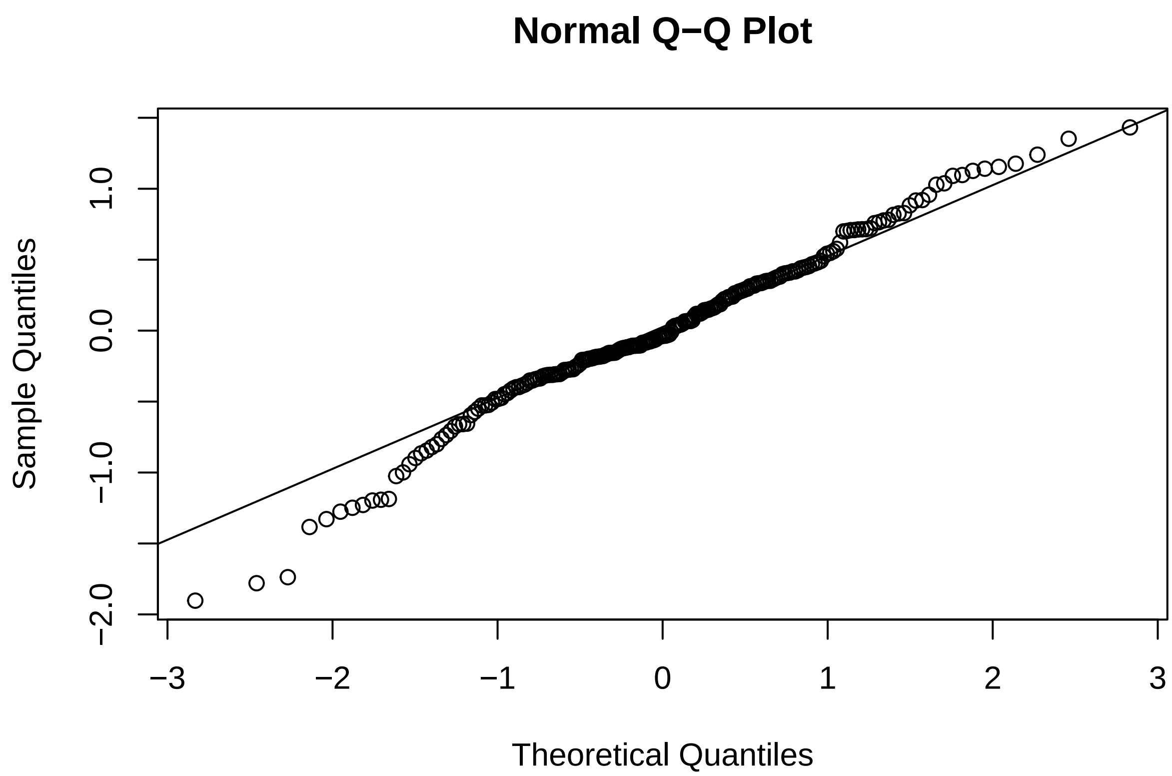
After an initial verification, I have found that the assumptions of normality for both the error term and random effects are violated. Looking at the shape of the histogram formed from the data, since there are some negative values in the data, I have added the smallest value of the data + 1 to all of my values, so that I can retain the shape of the distribution, and when I apply square root transformation on the data, I wouldn't get NA values.

```r
# Homogeneity of variance
plot(mod1)
```

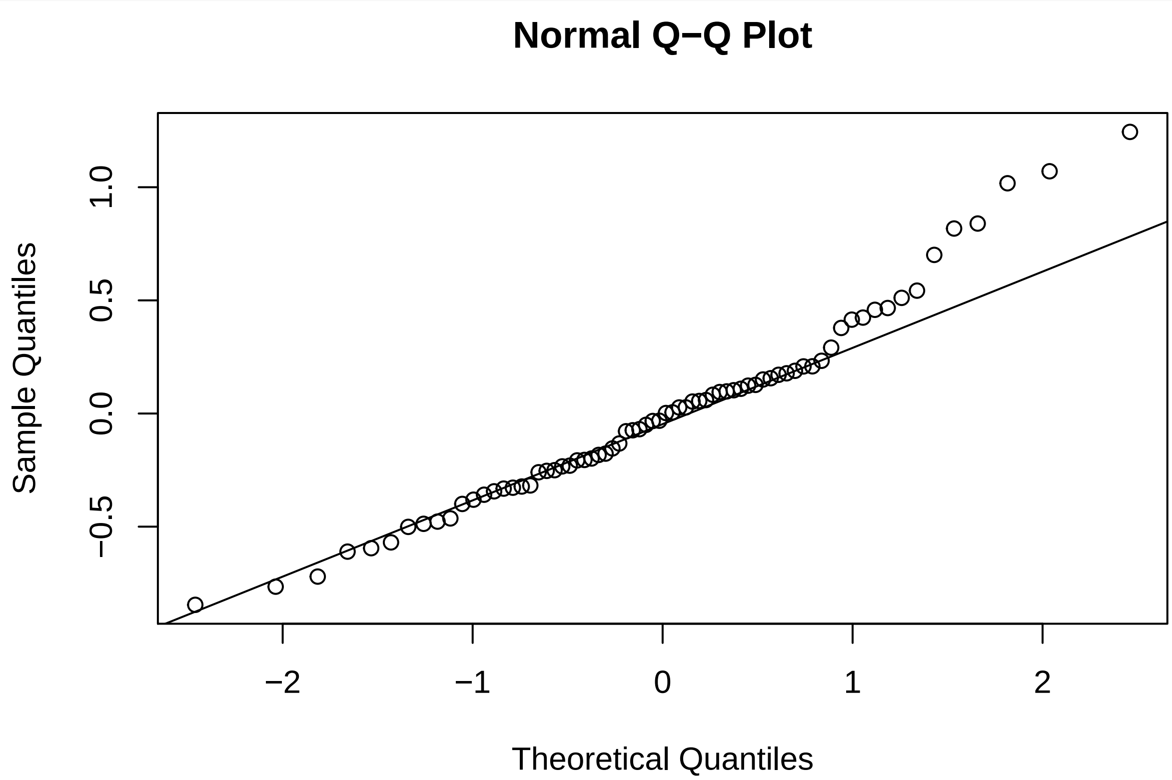


This residual plot does not indicate any deviations from a linear form. It also shows relatively constant variance across the fitted range. The slight reduction in apparent variance on the right f the graph are likely a result of there being fewer observation in these predicted areas.

```
# Normality of error term
qqnorm(residuals(mod1))
qqline(residuals(mod1))
```




```r
#Normality of random effects
qqnorm(ranef(mod1)$`(Intercept)`)
qqline(ranef(mod1)$`(Intercept)`)
```



The Q-Q Normal Plots do not raise any significant concern with normality of the weighted residuals.

```r
# Independence

LevelCon <- data.long$OnTime_minus_OffTime[data.long$distraction_level=="control"]
LevelLyr <- data.long$OnTime_minus_OffTime[data.long$distraction_level=="lyrics"]
LevelCla <- data.long$OnTime_minus_OffTime[data.long$distraction_level=="classical"]
T <- cbind(LevelCon,LevelLyr,LevelCla)
c<-cor(T)
round(c,3)
```

```
##          LevelCon LevelLyr LevelCla
## LevelCon    1.000    0.360    0.418
## LevelLyr    0.360    1.000    0.382
## LevelCla    0.418    0.382    1.000
```

The similarity among the correlations away from the diagonal, which range from 0.360 to 0.418. The correlation seems a bit weak, but verall, it seems that the assumption that correlations are constant across Levels of Distraction is a reasonable one.

1.2 Interpret results

```r
summary(mod1)
```

```
## Linear mixed-effects model fit by REML
##  Data: data.long
##        AIC      BIC    logLik
##   529.6608 546.4673 -259.8304
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept)  Residual
## StdDev:   0.5270718 0.6668805
##
## Fixed effects: OnTime_minus_OffTime ~ distraction_level
##                              Value Std.Error  DF  t-value p-value
## (Intercept)               3.359146 0.1001758 142 33.53250  0.0000
## distraction_levelcontrol 0.230589 0.1111468 142  2.07463  0.0398
## distraction_levellyrics  0.031013 0.1111468 142  0.27903  0.7806
##  Correlation:
##                          (Intr) dstrctn_lvlc
## distraction_levelcontrol -0.555
## distraction_levellyrics  -0.555  0.500
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.85333680 -0.46714018 -0.05618478  0.54413602  2.14747909
##
## Number of Observations: 216
## Number of Groups: 72
```

```r
knitr::kable(VarCorr(mod1))
```

	Variance	StdDev
(Intercept)	0.2778047	0.5270718

	Variance	StdDev
Residual	0.4447296	0.6668805

Looking at the estimates of variance, the error variance is estimated to be 0.5270718ˆ2 = 0.2778 while the variance due to the random intercept is estimated to be 0.6668805ˆ2 = 0.4447. The correlation of two measurements within the same subject is estimated to be: 0.2778/(0.2778+0.4447) = 38.45 %.

The differences among subjects attributes 38.45% of the variability in Difference.

```
knitr::kable(summary(mod1)$tTable[,-3],digital=3)
```

	Value	Std.Error	t-value	p-value
(Intercept)	3.3591462	0.1001758	33.5324979	0.0000000
distraction_levelcontrol	0.2305888	0.1111468	2.0746339	0.0398256
distraction_levellyrics	0.0310128	0.1111468	0.2790253	0.7806314

Looking at the fixed effects due to Levels of Distractions in Model 1, the p-value for lyrics group is 0.7806, indicating that it is not statistically significant, hence I can interpret that there is no significant difference in the cognitive flexibility whether we are listening to classical music and music with lyrics. The p-value for control group is 0.03983,indicating that it is statistically significant. This shows that there is significant difference in the cognitive flexibility when we are listening to music and when we are not listening to music. It is also interesting to point out that our cognitive flexibility actually increases when we are listening to music.

In summary, the correlation between two Difference measurements for 72 subjects is both relatively constant across different levels of distraction and moderate (approximately 0.38). In addition, the fixed effect for the control group in Model 1 is also highly statistically significant, while that for the lyrics group is not statistically significant.

1.3 Limitations

There are a few limitations to the study.

1. In the model, the variance due to random effect is relatively small and the error variance is relatively large, when this occurs, we can say that there is a "shrinkage". This means that there is less subject-level variance relative to the population variance, and thus the mixed model will produce a group-specific effect that is closer to the overall population effect.

2. By fitting individual id as a random intercept, although differences among individuals can be obtained, this also uses a degree of freedom for each individual id, severely limiting the power of the model. Thirdly,

3. In the model, we have applied transformation to the non-Gaussian data to fit the Gaussian model, and such transformation could affect the results and make it harder to detect the real effects.